

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

## Предсказательная аналитика и системы поддержки принятия решений

Аналитический сборник №5

СЕНТЯБРЬ  
2020



Все права принадлежат Центру компетенций Национальной технологической инициативы на базе МФТИ по направлению «Искусственный интеллект». Распространение Альманаха разрешено только в виде публикации ссылки на сайт Альманаха <http://www.aiReport.ru>, копирование и перепост содержимого Альманаха или его части или выкладывание файла с Альманахом или его части без письменного согласования правообладателя запрещается.

Данный отчет сверстан для просмотра на экране компьютера формата 16×9. Печатную версию данного Альманаха можно заказать по ссылке ниже.

Сайт



[www.AIreport.ru](http://www.AIreport.ru)

Печатная версия



[www.ozon.ru](http://www.ozon.ru)



OZN191961238

Альманах Искусственный Интеллект - Предиктивная аналитика и системы поддержки принятия решений

# Альманах ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

## Предсказательная аналитика и системы поддержки принятия решений

Аналитический сборник

Москва / №5 / Сентябрь 2020



Центр компетенций НТИ  
«Искусственный интеллект»



Skoltech  
Skolkovo Institute of Science and Technology



# Содержание

Центр Национальной технологической инициативы на базе МФТИ по направлению «Искусственный интеллект».....	6	<b>ПРИМЕНЕНИЯ</b> .....	<b>73</b>	<b>КЛЮЧЕВЫЕ КОМПАНИИ МИРА</b> .....	<b>123</b>	<b>КЛЮЧЕВЫЕ КОМПАНИИ РОССИИ</b> .....	<b>137</b>
От редактора.....	8	Big Data в современном ритейле: предиктивные технологии для роста Retention и LTV.....	74	Топ компании мира.....	124	Топ компании России.....	138
<b>ВВЕДЕНИЕ</b> .....	<b>9</b>	Предсказательная аналитика и системы поддержки принятия решений: применения для умного города.....	82	SAS.....	126	Яндекс.....	140
Предиктивная аналитика — ключевые идеи.....	10	Предсказательная аналитика и системы принятия решений в промышленности.....	91	SAP.....	127	Mail.ru Group.....	141
Предсказательная аналитика: классификация.....	16	Большие данные в финансах.....	97	IBM.....	128	Лаборатория Касперского.....	142
Этика предсказаний.....	18	Прогностика динамических процессов социума.....	100	Netflix.....	129	Сбербанк.....	143
<b>ТЕХНОЛОГИИ</b> .....	<b>23</b>	Сможет ли AI стать Индустрией?.....	100	Baidu.....	130	МТС.....	144
Предиктивная аналитика на пальцах.....	24	Предиктивная аналитика решает задачи бизнеса в области телекоммуникаций.....	107	Alibaba Group.....	131	X5 Retail Group.....	145
Предиктивная аналитика: обзор трендов и методов.....	27	<b>АКТУАЛЬНЫЕ КЕЙСЫ</b> .....	<b>111</b>	Google.....	132	Газпром Нефть.....	146
Обзор классических систем поддержки принятия решений.....	38	Спектр применения систем искусственного интеллекта для диагностики поражения легких при пандемии COVID-19 и после нее.....	112	Facebook.....	133	Цифра.....	147
Обзор систем принятия решений, построенных на правилах.....	46	Видео кейсы применения.....	120	Amazon.....	134	КРОК.....	148
Анатомия рекомендательных систем.....	55			Huawei.....	135	ЛАНИТ.....	149
Выбор архитектуры рекомендательной системы.....	65					<b>ТРЕНДЫ И АНАЛИТИКА</b> .....	<b>151</b>
						Тренды и прогнозы.....	152
						<b>ПРИЛОЖЕНИЯ</b> .....	<b>157</b>
						Основные ссылки и методика формирования отчета.....	158

## Центр Национальной технологической инициативы на базе МФТИ по направлению «Искусственный интеллект»

Данный Альманах по Искусственному интеллекту – один из результатов работы аналитической группы Центра НТИ на базе МФТИ по направлению «Искусственный интеллект» (далее – Центр)

Центр НТИ по направлению «Искусственный интеллект» создан на базе МФТИ в 2017 году в рамках реализации программы Национальная Технологическая Инициатива.

### ОСНОВНЫЕ ЦЕЛИ И ЗАДАЧИ ЦЕНТРА:

- Создание новых технологий, продуктов и услуг для рынков НТИ;
- Создание и вывод на рынки НТИ спин-офф-компаний, нацеленных на преодоление технологических барьеров и быстрый рост капитализации;
- Нарращивание человеческого капитала в области искусственного интеллекта и технологического предпринимательства;
- Развитие системы целевой подготовки научных и инженерных кадров, способных решать сложные задачи в области искусственного интеллекта;
- Развитие кооперации академических организаций с предприятиями различных отраслей реального сектора экономики.

### НАПРАВЛЕНИЯ РАБОТЫ

Деятельность Центра ведется с целью комплексного развития соответствующей сквозной технологии и осуществляется в рамках следующих групп мероприятий: выполнение научно-исследовательских и опытно-конструкторских работ (далее – НИОКР) и коммерциализация их результатов; разработка и реализация основных

образовательных программ высшего образования, программ дополнительного образования и дисциплин (модулей), направленных на формирование необходимых компетенций; развитие информационной инфраструктуры, а также инфраструктуры научной, научно-технической и инновационной деятельности; обеспечение правовой охраны, управления правами и защиты РИД, разработки необходимой нормативной базы; развития партнерских отношений с заинтересованными организациями.

В рамках Программы создания и развития Центра выделено 7 ключевых направлений, которые в основном определяют содержание технологии «Искусственный интеллект»:

1. разговорный искусственный интеллект;
2. распознавание текстов и речи, лингвистический анализ;
3. экспертные, рекомендательные, информационно-аналитические системы, автоматизация проектирования и управления;
4. специализированные процессоры и вычислительные системы для искусственного интеллекта, дизайн-центр по проектированию специализированных микросхем, микропроцессоров, микроконтроллеров, чипсетов и приборов для устройств и систем с искусственным интеллектом;
5. техническое зрение, обнаружение, распознавание, дешифрация, классификация изображений;
6. робототехника, умные машины;
7. «умные» сети и системы в энергетике, связи, городском хозяйстве и в других отраслях, «умный дом», «умный город».

### ОСНОВНЫЕ УЧАСТНИКИ КОНСОРЦИУМА И ПАРТНЕРЫ ЦЕНТРА НТИ:

#### ○ Крупные промышленные корпорации

ПАО Ростелеком, АО Глонасс, ОАО РЖД, ГК 1520, ПАО Россети, АО Концерн ВКО Алмаз-Антей, АО Росэлектроника, ПАО Газпром нефть, ГК ХимРар

#### ○ Малые и средние частные инновационные компании

Нейроботикс, BiTronics Lab, Миландр, АО Крибрум, Наносемантика, Нейросети Ашманова, АО МТЛ, АО Интеллект, ГК InfoWatch

#### ○ Финансовые организации и специализированные Фонды

Сбербанк России, Фонд Национальной технологической инициативы, Фонд перспективных исследований

#### ○ Высшие учебные заведения и исследовательские институты

НИУ ВШЭ, Сколтех, Университет Иннополис, АО НИИАС

Одной из задач Центра мы видим выпуск качественной аналитики для всего рынка. Этот альманах является одним из наших аналитических проектов.

**Центр НТИ на базе МФТИ по направлению «Искусственный интеллект» является центром экспертизы и компетенций по ИИ в России. Какие бы задачи вы не ставили перед собой – мы готовы вам помочь. Приглашаем Вас к сотрудничеству!**



### Игорь Пивоваров

Главный аналитик Центра НТИ на базе МФТИ по направлению «Искусственный интеллект», член Рабочей группы НТИ Нейронет, член Экспертного совета по инновационному развитию при Минэкономразвития РФ (группа искусственный интеллект), член Экспертного совета по научно-технологическому развитию и интеллектуальной собственности при комитете по науке и образованию Государственной Думы ФС РФ, генеральный директор ООО «АйПи Лаборатория», организатор Открытой конференции по искусственному интеллекту OpenTalks.AI.

## От редактора

Работая над этим выпуском, мы столкнулись с неожиданными трудностями. Казалось бы, темы предсказательной аналитики, рекомендательных систем и систем принятия решений совсем не новые, информации много и экспертов также много. Но оказалось, что эта область сильно более закрытая чем другие.

В предыдущих выпусках Альманаха мы описывали технологии по обработке естественного языка (№2) и компьютерного зрения (№3). Это бурно развивающиеся области, там много исследователей и публикаций и мы без труда нашли авторов статей и собрали много экспертных ответов на наши анкеты.

Но в области предсказательной аналитики практически все ведущие эксперты и компании по тем или иным причинам отказались написать нам статью. Практической информации оказалось удивительно мало. Мы разобрались в этой теме и пришли к неожиданным выводам, которые изложим в этом выпуске.

Забегая вперед, в нескольких словах: предсказательная аналитика стала настолько важной составляющей бизнеса лидирующих компаний, что все работы в этой области ведутся в закрытом режиме. Почему?

Примерно 100 тыс. лет назад *homo sapiens* научились использовать свой интеллект для прогнозирования мира и возможных последствий своих действий. Мы научились строить модель внешнего мира у себя в голове и прокручивать варианты, выбирая оптимальное действие. Похоже, что именно это дало нам колоссальное эволюционное преимущество, благодаря которому теперь *homo sapiens* — абсолютно доминирующий биологический вид. <sup>[1]</sup>

Похоже, что подобная история только что (за последние лет 20) произошла в экономике. Некоторые компании научились эффективно предсказывать будущее — движение рынка, реакцию своих покупателей и так далее. И это дало им колоссальное конкурентное преимущество перед другими. Зачем же им про него рассказывать?

В отличие от NLP и CV, предсказательная аналитика уже давно перестала быть «разработкой», которую не знают как применить и смотрят на опыт других. ПА стала 100% работающей корневой частью бизнеса, неотъемлемым элементом принятия бизнес-решений. Если Вы еще не используете предсказательную аналитику — этот выпуск для Вас!

## Введение

# 01

**Эрик Сигель**

PhD, эксперт по предикативному анализу

(отрывок из книги «Просчитать будущее. Кто кликнет, купит, соврет или умрет»)

## Предиктивная аналитика — ключевые идеи <sup>[2]</sup>

Количество данных, циркулирующих в сети каждый день, поистине сногшибательно. Каждое сообщение, которое вам нравится в Facebook, каждая покупка, которую вы совершаете в Интернете, и каждое нажатие на рекламу генерируют данные — океаны данных. Для компаний и правительств, желающих лучше понять и повлиять на ваше поведение, эти данные — золотая жила.

Используя предиктивную (прогнозную) аналитику (*далее ПА*), ветвь аналитики, которая занимается предсказанием будущих событий, компании могут использовать оставленные вами данные, чтобы предсказать ваше будущее поведение с невероятной точностью.

Однако эта растущая способность предсказывать индивидуальное поведение также поднимает важные моральные и этические вопросы. Хотим ли мы, чтобы наше будущее было предсказано?

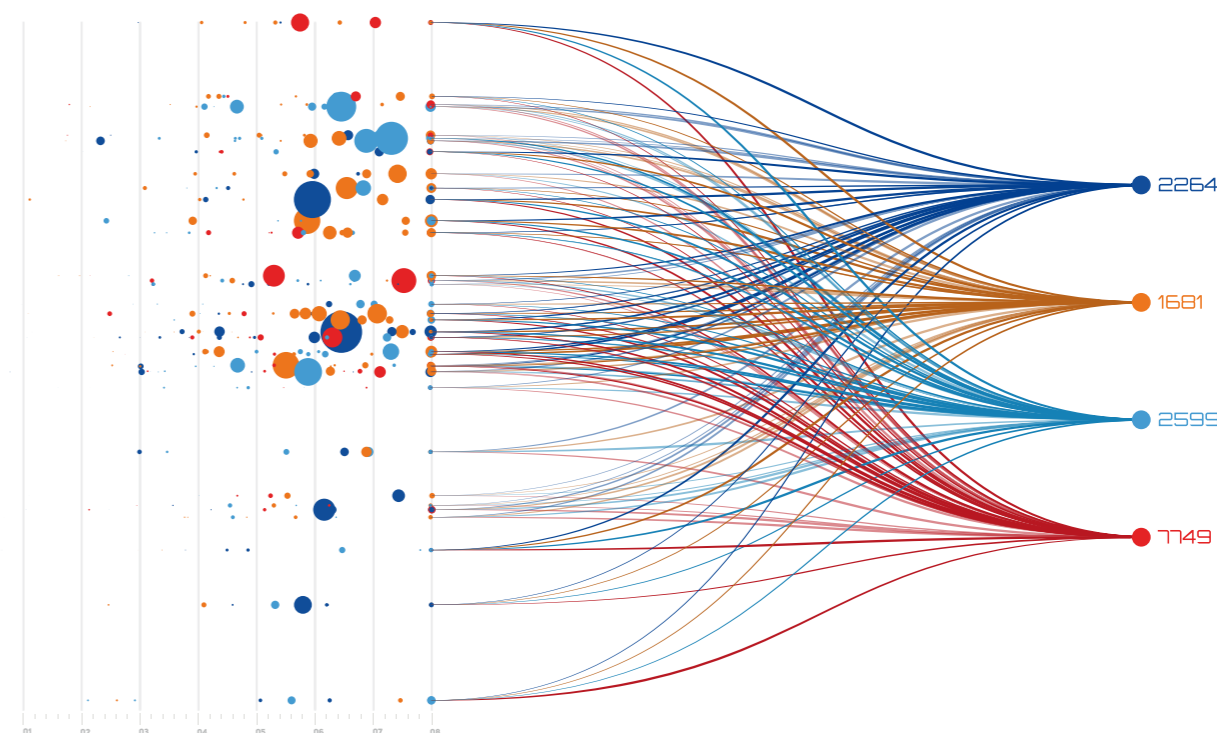
### 01. Предиктивная аналитика может помочь вам снизить риски и принять более точные решения

Каждый раз, когда компания инвестирует в дорогую рекламную кампанию, она рискует: всегда есть шанс, что кампания может провалиться, и миллионы долларов будут потрачены впустую. Однако благодаря использованию прогнозной аналитики компания может снизить этот риск.

Цель предиктивной аналитики состоит в том, чтобы изучить поведение человека и получить

представление о том, как люди будут реагировать на определенные ситуации, такие как просмотр рекламы.

ПА делает это, принимая во внимание широкий спектр статистических данных и человеческих характеристик, каждая из которых направлена на понимание индивидуального, а не общего поведения. Таким образом, вы не будете использо-



вать ПА, чтобы определить, какая реклама имеет наибольшую привлекательность; Вы будете использовать ее для определения вероятных откликов конкретных людей на конкретные рекламные объявления.

Точнее так: как только вы введете все свои переменные, вы получите прогнозный балл. Эта оценка не говорит вам о точном будущем, а говорит о том, насколько вероятны будут определенные индивидуальные реакции.

Например, вы хотите знать, каких людей в Соединенных Штатах онлайн-реклама будет наиболее эффективно призывать на-

жимать на объявление при поиске грантов и стипендий. Чем больше переменных вы указали, таких как возраст, пол и домен электронной почты, тем точнее будет прогнозный показатель.

Прогностическая модель, основанная на машинном обучении, является более динамичной, чем другие модели, поскольку она может изменяться, расти и адаптироваться в зависимости от типа данных, которые ей предоставляют. Она более точная, чем другие инструменты прогнозирования, поскольку всегда может быть перепроверена на имеющихся данных, чтобы определить, насколько точными будут результаты в будущем.

### 02. Прогнозирование поднимает вопросы ответственности, морали и предубеждений

По мере того, как наша способность использовать технологии в целях прогнозирования становится все лучше и сложнее, возникает важный вопрос: сколько данных вы готовы предоставить о своей жизни? И, что более важно, сколько данных о чужих жизнях вам нужно?

**Это больше, чем просто вопрос о вашем будущем; более насущной проблемой анализа данных является конфиденциальность.**

Когда пресса узнала, что корпорация Target использует ПА, чтобы определить, какие клиенты, скорее всего, забеременеют, многие почувствовали, что компания зашла слишком далеко. В то время как Target заявила, что просто хотела рекламировать товары для беременных определенным женщинам, этот вид маркетинга подвергает пользователей риску утечки личной информации людей своим друзьям, членам семьи и коллегам — информации, которой эти люди скорее всего не готовы делиться.

А вот применение ПА для противодействия преступности, безусловно, имеет огромный потенциал. Одна из американских ПА-компаний на основе исторических данных города Санта-Круз (штат Калифорния) продемонстрировала, что 25% краж могут быть точно предсказаны. Такая система поможет полиции выявлять различные «горячие точки», которые затем могут ежедневно патрулироваться.

Крупные города, такие как Чикаго, Мемфис и Лос-Анджелес, используют ПА для снижения уровня преступности. Они используют разнообразные данные, например, уже случившиеся преступления и совокупность условий, в которых они произошли (день недели, погода и т.д.).

### 03. По любым данным можно сделать прогноз, но для точного прогноза нужен большой объем сбалансированных данных

В наши дни данные являются ценным и важным товаром для бизнеса, и мы производим их все больше каждый день. В прогнозной аналитике философия такова: чем больше данных, тем лучше — при условии, что они хорошо сбалансированы.

Это означает, что вы должны быть избирательными в данных, которые используете, и добавлять в выборку данные каждого типа.

Например, один тип данных, который относится к нашим обычным задачам и поведению, может быть получен из таких источников, как телефонные записи, банковские операции и онлайн-покупки. Другой тип дан-

Однако некоторые считают, что предсказания поведения одного человека на основе используемых данных о действиях других людей могут быть слишком категоричными.

Например, некоторые города используют ПА для определения вероятности возвращения осужденного к преступной жизни. И многие люди считают, что ПА-модели в данном случае могут быть предвзятыми.

Представьте, что есть два преступника, которые совершили одно и то же преступление, и они оба хотят условно-досрочного освобождения, но один из них прибыл из местности с более высоким уровнем преступности. Из-за уровня преступности в этой местности считается, что этот преступник с большей вероятностью вернется к преступной жизни. Это явно предвзятый прогноз, и, поскольку в городских кварталах с лишними гражданскими правами меньшинствами, как правило, уровень преступности выше, больше всего пострадают люди из этих районов. Такой прогноз может стать очередным звеном в цепи расового неравенства.

ных, часто используемый в моделях ПА, это история социальных сетей и блогов. Ежедневно публикуется ≈ 864 000 постов и записей, которые, по сути, превращают личные мысли человека в общедоступные данные. По состоянию на 2011 год было 100 миллионов отдельных блогов только на WordPress и Tumblr.

Это много данных. Фактически, если бы вы взяли все данные, хранящиеся на компьютерах в 1986 году, и распечатали их на двусторонней бумаге, у вас было бы достаточно листов, чтобы покрыть всю территорию Земли. Но сегодня вы уже можете покрыть всю планету слоем листов толщиной в несколько книг!

Хотя это изобилие данных и делает наш анализ таким сложным, оно также приводит к большему количеству потенциальных ошибок, когда данные не сбалансированы.

По мере увеличения данных возрастает вероятность того, что случайное событие будет принято за что-то значимое. Большинство ошибок, возникающих в ПА, являются результатом слишком большого количества переменных в одной области, что приводит к ложной корреляции, но этого можно избежать, создав сбалансированные датасеты, то есть добавив большое количество разнородных данных.

## 04. Благодаря машинному обучению могут быть выявлены упущенные из виду риски, однако также существуют риски для самого машинного обучения

Как мы уже видели, интеллектуальная аналитика выигрывает за счет применения машинного обучения, и обучаемая модель со временем становится все точнее в своих предсказаниях.

Но есть еще одно очень важное преимущество машинного обучения, которое заключается в его способности распознавать замаскированные риски или «микрориски».

Замаскированные риски являются стандартной опасностью для бизнеса. Они, как правило, представляют собой крошечные упущения, которые легко пропустить или проигнорировать, пока они не превратятся в огромную проблему.

Например, когда Chase Bank начал использовать ПА для составления долгосрочных прогнозов выплат по ипотечным кредитам, он осознал, в каком количестве упускаются будущие проценты при предоставлении клиентам возможности предоплаты или досрочных платежей по кредитам. Поначалу эти платежи выглядели как незначительные потери, но, когда они суммировались с прогнозируемой прибылью, они становились болезненно большой проблемой.

Согласно одному из исследований, проведенных с использованием методов ПА, у вас меньше шансов купить неисправный автомобиль, если этот автомобиль оранжевого цвета. Это явная бессмыслица, но данные подтверждали полученный результат. Проблема заключалась в том, что было недостаточно продаж автомобилей, чтобы уравновесить данные, и когда выборка была увеличена, стало ясно, что цвет краски не имеет ничего общего с шансами покупки автомобиля.

Благодаря интеллектуальному анализу и машинному обучению компьютеры буквально программируют сами себя. Ни одна деталь не является слишком несущественной, чтобы выйти из рассмотрения. В результате, ни один микрориск не останется незамеченным, поскольку модель всегда учитывает долгосрочные последствия. Таким образом, такая организация, как Chase Bank, сможет что-то предпринять, пока не стало слишком поздно. Вот почему банки теперь используют ПА, чтобы учесть все небольшие риски, связанные с кредитами.

Но обратной стороной такого подхода является попытка учесть слишком много факторов, и чрезмерное обучение может также привести к ошибочным прогнозам.

Профессор из Беркли однажды привел отличный пример такого ошибочного прогноза, представив данные, которые подтверждают любопытное утверждение: колебания фондового рынка в США содержат такие же паттерны, как и динамика производства масла в Бангладеш.

Выход из этой подобной ситуации очень человеческий: дайте возможность машине делать ошибки, чтобы она могла учиться на них и в следующий раз распознать неправильную логическую связь.

## 05. Объединение нескольких источников данных и моделей ПА повышает точность результата

Как начинающие художники и предприниматели, ПА выигрывает от краудсорсинга, применяя идею за счет использования ансамбля методов. Ансамбль представляет собой комбинацию прогнозирующих моделей, работающих одновременно в конкуренции и сотрудничестве. И родилась эта модель в результате краудсорсинговых конкурсов.

Принято считать, что «рассвет ансамблевых моделей», произошел в 2008 году, когда Netflix запустила краудсорсинговый конкурс для повышения точности своей системы рекомендаций на 10%. На заключительных этапах соревнования объединились две большие команды из более чем 20 участников, объединив при этом две очень мощные прогностические модели, каждая из которых достигла цели Netflix.

Это стало возможным благодаря атмосфере дружеской конкуренции, которая была создана во время соревнования. В конкурсе участвовали посетители интернет-форумов, которые способствовали обмену новыми идеями и постоянному открытому диалогу.

Исследования показывают, что при переходе от одной модели к ансамблю производительность возрастает на 5–30%. Кроме того, ансамблевые модели, постоянно совершенствуются за счет интеграции большого числа моделей. Это явление стало называться ансамблевым эффектом, и оно используется для решения все более сложных проблем

## 06. Человеческий язык — сам по себе сложная задача, но на пути его моделирования уже достигнуты большие успехи

Увеличенная мощь ансамблевых моделей была использована в некоторых удивительно сложных проектах, включая способность машин обрабатывать естественный язык. Но одной из самых больших проблем в любом проекте, связанном с компьютерной лингвистикой, является выявление всех нюансов человеческой речи.

Когда два человека разговаривают друг с другом на работе, возникает множество контекстов, каждый из которых помогает определить истинное значение сказанного. Например, человек может сказать «это здорово», однако сказанное им может оказаться сарказмом, и быть полностью противоположным исходному смыслу.

Тем не менее, считается, что текстовые данные составляют 80% всех данных. Таким образом, понимание текста — это как самая большая возможность ПА, так и его самая большая проблема.

Но за последние годы в этой области достигнут огромный прогресс. Такие модели как BERT, архитектура Transformer а в последние год и новые модели GPT-2 и GPT-3 позволяют не только качественно распознавать семантику и смысл текста, но даже генерировать новые тексты на заданные темы. Есть ощущение, что в ближайшие несколько лет эта задача будет полностью решена.

## 07. ПА может выявить незаметное путем количественной оценки силы убеждения

Вы ненавидите получать спам от мобильных операторов и кредиторов? Хорошая новость заключается в том, что достижения в области ПА теперь позволяют компаниям узнать, какие люди восприимчивы к холодной рекламе, а каких лучше оставить в покое.

Ни одна компания не хочет раздражать свою целевую аудиторию, которая может из-за этого отказаться от их продуктов и услуг, поэтому необходимо быть тактичным и убедительным одновременно, и именно в этом заключается искусство ПА.

Одним из лучших примеров данного подхода является деятельность норвежской телекоммуникационной компании Telenor, сотрудники которой осознали, что, отправляя рекламу тем, кто хочет отказаться от их услуг, они зачастую попадают на клиентов, не собиравшихся прекращать использование услуг компании и понижают таким образом лояльность собственной аудитории. И, делая это, Telenor фактически провоцирует их. Это поднимает еще одну проблему, стоящую перед отраслью ПА: возможно ли одновременно предсказать, как целевой и нецелевой клиент отреагирует на одно и то же сообщение?

Эта тема породила модель uplift, которая оценивает силу убеждения. Модель uplift использует два набора данных, чтобы для конкретного клиента найти ответ на вопрос: какова будет сила убеждения для него?

Как правило, один из наборов данных — это контрольная группа, например, группа клиентов, с которыми компания вообще не контактирует. Такая модель похожа на медицинские клинические исследования, в которых есть контрольная группа, дающая эталонные результаты для измерения всего остального.

В дополнение к измерению силы убеждения, такая модель может также разделить клиентов на три группы: те, о ком вам не нужно беспокоиться; люди, которые не нуждаются в дополнительном убеждении, и те, кого никогда не убедить.

Эта модель творила чудеса для таких компаний, как Банк США, Fidelity и Telenor, и увеличила эффективность маркетинга на целых 36%.

Возможно, вы не знаете о влиянии ПА на вашу повседневную жизнь, но она затрагивает почти все ее аспекты и не только влияет на то, как технологии взаимодействуют с вами; это также движущая сила многих наших текущих технологических достижений. Если вы хотите знать, какие процессы и как происходят в современном мире, вы должны быть знакомы с прогнозной аналитикой.



# Предсказательная аналитика: классификация

**ЧТО СЛУЧИЛОСЬ?**

**ОПИСАТЕЛЬНАЯ АНАЛИТИКА**  
При помощи описательной аналитики создается сводка исторических данных для их дальнейшего анализа. Например, непрерывный сбор информации с производственного оборудования с помощью датчиков позволит точно идентифицировать момент сбоя в технологическом процессе.

**ПОЧЕМУ ЭТО СЛУЧИЛОСЬ?**

**ДИАГНОСТИЧЕСКАЯ АНАЛИТИКА**  
В диагностической аналитике используются различные методы анализа данных для выявления основных факторов влияния на результаты. Например, в промышленности диагностическая аналитика позволяет понять, по какой причине случилась авария на производстве.

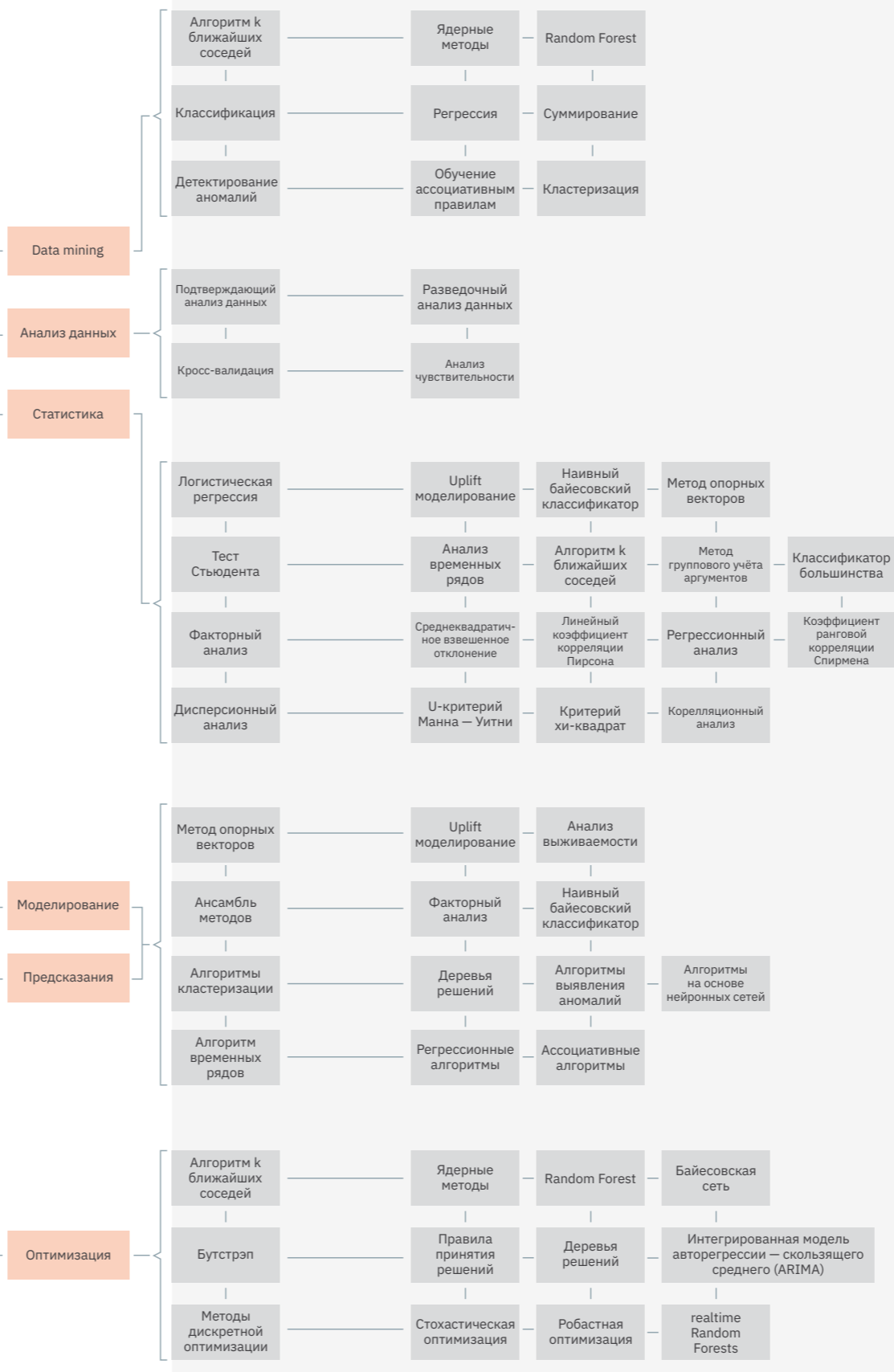
**ЧТО МОЖЕТ СЛУЧИТЬСЯ?**

**ПРЕДСКАЗАТЕЛЬНАЯ АНАЛИТИКА**  
Предсказательная аналитика позволяет на основе исторических данных строить модели, делающие предсказания. Например, определять сроки профилактического ремонта оборудования с целью избежать его поломки.

**ЧТО ДЕЛАТЬ, ЧТОБЫ ЭТО СЛУЧИЛОСЬ?**

**ПРЕДПИСЫВАЮЩАЯ АНАЛИТИКА**  
Предписывающая аналитика позволяет принимать максимально эффективные управленческие решения для минимизации или максимизации вероятности конкретного события. Например, модель предписывающей аналитики может показать, какие именно детали производственного оборудования надо заменить и в какие сроки, чтобы минимизировать вероятность поломки.

## Основные методы



- Топ ПО/инструментов/библиотек**
- Топ платных**
    - Sisense, Neutal Designer, Rapid Insight Veera, Alteryx Analytics, RapidMiner Studio, Dataiku RSS, KNIME Analytics Platform, SAS Enterprise Miner, Oracle Data Mining ODM, Altair, TIBCO Spotfire, AdvancedMiner, Microsoft SQL Server Integration Services, Analytic Solver, PolyAnalyst, Viscosity Software Suite, Salford Systems SPM, HP Vertica Advanced Analytics, TIMi Suite, Genedata Analyst, LIONoso, Teradata Warehouse Miner, pSeven, Civis Platform
  - Топ бесплатных**
    - Orange data mining, Anaconda, R Software Environment, Scikit-learn, Weka Data Mining, Shotgun, DataMelt, Natural Language Toolkit, Apache Mahout, GNU Octave, Graphlab Create, ELKI, Apache UIMA, KNIME Analytics, Platform Community, TANAGRA, Rattle GUI, CMSR Data Miner, OpenNN, Dataiku DSS Community, DataPreparator, LIBLINEAR, Chemicalize.org, Vowpal Wabbit, mlpy, Dlib, CLUTO, TraMineR, ROSETTA, Pandas, Fityk, KEEL, ADaMSoft, Sentic API, ML-FlexDataBionic, ESOM, MALLET, streamDM, MiningMart, Modular Toolkit for Data Processing, Jubatus, LIBSVM, Arcadia Data Instant
  - Топ платных**
    - Nvivo, ATLAS.ti, Provalis Research Text Analytics Software, Quirkos, MAXQDA, Dedoose, Raven's Eye, Qiqa, webQDA, HyperRESEARCH, Transana, F4analyse, Annotatons, Datagrav, Sisense
  - Топ бесплатных**
    - Orange Data Mining, Anaconda, R Software Environment, Scikit-learn, Weka Data Mining, Shotgun, Tableau Public, DataMelt, Microsoft R, Trifacta, SciPy, ELKI, KNIME, Analytics Platform Community, Scilab, TANAGRA, Dataiku DSS Community, DataPreparator, ITALASSI, HP Vertica Advanced Analytics, Google Fusion Tables, NodeXL, Fluentd, Display, NumPy, OpenRefine, JuliaMassive Online Analysis, DataWrangler, EasyReg, Matplotlib, Ipython, SymPy, FreeMat, jMatLab, PAW, ILNumerics, ROOT, NetworkX, Arcadia Data Instant, SIGVIEW, Gephi
  - Топ платных**
    - IBM SPSS Modeler, Minitab, TIBCO Spotfire, STATISTICA, Analyse-it, AcaStat, STATISTICA, SAS Visual Statistics, Forecast Pro, Regression Analysis of Time Series, MATLAB, PolyAnalyst, EasyFit, JMP, Data Applied, Mathematica, MaxStat, Data Desk, XLSTAT, Analytica, Eviews, GraphPad InStat, GeneXproTools, Develve, NCSS, NLOGIT, GAUSS, Statgraphics Centurion, Qlucore, Statistix, TurboStats, Maple, WinSPC, Origin, SmartPLS, StartsDirect, StatPlus, MedCalc, Genedata Analyst, UNISTAT, LIMDEP, statistiXL, NLREG, ESBStats, SigmaPlot, PASS, Statwing, Number Analytics, SUDAAN, SuperCross, Prisync, StstXact, Genstat, Genstat, LISREL, Sigma Stat, Wizard
  - Топ бесплатных**
    - SAS University Edition, GNU PSP, Statistical Lab, Shotgun, Develve, GNU Octave, DataMelt, Arc, SciPy, Dataplot, MacAnova, Past, Zelig, Gretl, GNU Data Language, CumFreq, OpenMx, MaxStat Lite, SageMath, Dashboard of Sustainability, Salstat, SOFA Statistics, Perl Data Language, NIMBLE, OpenEpi, ADaMSoft, BV4.1, CSPro, Dap, Epi.Info, First Bayes, Ploticus, Yorick, Statcato, ViSta, StatCVS, JAGS, Stan, pbdR, OpenStat, Simfit, MicrOsiris, NCAR Command Language, EasyReg, IVEware, WinBUGS, WINPEPI, IDAMS, ADMB, Scilab
  - Топ платных**
    - Science for Cloud Data Teams, Google AI Platform, Rapid Insight Veera, Microsoft Azure, SAP Predictive Analytics, Alteryx Analytics, DataRobot, IBM Predictive Analytics, RapidMiner Studio, KNIME Analytics Platform, SAS Predictive Analytics, H2O.ai, Oracle Data Mining ODM, GoodData, WebFOCUS Platform, Altair, TIBCO Spotfire, MATLAB, Mathematica, AdvancedMiner, Analytic Solver, ABM, Viscosity Software Suite, FICO Model Central, Salford Systems SPM, TIMi Suite, DMWay, Feature Labs, Vanguard Business Analytics Suite, CMSR Data Miner Suite
  - Топ бесплатных**
    - Anaconda, Orange Data Mining, R Software Environment, Scikit-learn, Weka Data Mining, Microsoft R, Apache Mahout, GNU Octave, GraphLab Create, SciPy, KNIME Analytics Platform Community, Apache Spark, Dataiku DSS Community, LIBLINEAR, Vowpal Wabbit, NumPy, PredictionIO, TANAGRA
  - Топ платных**
    - IBM Prescriptive Analytics, Profitect, Ayata, AIMMS, LIONoso, NGData



Игорь Пивоваров

Центр Науки и Технологий  
Искусственного Интеллекта МФТИ

# Этика предсказаний

## 1. ПРИ ЧЕМ ЗДЕСЬ ЭТИКА?

Предсказательная аналитика прочно вошла в нашу жизнь. Множество умных алгоритмов формируют нашу новостную ленту, показывают рекламу, рекомендуют к покупке товары, книги, фильмы, даже друзей. Все это базируется на постоянном сборе информации и анализе нашего поведения. Казалось бы, причем здесь этика?

Но при ближайшем рассмотрении оказывается, что бизнес многих ведущих технологических компаний построен на достаточно скользких этических основаниях, зачастую на эксплуатации очень опасных и низменных инстинктов человека.

И чем дальше мы движемся, тем больше становится фактов, вызывающих озабоченность. Скандал с Cambridge Analytica; слежка АНБ за пользователями соцсетей, включая их переписку, а также сведения о местонахождении; регулярные новости о том, как большие корпорации собирают информацию о пользователях и предоставляют ее разработчикам приложений...

*Вам все еще кажется, что этика здесь не причем?*

## 2. МИЛЛИАРДЫ КУКОЛ «ВУДУ»

Большие корпорации — такие как Google, Facebook, Яндекс и другие — собирают гигантское количество данных о каждом из нас — наши запросы в интернете, сайты которые мы посещаем, наши покупки, наши контакты, что мы «лайкаем» и смотрим и т. д. И на базе всей этой информации алгоритмами искусственного интеллекта строится фактически «аватар» челове-

ка, его цифровой двойник, наше представление на серверах компании, которое Тристан Харрис остроумно называет «куклой вуду»<sup>[3]</sup>. Этот цифровой двойник достаточно точно отражает наши пристрастия и увлечения, наши пороки и слабости. И «держит в руках» эту «куклу вуду» гигантский суперкомпьютер, который, каждый раз, когда мы открываем эту социальную сеть, решает, что нам показать в данный момент времени. При этом он принимает во внимание все факторы, включая время суток, погоду за окном и проигрыш вашей любимой футбольной команды накануне. 70% трафика Youtube сейчас смотрят благодаря рекомендациям машинных алгоритмов, а не благодаря самостоятельному выбору человека.

**Вы можете возразить: «это же просто рекомендация! В конечном итоге, я же сам принимаю решение — смотреть или не смотреть!».**

Это так. Но нужно понимать, что против вас играет огромный суперкомпьютер, располагающий данными не только о вас, но и о ваших друзьях, близких, об их интересах и реакциях.

Открывая сеть, каждый из нас оказывается один на один с мощной системой искусственного интеллекта, который 24 часа в сутки не уставая, без эмоций и лишних рассуждений, методично анализирует ваше поведение и ищет подходы к вашему сердцу, чтобы завоевать ваше внимание.

## 3. ПРЕДСКАЗАНИЕ КАК БИЗНЕС

Зачем нужны эти «куклы вуду»? С их помощью можно предсказать множество вещей о вас и вашем поведении. Например, какую рекламу вам показать. Скажем, вы искали недавно отель на море, чтобы съездить в отпуск? Пожалуйста, вот реклама туров в Египет, купальников и солнцезащитных очков. Думаете какой фильм посмотреть? Вот пожалуйста, рекомендация на основе больших данных анализа множества пользователей. Заходите в новостную ленту или Instagram? Вот подборка новостей или фотографий, соответствующих вашим интересам. На этих предсказаниях выросли самые большие компании сегодняшней мировой экономики. Google, Facebook, Amazon — все они специализируются именно на предсказании того, какой контент вам показать, чем вас зацепить. Их огромная капитализация показывает, что инвесторы верят в эту бизнес-модель, она дает реальные результаты. Казалось бы — ну и что в этом плохого? Человек же получает то, что ему интересно. И он сам выберет дальше, что с этим делать. Сделаем еще один шаг чтобы понять, как для каждого человека подбирать контент, который будет наверняка цеплять его, задевать, провоцируя на реакцию. Каким образом это сделать?

## 4. КОНФЛИКТЫ И ПРОТИВОРЕЧИЯ КАК БИЗНЕС-МОДЕЛЬ

При внимательном рассмотрении таких социальных сетей, как Facebook, Instagram, Twitter или Youtube у них видны одинаковые черты. Все эти сервисы бесплатны для пользователя и существуют за счет рекламы. А рекламы можно показать тем больше, чем больше пользователь проводит времени в сети. А пользователь проводит тем больше времени, чем больше интересного контента он видит. То есть бизнес-модель соцсетей построена на максимальном удержании («engagement») пользователя. Первой понятной стратегией подбора конечно является показывать контент схожий с тем, что пользователю нравится. Если пользователь «лайкнул» пост, посмотрел фильм или купил книгу — значит найдена тема, которая ему нравится и ее можно развивать дальше. Но практика показывает, что для многих пользователей эта простая стратегия быстро перестает работать — человеку надоедает смотреть и читать одно и то же. Но есть другой действенный способ понять, что является самым интересным контентом и вызывает самую живую реакцию пользователя и заставляет его оставать-

ся «plugged in». Замечательно об этом рассказал недавно Tristan Harris на слушаниях в сенате США<sup>[4]</sup>. Захват и удержание внимание пользователя осуществляется в два приема:

- 1. Первое** — вместо попыток удержать внимание пользователя нужно позволить и помочь ему самому привлечь внимание к себе. Это сильнейший социальный наркотик — каждый из нас стремится привлечь к себе внимание. И механизмы всевозможных «лайков» и «ретвитов» прямо показывают человеку уровень внимания к нему и заставляют снова и снова «постить в сеть», чтобы не терять уровень внимания.
- 2. И второе.** Что будет привлекать максимальное внимание? Что человек будет репостить сам? И что он будет комментировать в других постах? Оказывается, что максимальную кликабельность и ретвиты вызывает противоречивая информация, содержащая в себе скандал или конфликт, полярные и даже провокационные высказывания и позиции. Именно это наиболее эффективно привлекает внимание окружающих и приносит лайки и репосты.

**Для удержания внимания, надо показывать вам то, что будет вас провоцировать, вызывать острую реакцию, желание ответить или перепостить. А реакция других людей на ваши посты сработает как сильнейший стимулятор.**

## 5. HUMAN IS A HACKABLE ANIMAL

Это выражение ввел в обиход известный историк, профессор Юваль Ной Харари. Под этим он понимает то, что человек оказался уязвим перед этими новыми технологиями. Когда человека анализирует суперкомпьютер на основе больших данных, в том числе биометрических, то можно не только предсказывать его поведение, но и воздействовать на него. А еще лучше, если это новое поведение получает «подкрепление», как это называется в биологии. В данном случае — те самые «лайки» и репосты, действующие на человека как наркотик. Результат понятен. 99% людей оказываются неспособными противостоять этому воздействию. В 1954 году в мозге был открыт так называемый «центр удовольствия», вырабатывающий нейромедиатор дофамин, вызывающий у человека чувство удовольствия (или удовлетворения). Дофамин естественным образом вы-

рабатывается в больших количествах во время положительного, по субъективному представлению человека, опыта — к примеру: секса, приёма вкусной пищи, приятных телесных ощущений. В опытах на мышах, которые проводили Джеймс Олдс и Питер Милнер, им в мозг вживляли электроды, которые, когда мышь нажимала на педаль, стимулировали «центр удовольствия», вызывая выработку дофамина. И мышам это настолько нравится, что они постепенно переставали есть и пить и только нажимали на педаль до 700 раз в час...

Мы становимся похожими на этих подопытных мышей, которые нажимают на педаль, получая дофамин. Люди реально подсаживаются на ленту соцсетей, продолжая прокручивать ленту, каждый раз получая новую порцию информации.

*Но и это еще не всё.*

## 6. МИЛЛИАРДЫ ЛИЧНЫХ МИРОВ

Противоречия и конфликты, похоже, встроены в бизнес-модель социальных сетей. Но еще сильнее на общество влияет тот факт, что каждый человек получает свою собственную «информационную ленту». Для каждого из нас суперкомпьютеры больших корпораций формируют собственную «реальность». Мне могут сегодня повсюду показывать новости об экологической катастрофе в Норильске, а вы весь день видите сообщения о коронавирусе. А какой-нибудь сторонник Навального будет везде видеть сообщения о политических протестах. И каждому из нас будет казаться, что все остальные видят то же самое, а на самом деле каждый видит свою картинку... Получается, что сервисы с девизами типа «объединяем людей» на деле их разъединяют на множество маленьких групп. Каждый из нас видит лишь тот «информационный сектор», который выбирается для нас алгоритмами и не видит других. Каждый живет в собственном мире, собственном информационном пузыре, собственной матрице.

## 7. УПРАВЛЕНИЕ ЧЕРЕЗ ПРЕДСКАЗАНИЕ

Но самое неприятное, что в таком мире каждым конкретным человеком становится проще управлять. Дозируя информацию, которую он получает, управляя ее последовательностью и источниками, мнениями людей оказывается достаточно просто можно манипулировать. И не только мнением по поводу конкретного товара или знаменитости, но и управлять политическим или даже жизненным выбором.

Последний самый яркий пример такой манипуляции — компания Cambridge Analytica, которая, используя данные Facebook, манипулировала его пользователями с целью воздействия на выборы президента США. А сколько таких историй нам неизвестно?

*Я не утверждаю, что каждым из нас целенаправленно манипулируют. Но вы готовы утверждать обратное?*

## 8. ВЕЛИКИ ЛИ РИСКИ

Механизмы формирования наших «информационных лент» надежно скрыты внутри больших корпораций. Можем ли мы быть уверены в их непредвзятости и отсутствии ошибок? Кто пишет эти алгоритмы?

Вы скажете, что волноваться не нужно, их пишут замечательные умные люди. У которых как правило нет цели кем то манипулировать. Ими движут стремление решить интересную задачку и в конечном итоге заработать денег. Я соглашусь, но лишь отчасти. Потому что на выходе не всегда может получиться то, что ожидается.

Классическим примером в этой области стал случай с алгоритмом компании Northpointe, который предсказывал вероятность повторного преступления подсудимым в суде в округе Броуард (Флорида), помогая судье определить, наказать преступника по минимальной или максимальной строгости, предусмотренной законом. Программа на основе имеющихся данных вычисляет так называемый «показатель оценки риска» (risk assessment score). Этот показатель определяет вероятность повторения преступления и, соответственно, влияет на тяжесть приговора. Так вот, тщательный анализ результатов работы программы, проведенный журналистами ProPublica<sup>[5]</sup>, выявил странную тенденцию: по какой-то причине для афроамериканцев (негров) программа выдавала более высокую оценку риска, чем для белых. Другими словами, алгоритм проявлял расовую дискриминацию!

Как выяснилось, только 20% подозреваемых, по которым программа определила высокий риск совершения преступлений, действительно совершили его в течение двух лет. Но ошибочное предсказание рецидива для чернокожих было вдвое выше, чем для белых!

Это резонансный случай, подробно разобранный, потому что дело казалось судьбы многих людей. А сколько таких случаев неизвестно?

**Чем больше алгоритмы входят в нашу жизнь, чем больше случаев, в которых им отдается решение (или рекомендация решения) — тем больше будет возникать «серых зон» — неоднозначных ситуаций, в которых неочевидно почему алгоритм выдал такую рекомендацию или решение. И тем больше будет возникать неэтичных ситуаций.**

## 9. ЧТО ДЕЛАТЬ?

Очевидно, что нужно какое то регулирование таких алгоритмов. Не очень верится в саморегулирование больших корпораций — бизнес есть бизнес и для коммерческой компании главная задача — увеличение выручки и прибыли. Также не очень верится и в регулирование со стороны государства.

Проблема в том, что государство (а конкретно отдельные государственные агентства, например спецслужбы) являются прямыми выгодоприобретателями этих больших данных и информации о каждом человеке и его взаимоотношениях с миром. Государству не очень выгодно сильно регулировать эту область.

## 10. ЭТИЧЕСКИЙ КОМИТЕТ

Возможным подходом к решению может быть введение «этических комитетов» в каждой большой корпорации, собирающей и обрабатывающей персональную информацию. Такие этические комитеты, например, существуют во всех медицинских учреждениях и через них в обязательном порядке проходят сложные решения по клиническим исследованиям, когда определяется какие данные и как нужно собирать и взвешиваются возможные результаты с одной стороны и необходимые для этого манипуляции (в том числе инвазивные) с пациентами с другой стороны.

Такой этический комитет должен формироваться широким общественным голосованием из людей, являющихся безусловными моральными авторитетами. Люди склонны не верить правительству или большим корпорациям, но верят известным ученым, писателям, врачам. Не обязательно, чтобы эти люди были специалистами в области машинного обучения или больших данных, потому

что обсуждаемые здесь вопросы имеют понятное общечеловеческое значение.

Такой этический комитет может существовать в формате отдельной некоммерческой организации, в которую сама корпорация может внести пожертвования, но не сможет влиять на непосредственные выплаты членам этого комитета, чтобы исключить конфликт интересов.

Через такой этический комитет должны проходить все изменения алгоритмов поиска, рекомендации и выдачи информации. Он должен акцептовать новые алгоритмы и необходимость сбора тех или иных данных.

Некоторые компании скажут, что у них и так настроены процедуры обсуждения и утверждения алгоритмов и сбора данных. Это так, но простой гражданин не может быть уверен в объективности сотрудников компании, которые зависят от нее. А в этическом комитете должны быть независимые авторитетные люди.

Компании не могут раскрывать публично данные и алгоритмы, в силу конфиденциальности и конкуренции. Вместо этого они обязаны будут раскрывать любую информацию этическому комитету, который будет в некотором роде выполнять роль священника, хранящего тайну исповеди и помогающего компании исправиться и стать лучше.

## 11. ПОСЛЕСЛОВИЕ

Эта статья написана не для того, чтобы напугать или запретить. И то и другое не работает. Хотелось бы найти компромисс, чтобы мы смогли больше доверять компаниям, хранящим наши данные и создающим алгоритмы, которые оказывают колоссальное влияние на нашу жизнь.

Врачи, которые лечат нас и знают про нас все, дают клятву Гиппократу, что они не будут использовать информацию и свои знания во вред пациенту. И тем не менее, в сложных ситуациях, врачи прибегают к консилиуму, чтобы быть уверенными в правильности и этичности принимаемого решения.

Как же получилось, что огромные корпорации, собравшие про нас гигантское количество информации и влияющие каждый день на наши покупки, действия и психологическое состояние, оказались без этического контроля?



# Технологии

02

Предсказательная аналитика и системы принятия решений — достаточно старая область, в ней есть как давно используемые технологии (например, экспертные системы, базирующиеся на правилах) так и новые (например, машинное обучение на нейросетях). Мы решили дать слово всем, чтобы показать весь спектр работающих сегодня решений.



Алексей Соколов

CIO Redmadrobot Data Lab

# Предиктивная аналитика на пальцах<sup>[6]</sup>

Приходит к дата-аналитику, скажем, владелец большой розничной сети с вполне конкретным запросом: хочу знать, где открыть новую точку и сколько выручки я с нее получу. Реально ли это? Вполне.

Сначала нужно посмотреть, какие данные уже есть у заказчика. Их еще называют внутренними данными.

## ВНУТРЕННИЕ ДАННЫЕ

У магазина обычно уже есть какие-то данные по существующим точкам: ассортимент, товарооборот, площадь торгового зала и так далее. Используя только эти данные, мы можем обучить модель и попытаться предсказать, например, выручку для каждой точки: для этого мы делим существующие данные в пропорции 70/30, обучаем модель на 70% данных, а на оставшихся 30% проверяем, насколько точно наша модель научилась предсказывать выручку для точки.

Проблема в том, что точность такой модели может быть невысокой: ей просто не хватает данных для обучения. Другими словами, если у нас есть только внутренние данные от магазинов, этого может быть недостаточно, чтобы с приличной точностью предсказать, сколько магазин будет выручать за месяц.

Что делать в этом случае? Обогащать данные, то есть дополнять то, что уже есть у клиента, внешними данными.

## ВНЕШНИЕ ДАННЫЕ

Внешних данных бывает огромное множество.

Погода, курсы валют, график запуска ракет SpaceX — все это внешние данные по отношению к нашему клиенту.

Понятно, что не все внешние данные нам нужны, и не все из них мы можем достать. На этом этапе к нам подключается аналитик: он хорошо разбирается в типах и источниках внешних данных, и может дать экспертную оценку, какие из них будут релевантны. Перед разработкой модели проводится исследование, которое помогает понять, какие данные нам будут полезны, а какие нет.

В случае с магазином нам могут быть полезны, например, такие данные, как проходимость конкретной точки, какие конкуренты стоят рядом, сколько денег выручают торговые точки в этом районе.

На основе этих гипотез мы можем подтянуть внешние данные и обучить модель, уже используя их. Предсказательная сила в этом случае обычно улучшается. Мы можем обучать модель несколько раз, добавляя и убирая какие-то наборы данных, добиваясь все большей точности.

## КАК ПОЛУЧИТЬ ВНЕШНИЕ ДАННЫЕ?

Некоторые сервисы-агрегаторы данных отдают их свободно, иногда даже в удобном формате xml или json — как, например, сервис OpenStreetMap, где можно получить географические данные об объекте. Бывают публичные базы данных, например от Google — это уже собранные большие наборы данных по различным тематикам, которые можно найти в открытом доступе и свободно использовать для обучения своей модели.

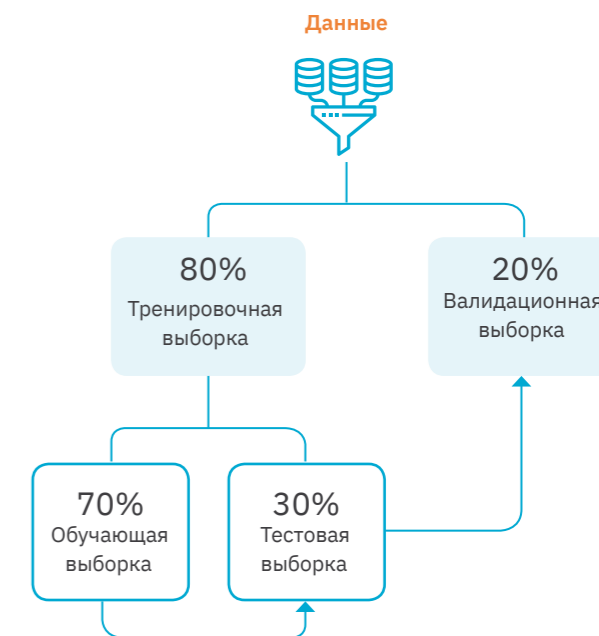
Некоторые данные находятся в открытом доступе, но их неудобно использовать. Тогда приходится парсить сайты, то есть вытаскивать данные в автоматическом режиме (до тех пор, пока это законно, конечно — но в большинстве случаев это законно).

А некоторые данные приходится покупать или договариваться об их использовании — например, если работать с операторами фискальных данных, которые могут разрешить использовать некоторую информацию о чеках.

В каждом случае мы решаем, насколько нам нужны эти данные, насколько они повысят точность модели и насколько это важно для заказчика. Предположим, какой-то набор данных позволит нам сделать модель на 10% точнее. Насколько это хорошо для заказчика? Сколько денег он сэкономит или получит, если предсказания нашей модели будут на 10% точнее? Стоит ли это того, чтобы покупать этот набор данных? Чтобы понимать это, нам нужно действительно много знать про клиента — поэтому на этапе понимания задачи мы задаем много вопросов про его бизнес, источники прибыли и особенности работы.

## КАК ПРОВЕРИТЬ ТОЧНОСТЬ МОДЕЛИ?

Как проверить (и доказать клиенту), что наша модель действительно имеет смысл? Что она предсказывает результат с нужной нам вероятностью?



Делим все данные, которые у нас есть, случайным образом в пропорции 80/20. С 80% мы будем работать и обучать на них модель, это наша тренировочная выборка. 20% пока отложим — они нам понадобятся позже, чтобы проверить на них модель и убедиться, что все работает. Это валидационная выборка.

Тренировочную выборку делим на обучающую и тестовую выборки (70/30). На 70% обучаем модель. На 30% проверяем. Когда точность нас устраивает — проверяем модель теперь уже окончательно, на валидационной выборке, то есть на тех данных, которые модель никогда не видела. Это позволяет нам убедиться, что модель действительно предсказывает с заданной точностью.

**Как правило, точность модели на тестовой и валидационной выборке почти совпадает. Если они сильно отличаются — скорее всего, дело в данных: возможно, они были поделены на обучающую и валидационную выборки не случайным образом, либо данные неоднородны.**

## MVP И ПРОМЫШЛЕННОЕ РЕШЕНИЕ

Когда мы обсуждаем с клиентом задачу, мы среди прочего определяем с ним критерии успешности проекта. Как понять, что мы выполнили задачу? Какая точность должна быть у получившейся модели и почему именно такая?

Проект мы всегда начинаем с MVP — это относительно дешевая проверка наших гипотез, это модель, которая уже может приносить ценность. Пробуем обучать модель на имеющихся данных и находим некий baseline — минимальную точность модели (например, 75%). Эту точность мы будем все время стараться повышать — до тех пор, пока это рентабельно и имеет смысл.

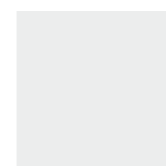
Когда точность модели нас наконец устраивает, мы упаковываем получившуюся модель в веб-сервис или мобильное приложение с удобным интерфейсом. В нашем примере с открытием магазина и прогнозированием его выручки веб-сервис мог бы выглядеть как интерактивная карта, где разные районы подсвечивались бы разными цветами в зависимости от перспективности открытия магазина здесь, а для каждой выбранной точки отрисовывалась бы плашка с прогнозом выручки магазина, поставленного в этой точке.

Отличие MVP от промышленного решения в том, что модель MVP не может дообучаться. А точность любой модели со временем деградирует, и ее надо дообучать. Поэтому для промышленного решения мы реализуем один из двух вариантов поддержки: либо мы поддерживаем ее самостоятельно, постоянно дообучая модель (и увеличивая ее точность), либо реализуем цикл переобучения модели непосредственно внутри самого софта.

Поддержка со стороны живой команды, конечно, дороже. Но минус автоматического переобучения в том, что оно не может учесть внезапных изменений характера данных. Оно не учтет, например, что в результате каких-нибудь санкций магазин перестал продавать определенные типы товаров и его выручка снизилась. Тогда точность модели сильно упадет, и ее надо будет переобучать вручную, добавляя недостающие данные.

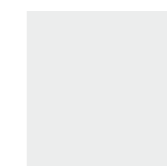
#### ПРЕДИКТИВНАЯ АНАЛИТИКА: ЧТО ПОЛУЧАЕМ НА ВЫХОДЕ

1. Веб-сервис или мобильное приложение с удобным интерфейсом, которое наглядно показывает клиенту ответ на его вопрос (например, где открывать магазин и сколько у него будет выручки).
2. Под капотом — модель, которая с заданной (и оговоренной) точностью выдает предсказания на основе имеющихся данных — внутренних данных клиента и внешних данных, которые мы приняли решение собирать и использовать в этой модели.
3. Поддержку модели, реализованную либо как постоянные доработки модели со стороны живой DS-команды, либо как встроенная функция периодического переобучения внутри самой программы. Модель поддержки выбирается в зависимости от характера данных и бизнес-задач, которую решает модель.
4. Наглядное подтверждение тому, что Data Science и Machine Learning — не просто модные технологии, а инструменты, которые помогают быстро и точно решать реальные задачи бизнеса.



**Vaibhav Kumar**

DIT University



**Madan L. Gang**

DIT University

## Предиктивная аналитика: обзор трендов и методов <sup>[7]</sup>

Предиктивная аналитика — это термин, используемый в основном в статистических и аналитических методах. Этот термин взят из статистики, машинного обучения, баз данных и методов оптимизации. Будущие события и поведение переменных могут быть предсказаны с помощью модели прогнозной аналитики путем анализа текущих и исторических данных.

Результатом модели предиктивной аналитики является оценка вероятности (или балл). Высокий балл указывает на высокую вероятность возникновения события, низкий балл указывает на низкую вероятность. Такие модели полезны при определении риска и возможностей для каждого отдельного процесса или технологии, клиента или сотрудника организации. С увеличением внимания к системам поддержки принятия решений, аналитические модели доминируют в этой области. В этой статье, мы представим обзор трендов, методов и приложения прогностической аналитики.

#### КЛЮЧЕВЫЕ СЛОВА

Прогнозная аналитика, предиктивная аналитика, статистика, машинное обучение

#### 01. ВВЕДЕНИЕ

Предиктивная аналитика используется при прогнозировании будущих событий. ПА анализирует текущие и исторические данные для того, чтобы делать прогнозы о будущем, используя методы из статистики, интеллектуальный анализ данных, машинное обучение и искусственный интеллект <sup>[8]</sup>. ПА объединяет вместе математику, информационные технологии и бизнес процессы производства и управления, чтобы сделать предсказание о будущем. Бизнес может эффективно использовать большие данные для увеличения прибыли путем успешного применения предиктивной аналитики. Это может помочь организациям стать про-активными, прогнозировать тренды или поведение на основе данные. Возможности предиктивной аналитики значительно выросли вместе с развитием больших данных <sup>[9]</sup>.

Предположим, интернет-компания XYZ Inc. ведет свой розничный бизнес по всему миру и продает различные продукты. Миллионы клиентов посещают веб-сайт XYZ, чтобы найти интересующий их продукт. Они ищут характеристики, цены, предложения, связанные с этим продуктом на сайте XYZ. Есть много продуктов, продажи которых зависят от сезона. Например, спрос на кондиционеры увеличивается летом, а спрос на обогреватели увеличивается зимой. Компания XYZ собирает все данные таких поисков:

- в каком сезоне клиенты заинтересованы в том или ином продукте;
- какой ценовой диапазон интересует конкретного клиента;
- как клиенты реагируют, увидев предложения по продукту ;
- какие другие продукты покупают клиенты в сочетании с этим продуктом;

На основе собранных данных компания XYZ может проанализировать и определить запросы человека. Она выяснит, чем привлекла конкретного клиента, какая рекомендация его заинтересовала, а затем обратиться к нему по электронной почте. Компания сообщит клиенту, что на ее сайте для него есть специальное предложение. Если клиент снова заходит на сайт, чтобы купить этот продукт, тогда компания предложит другие продукты, которые другие клиенты покупают в сочетании с этим продуктом. Если клиент начинает покупать часто, компания может уменьшить для него предложение или увеличить цену. Это простой кейс применения предиктивной аналитики.

Предиктивная аналитика имеет широкий спектр применения во многих областях. Страховые компании собирают данные работающих специалистов от третьей стороны и определяют, какой тип работы и страховой план интересует его и, таким образом, они приближаются к клиенту, чтобы привлечь его своим предложением <sup>[10]</sup>.

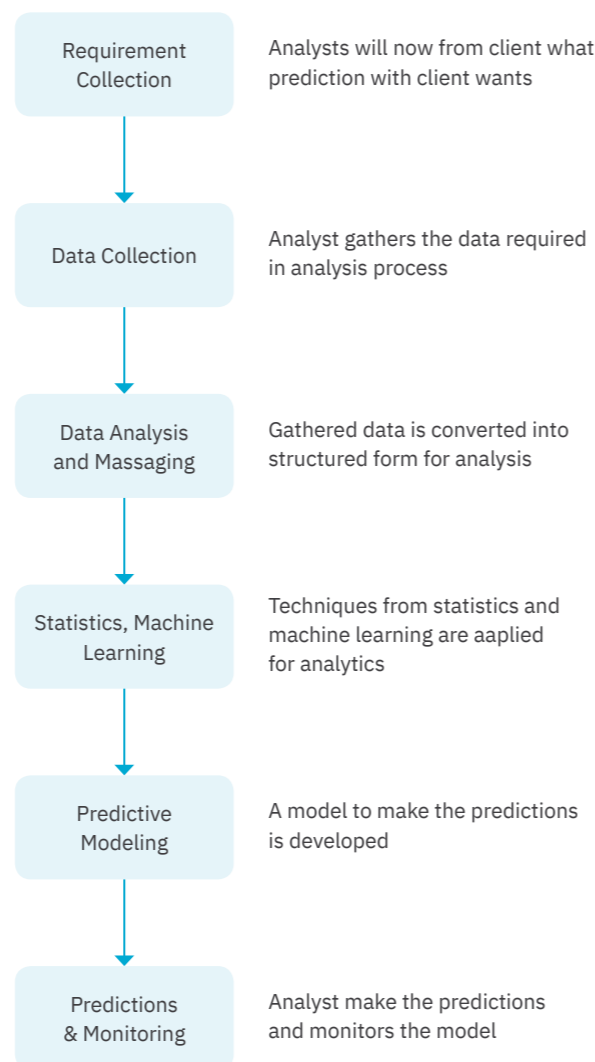
Банки применяют предиктивную аналитику, чтобы выявить риски кредитных карт и мошенничества клиентов, и быть оповещенными о таких клиентах. Организации, осуществляющие финансовые инвестиции определяют ресурсы, которые могут дать хорошую прибыль от ин-

вестиций, и они даже предсказывают будущие результаты акций на основе прошлых и текущих данных. Многие компании применяют прогнозные модели в прогнозировании продаж своей продукции, если они планируют инвестировать в производство <sup>[4]</sup>.

## 02. ПРОЦЕСС ПРОГНОЗИРОВАНИЯ

Прогнозная аналитика включает в себя несколько этапов, через которые аналитик может предсказать будущее на основе текущих и исторических данных. Этот процесс представлен на рисунке 1, приведенном ниже. (Рисунок переписать)

Рис. 1. Процесс прогнозирования



### 2.1. Сбор требований

Чтобы разработать прогностическую модель, необходимо прояснить, что является целью прогноза. Например, фармацевтическая компания хочет знать прогноз продаж лекарства в конкретной области, чтобы избежать истечения срока хранения этих лекарств. Аналитик обсуждает с заказчиком прогноза требования к прогностической модели и как заказчик будет использовать эти прогнозы. Исходя из этого становится понятно, какие данные от заказчика потребуются при разработке модели.

### 2.2. Сбор данных

Зная требования заказчика, аналитик собирает наборы данных, (которые могут быть в разных форматах из разных источников), необходимые при разработке прогностической модели. Это может быть полный список клиентов компании. Эти данные могут быть в структурированном формате или в неструктурированной форме. Желательно проверить данные заказчика с его же специалистами в связке с другими данными, чтобы избежать ошибок.

### 2.3. Подготовка и анализ данных

Аналитики данных анализируют собранные данные и готовят их для использования в модели. На этом этапе неструктурированные данные преобразуются в структурированную форму. Когда все данные становятся доступны в структурированном виде, проверяется их качество. Есть вероятность, что ошибочные данные присутствуют в основном наборе данных или есть много пропущенных значений определенных атрибутов, все ошибки должны быть исправлены.

Эффективность прогностической модели полностью зависит от качества данных. Фаза анализа иногда упоминается как обработка данных или даже «массаж данных», что означает преобразование необработанных данных в формат, который используется для аналитики.

В процессе подготовки данных, из них как правило выделяются две части:

1. Обучающая выборка (training set) — на этом наборе данных будет проводиться настройка и обучение модели
2. Тестовая выборка (testing set) — этот набор данных будет использоваться для проверки качества предсказания модели.

### 2.4. Статистика, машинное обучение

В процессе прогнозной аналитики используется много статистических методов и методов машинного обучения. Теория вероятностей и регрессионный анализ являются наиболее важными методами, которые популярны в предиктивной аналитике. Кроме того, искусственные нейронные сети, дерево решений, метод опорных векторов являются инструментами машинного обучения, которые широко используются во многих прогностических аналитических задачах. Все модели прогнозной аналитики основаны на статистических и / или машинных методах обучения. Машинное обучение часто имеет преимущество перед традиционными статистическими методами, но методы статистики как правило всегда вовлекаются в разработку любой прогнозной модели.

### 2.5. Прогнозная модель

На этом этапе модель разрабатывается на основе статистических методов и машинного обучения и обучающей выборки данных. После разработки модели, она тестируется для проверки правильности модели на тестовом наборе данных и, в случае успеха, модель считается пригодной. После внедрения модель может делать точные прогнозы на основе новых данных. Для многих применений для более точного применения применяется ансамбль моделей (см. 5.5).

### 2.6. Прогнозирование и мониторинг

После успешных испытаний модель разворачивается у заказчика для ежедневных прогнозов и процесса принятия решений. Результаты и отчеты генерируются моделью для управленческого процесса. Модель регулярно контролируют, чтобы убедиться, что она дает правильные результаты и делает точные прогнозы.

Таким образом, прогностическая аналитика — это пошаговый процесс, который включает в себя несколько этапов от формулирования требований к модели до ее разработки и контроля для ее эффективного использования в процессе принятия решений.

### 03. ВОЗМОЖНОСТИ ПРЕДСКАЗАТЕЛЬНОЙ АНАЛИТИКИ

Хотя у предиктивной аналитики давняя история и она широко применяется во многих областях десятилетиями, сегодня наступила новая эра ПА из-за повышения вычислительных мощностей и наличия больших данных [12].

Многие компании стремятся внедрить прогнозную аналитику, чтобы увеличить выручку и прибыль. Есть несколько причин для этого:

- Конкурентная среда и экономические условия подталкивают организации к использованию ПА.
- Рост объема и типов данных дает возможность использовать ПА, чтобы найти новые идеи из анализа больших данных.
- Компьютеры стали быстрее, дешевле и удобнее в использовании
- Доступно разнообразное программное обеспечение, простое в использовании.

Некоторые из наиболее распространенных применений ПА перечислены здесь:

- 1. Обнаружение мошенничества:** обнаружение и предотвращение преступного поведения могут быть улучшены благодаря объединению нескольких методов анализа. Рост кибер-мошенничества становится проблемой. Поведенческая аналитика может применяться для мониторинга действий в сети в режиме реального времени. Она может идентифицировать аномальные действия, которые могут привести к мошенничеству. Новые угрозы также могут быть обнаружены путем применения этой концепции [13].
- 2. Снижение риска:** прогнозная аналитика применяется при оценки вероятности дефолта покупателя или потребителя, которым присваивается кредитный рейтинг. Кредитный рейтинг генерируется предиктивной моделью, используя все данные, связанные с кредитоспособностью человека/компании. Предиктивные модели применяются банками при выпуске кредитных карт и страховыми компаниями для выявления мошеннических действий клиентами [14].

- 3. Оптимизация маркетинговой кампании:** применение прогнозной аналитики может выявить потенциальных клиентов и побудить их купить продукт. ПА активно используется для продвижения перекрестных продаж. Она помогает компаниям привлекать и удерживать выгодных клиентов [15].
- 4. Улучшение работы:** прогнозирование запасов и управление ресурсами может быть достигнуто путем применения ПА. Авиакомпании могут использовать ПА чтобы установить цены на билеты. Отели могут использовать прогнозистические модели для прогнозирования количества гостей на определенную ночь, чтобы максимизировать занятость и увеличить доход. Организации могут применять предиктивную аналитику для более эффективного функционирования [16].
- 5. Система поддержки принятия клинических решений:** экспертные системы на основе прогнозных моделей могут быть использованы для установления диагноза у пациента. Также, ПА может использоваться при разработке лекарств [17].

### 04. КАТЕГОРИИ МОДЕЛЕЙ ПРЕДИКТИВНОЙ АНАЛИТИКИ

Основной смысл термина предиктивная аналитика — предиктивное моделирование, то есть оценка имеющихся данных с использованием предиктивной модели, а затем выдача прогноза. Но в целом термин используется для обозначения ряда дисциплин, связанных с аналитикой. Эти дисциплины включают процесс анализа данных и используются в принятии решений. Их можно классифицировать следующим образом:

- **Прогностические модели:** связь между результатами работы объекта и его характеристиками моделируется прогностическими моделями. Такая модель оценивает вероятность того, что аналогичный объект с такими же характеристиками покажет такой же результат. Например, такая модель широко применяется в маркетинге, где ожидаются ответы о предпочтениях клиентов. Она моделирует поведение человека, чтобы дать ответы конкретный вопрос. Такая модель обычно считается прямо во время проведения транзакции с конкретным клиентом, рассчитывая его реакцию или, например, риск мошенничества.

- **Описательные модели:** Описательная модель выявляет связи между данными для нахождения потенциальных клиентов или групп клиентов. Как и прогностические модели, которые идентифицируют одного клиента, описательные модели определяют множественные отношения между продуктом и клиентами. Вместо ранжирования клиентов по их действиям, она классифицирует клиентов по их предпочтениям. Большое количество отдельных клиентов могут быть объединены в группу, чтобы сделать для нее прогноз в описательной модели.
- **Модели принятия решений:** описывают связь между данными, результатом прогноза и решением. Если на основе результатов прогноза необходимо принять решение, которое включает в себя множество параметров, для этого строят модель принятия решения. Такие модели используются, чтобы максимизировать определенный результат или минимизировать конкретные риски. Эти модели встраиваются в бизнес-правила, которые выдают решения для каждой ситуации при любых обстоятельствах.

Прогнозирующая модель строится для предсказания на необходимом уровне детализации. Такая модель генерирует прогностический балл для каждого конкретного объекта (человека, компании, транзакции). Это технология, которая учится на прошлом опыте, чтобы сделать предсказание о будущем поведении. Точность результатов модели зависит от качества данных.

### 05. СПОСОБЫ ПРЕДИКТИВНОГО АНАЛИЗА

Все модели прогнозной аналитики сгруппированы на классификационные модели и регрессионные модели. Классификационные модели предсказывают принадлежность значений к определенному классу, в то время как регрессионные модели предсказывают число. Мы перечислим здесь наиболее популярные методы, которые используются при разработке прогностических моделей.

#### 5.1. Дерево решений

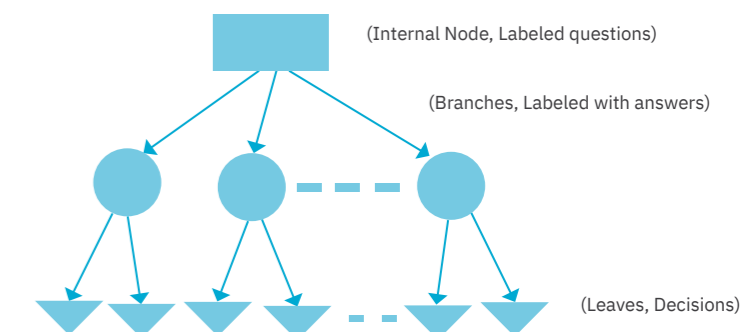
Дерево решений — это модель классификации, но его можно использовать и в регрессии. Это древовидная модель, которая связывает решения и их возможные последствия, например ис-

ход события или результирующая стоимость ресурса [18]. В этой древовидной структуре каждая ветвь представляет собой выбор между рядом альтернатив и каждый лист представляет решение. На основе категорий вводных переменных, оно разбивает данные на подмножества. Это помогает в анализе решений. Простота понимания и интерпретации делают деревья решений популярными в применении. Типичная модель дерева решений представлена на рисунке 2 ниже.

Дерево решений представлено на рисунке 2 в виде структуры. Внутренние узлы это вопросы, связанные с решениями. Все ветви являются возможными ответами на вопрос. Внешние узлы дерева, называемые листьями, это решение проблемы. Эта модель имеет свойство обработки недостающих данных, и это также полезно при выборе предварительных переменных. Деревья часто называются генеративными моделями правил индукции, которые работают на эмпирических данных. Они используют большую часть данных из датасета и минимизируют количество вопросов.

Деревья решений имеют свои преимущества и недостатки. Новые возможные сценарии могут быть легко добавлены в модель, что отражает гибкость и адаптивность модели. Дерево решений легко может быть интегрировано с другими моделями. Но у них есть ограничения на изменения. Небольшое изменение в данных может привести к большим изменениям в структуре дерева. Деревья уступают в точности прогнозирования по сравнению с другими прогнозными моделями. Эта модель сложна в расчете, особенно при использовании нечетко определенных данных.

Рис. 2

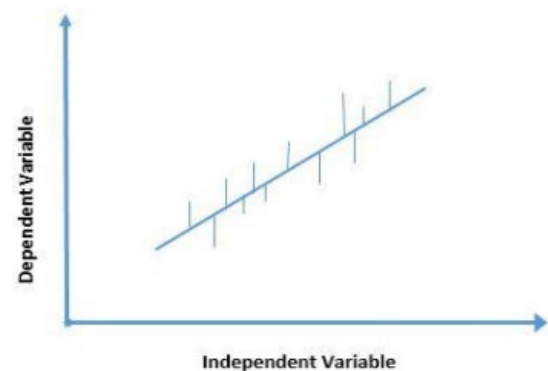




## 5.2. Модель регрессии

Регрессия является одним из самых популярных статистических методов, который оценивает отношения между переменными. Это модели связи между зависимой переменной и одной или более независимыми переменными. Она анализирует, как значение зависимой переменной изменяется при изменении значений независимых переменных [19]. Отношение между зависимыми и независимыми переменными представлена на рисунке 4.

Рис. 4. Regression model



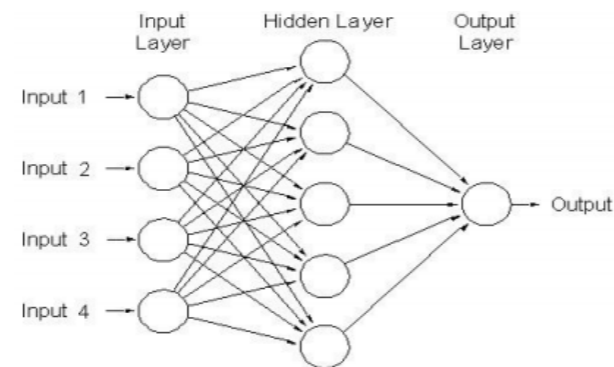
Регрессионная модель находит ключевые паттерны в наборах непрерывных данных, которые предположительно имеют нормальное распределение. Такая модель используется, чтобы узнать как специфические факторы влияют на интересующую нас переменную, анализируя все независимые переменные и сопоставляя их с зависимыми переменными. На выходе мы получаем прогноз изменения значения зависимой переменной в форме регрессионной функции, дающей вероятность распределения.

Есть два типа регрессионных моделей, используемых в предиктивной аналитике: линейная регрессионная модель и модель логистической регрессии. Модель линейной регрессии применяется для моделирования линейного отношения между зависимыми и независимыми переменными с помощью линейной функции. Модель логистической регрессии используется, когда есть категории зависимых переменных. Благодаря этой модели неизвестные значения дискретных переменных, прогнозируются на основе известных значений независимых переменных. Это предполагает ограниченное количество значений переменных.

## 5.3. Искусственная нейронная сеть

Искусственная нейронная сеть имитирует нейроны нервной системы человека и ее системные возможности обработки входных сигналов и вычисление выходных сигналов [20]. Архитектура искусственной нейронной сети общего назначения представлена на рисунке 5.

Рис. 5. Artificial Neural Network



Искусственные нейронные сети используются в прогнозной аналитике как мощный инструмент прогнозирования новых данных на основе имеющегося набора данных. На входной слой сети подаются паттерны данных из обучающей выборки, которые далее передаются скрытым слоям сети, тоже являющихся векторами нейронов. Выход каждого нейрона передается на вход нейрона следующего слоя. В сети могут использоваться различные виды суммирования и функции активации нейронов в зависимости от требований. На выходном слое получают прогнозные новые данные.

Существуют различные модели искусственных нейронных сети и каждая модель использует различные алгоритмы. Метод обратного распространения ошибки — основной алгоритм, используемый для обучения и настройки сети. Искусственные нейронные сети используются также в задачах обучения без учителя. Они могут легко справиться с нелинейными зависимостями в данных. Они также используются в оценке результатов регрессионных моделей и деревьев решений.

## 5.4. Байесовская статистика

Этот метод относится к статистическим, который использует параметры в качестве случайных величин и используется для поиска причинно-следственных связей для определения вероятности наступления события [21]. Байесовская статистика основана на теореме Байеса, которая позволяет как бы «переставить причину и следствие»: по известному факту события вычислить вероятность того, что оно было вызвано данной причиной. События, отражающие действие «причин», в данном случае обычно называют гипотезами, так как они — предполагаемые события, повлекшие данное. Безусловную вероятность справедливости гипотезы называют априорной (насколько вероятна причина вообще), а условную — с учетом факта произошедшего события — апостериорной (насколько вероятна причина оказалась с учетом данных о событии). С одной стороны, подход позволяет выяснить вероятность апостериорного события, учитывая, что произошло событие априори. С другой стороны, теорема Байеса находит вероятность априорного события, учитывая, что апостериори уже произошло. Графически метод представлен на Рисунке 6.

Рис. 6. Bayesian Statistics



Он использует вероятностную графическую модель, которая называется Байесовская сеть, которая представляет собой условную зависимость среди случайных величин. Эта концепция успешно применяется для выяснения причинно-следственных связей. Например, ее можно применять при поиске болезни, основываясь на сим

## 5.5. Ансамбль методов

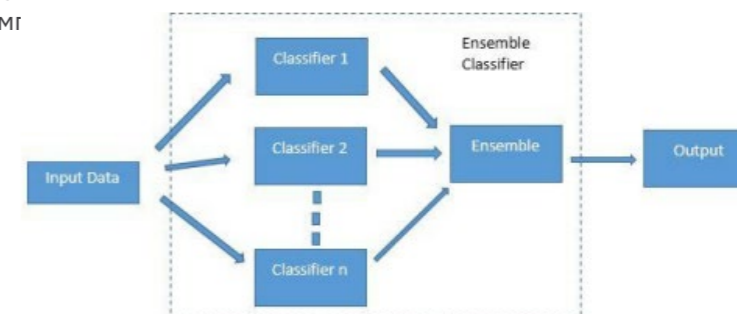
Используется для повышения точности результирующей модели. Несколько предсказателей, которые пытаются получить одну и ту же переменную дадут более точный результат, нежели одиночный предсказатель, поэтому мы объединяем несколько моделей и их предсказания. Применение такого подхода увеличивает точность и уменьшает дисперсию [22]. Пример классификации с использованием ансамблевого обучения представлен на рисунке 7.

Техники ансамблирования впоследствии классифицируются в Бэггинг и Бустинг.

**Бэггинг** — простая техника, в которой мы строим независимые модели и комбинируем их, используя некоторую модель усреднения (например, взвешенное среднее, голосование большинства или нормальное среднее). Обычно берут случайную подвыборку данных для каждой модели, так все модели немного отличаются друг от друга. Примером бэггинга служит модель случайного леса (Random Forest, RF).

**Бустинг** — это техника построения ансамблей, в которой предсказатели построены не независимо, а последовательно. Идея здесь в том, что следующая модель будет учиться на ошибках предыдущей. Они имеют неравную вероятность появления в последующих моделях, и чаще появятся те, что дают наибольшую ошибку. Предсказатели могут быть выбраны из широкого ассортимента моделей, например, деревья решений, регрессия, классификаторы и т.д. Из-за того, что предсказатели обучаются на ошибках, совершенных предыдущими, требуется меньше времени для того, чтобы добраться до реального ответа. Но мы должны выбирать критерий остановки с осторожностью, иначе это может привести к переобучению. Один из самых популярных таких методов — градиентный бустинг.

Рис. 7. Ensemble Classifier

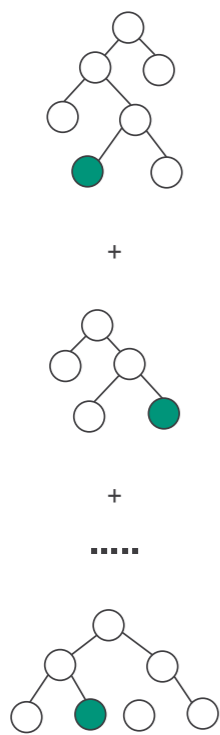


## 5.6. Градиентный бустинг

Градиентный бустинг — это техника машинного обучения для задач классификации и регрессии, которая строит модель предсказания в форме ансамбля слабых предсказывающих моделей, обычно деревьев решений [23]. Причем, каждая итерация моделей работает с ошибками предыдущей итерации, улучшая предсказания. Как только мы достигли момента, когда улучшения нет, мы прекращаем дорабатывать нашу модель (иначе это может привести к переобучению).

Эта модель образно представлена на рисунке 8.

Рис. 8. Gradient Boosting



Алгоритм построения ансамбля примерно такой:

- Сначала строим простые модели и анализируем ошибки;
- Затем определяем точки, которые не вписываются в простую модель;
- Добавляем новые модели, которые обрабатывают сложные случаи, которые были выявлены на модели предыдущего уровня;
- Собираем все построенные модели, определяя веса для каждого предсказателя.

## 5.7. Метод опорных векторов [24, 25]

Идею метода удобно проиллюстрировать на следующем простом примере: даны точки на плоскости, разбитые на два класса (рис. 9а). Проведем линию, разделяющую эти два класса (красная линия на рис. 9а). Далее, все новые точки (не из обучающей выборки) автоматически классифицируются следующим образом:

- точка выше прямой попадает в класс А,
- точка ниже прямой — в класс В.

Рис. 9А

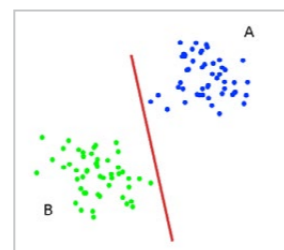
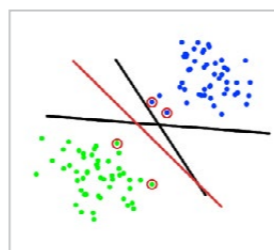


Рис. 9Б



Такую прямую назовем разделяющей прямой. Однако, в пространствах высоких размерностей прямая уже не будет разделять наши классы, так как понятие «ниже прямой» или «выше прямой» теряет всякий смысл. Поэтому вместо прямых необходимо рассматривать гиперплоскости — пространства, размерность которых на единицу меньше, чем размерность исходного пространства. В трехмерном пространстве, например, гиперплоскость — это обычная двумерная плоскость.

Допустим, существует несколько прямых, разделяющих два класса (рис. 9б). С точки зрения точности классификации лучше всего выбрать прямую, расстояние от которой до каждого класса максимально. Другими словами, выберем ту прямую, которая разделяет классы наилучшим образом (красная прямая на рис. 9б). Такая прямая (а в общем случае — гиперплоскость), называется оптимальной разделяющей гиперплоскостью.

Вектора, лежащие ближе всех к разделяющей гиперплоскости, называются опорными векторами (support vectors). На рисунке 9б точки, образующие эти вектора отмечены красным.

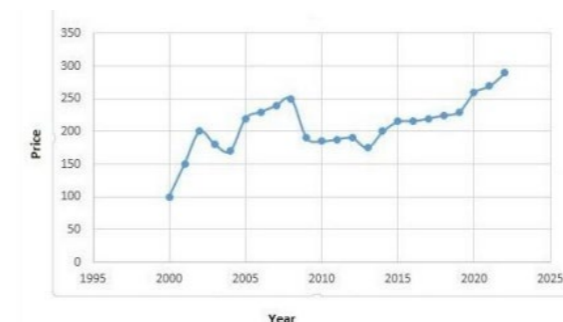
## 5.8. Анализ временных рядов

Анализ временных рядов является статистическим методом, который использует значения переменной во времени, которые собираются в течение определенного временного интервала. Анализ временных рядов разделен на две категории, а именно частотную и временную [26]. Он предсказывает будущее значение переменной в определенном интервале времени, основываясь на анализе значений за прошедшие интервалы времени.

Анализ временных рядов предполагает, что данные содержат систематическую составляющую (обычно включающую несколько компонент) и случайный шум (ошибку), который затрудняет обнаружение регулярных компонент. И нужно отфильтровать шум, чтобы увидеть регулярную составляющую более отчетливо. Регулярная составляющая обычно принадлежит к двум классам: она является либо трендом, либо сезонной составляющей. Тренд представляет собой общую систематическую линейную или нелинейную компоненту, которая может изменяться во времени. Сезонная составляющая — это периодически повторяющаяся компонента. Оба эти вида регулярных компонент часто присутствуют в ряде одновременно. Например, продажи компании могут возрасти из года в год, но они также содержат сезонную составляющую (как правило, 25% годовых продаж приходится на декабрь и только 4% на август).

Анализ временных рядов используется в прогнозировании фондового рынка и прогноза погоды. Пример изменения цены на некоторые продукты за период времени и прогноз будущих изменений представлены на рисунке 10.

Рис. 10. Time Series Analysis



## 5.9. Метод k-ближайших соседей

Это непараметрический метод, используемый в классификации объектов или регрессии. В случае использования метода для классификации объект относится к тому классу, который является наиболее распространённым среди k соседей данного элемента, классы которых уже известны. Результатом является принадлежность объекта к классу. В случае использования метода для регрессии, объекту присваивается среднее значение по k ближайшим к нему объектам, значения которых уже известны, результатом является выявленное свойство объекта [27].

Алгоритм может быть применен к выборкам с большим количеством атрибутов (многомерным). Для этого перед применением нужно определить функцию расстояния; классический вариант такой функции — евклидова метрика.

Это один из самых простых видов машинного обучения. Пример регрессии с использованием k-NN представлен на рисунке 11.

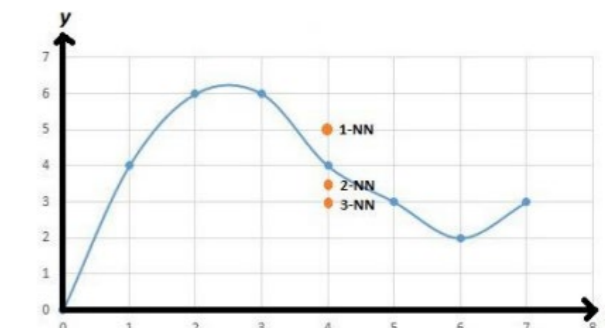


Рис. 11. Regression using k-NN

### 5.10. Метод главных компонент

В анализе данных, как и в любом другом анализе, порой бывает нелишним создать упрощенную модель, максимально точно описывающую реальное положение дел. Часто бывает так, что признаки довольно сильно зависят друг от друга и их одновременное наличие избыточно [28]. Но гораздо чаще бывает так, что признаки зависят друг от друга не очень строго и не очень явно.

Зная зависимости и их степень, мы можем выразить несколько признаков через один, слить воедино, так сказать, и работать уже с более простой моделью. Конечно, избежать потерь информации, скорее всего не удастся, но для минимизации потерь как раз и поможет метод главных компонент.

Данный метод аппроксимирует  $n$ -размерный набор данных до эллипсоида (тоже  $n$ -мерного), полуоси которого и будут являться будущими главными компонентами. И при проекции облака данных на такие оси (то есть снижении размерности) сохраняется наибольшее количество информации.

### 06. ПРИМЕНЕНИЕ ПРЕДИКТИВНОЙ АНАЛИТИКИ

Есть много применений ПА в разнообразных случаях. От постановки диагноза, где болезнь может быть предсказана на основе симптомов до прогноза рынка акций и их доходности. Ниже мы перечислим некоторые из популярных применений.

#### Банковские и финансовые услуги

В банковской и финансовой отраслях существует множество вариантов применения ПА. В обеих отраслях данные и деньги являются важной составляющей, и анализ данных и движения денег является обязательной составляющей деятельности организации. ПА помогает в обнаружении клиентов-мошенников и подозрительных сделок. Минимизирует кредитный риск клиента, по которому компании отрасли предоставляют кредиты. Она помогает в перекрестных продажах и в сохранении и привлечении ценных клиентов. Для финансовых отраслей, где деньги инвестируются в акции или другие активы, прогнозная аналитика прогнозирует возврат инвестиций и помогает принять решение об инвестициях.

#### Розничная торговля

Предиктивная аналитика помогает розничной отрасли определить клиента и понимать то, что ему нужно и что он хочет. Применяя предиктивную аналитику, компания предсказывает поведение покупателей по отношению к продукту. Компании могут фиксировать цены и устанавливать специальные предложения на продукты после выявления паттернов покупательского поведения клиентов. Это также помогает розничной торговле в прогнозировании того, как конкретный продукт будет быть успешным в определенном сезоне. Они могут продвигать свои продукты конкретным клиентам с индивидуальным предложением. Предиктивная аналитика также помогает розничным отраслям в улучшении их цепочки поставок. Они идентифицируют и прогнозируют спрос на продукт в конкретной области могут улучшить поставку продукции [29].

#### Здоровье и Страхование

Фармацевтическая отрасль использует прогнозную аналитику при разработке лекарств и улучшение цепочки поставок. Используя этот метод, компании могут предсказать истечение срока годности лекарств в конкретном регионе из-за спада продаж и точнее спланировать производство. Страховая отрасль использует модели предиктивной аналитики при выявлении и прогнозировании мошеннических претензий клиентов. Здравоохранение применяет предиктивную аналитику, чтобы узнать клиентов которые подвергаются наибольшему риску серьезного заболевания и при продаже страховых планов, подбирают подходящий [30].

#### Нефть, Газ и Коммунальные Услуги

Нефтяная и газовая промышленность используют прогнозную аналитику для прогнозирования отказа оборудования с целью минимизировать риска поломок. Они предсказывают потребность в ресурсах. Потребность в обслуживании прогнозируется, чтобы избежать каких-либо аварий в будущем [31].

#### Государственный сектор

Правительственные агентства используют ПА на основе больших данных для выявления возможных криминальных действий в определенной области. Они анализируют данные социаль-

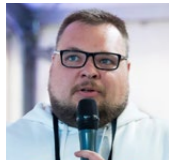
ных сетей, чтобы определить подозрительных лиц и делают прогноз их будущего поведения. Правительства используют предиктивную аналитику для прогнозирования будущих изменений населения на уровне страны и государства. Предиктивная аналитика активно используется в кибербезопасности [32].

### 07. ЗАКЛЮЧЕНИЕ

Существует долгая история использования предиктивной аналитике в осуществлении прогнозов. Ранее статистические модели применялись в качестве прогностических моделей, основанных на выборке данных из большого набора данных. С улучшениями в область компьютерных

наук и развитии компьютеров, были разработаны новые методы и улучшены алгоритмы.

Разработки в области искусственного интеллекта и машинное обучение изменили мир вычислений, где появились интеллектуальные методы вычислений и алгоритмы. Модели машинного обучения очень хорошо используются в качестве прогнозных моделей. Искусственные нейронные сети произвели революцию в области предиктивной аналитике. На основе входных данных, можно предсказать будущее любого параметра. В связи с достижениями в области машинного обучения и развитием методов глубокого обучения, оно все больше используется в предиктивной аналитике.



Голубев П.А.

Principal Data Scientist в Reaktor Dubai,  
Основатель стартапов MaritimeAI.net и pifia.io

# Обзор классических систем поддержки принятия решений

## ДИСКЛЕЙМЕР

Целью написания этой статьи было сделать краткий обзор принципов построения Интеллектуальных Систем Поддержки Принятия Решений (ИСППР), роли машинного обучения, теории игр, классического моделирования и примеров их использования в СППР.

## ВВЕДЕНИЕ

В общем виде, ИСППР — это система, которая ассистирует ЛПР (Лицам, Принимающим Решения) в принятии этих решений, используя инструментарии дата майнинга, моделирования и визуализации, обладает дружелюбным (G)UI, устойчива по качеству, интерактивна и гибка по настройкам.

## ЗАЧЕМ НУЖНЫ СППР:

- Сложность в принятии решений
- Необходимость в точной оценке различных альтернатив
- Необходимость предсказательного функционала
- Необходимость мультипоточного входа (для принятия решения нужны выводы на основе данных, экспертные оценки, известные ограничения и т.п.)

Первые СППР (тогда еще без И) выросли из СПТ (Систем Процессинга Транзакций), в середине 60-х — начале 70-х. Тогда эти системы не обладали никакой интерактивностью, представляя со-

бой, по сути, надстройки над РСУБД, с некоторым (совсем не большим) функционалом численного моделирования. Одной из первых систем можно назвать DYNAMO, разработанную в недрах MIT и представляющую собой систему симуляции каких-либо процессов на основе исторических транзакций. После выхода на рынок мейнфреймов IBM 360 стали появляться и условно-коммерческие системы, применявшиеся в оборонке, спецслужбах и НИИ [33].

С начала 80-х уже можно говорить о формировании подклассов СППР, таких как MIS (Management Information System), EIS (Executive Information System), GDSS (Group Decision Support Systems), ODSS (Organization Decision Support Systems) и др. По сути, эти системы представляли собой фреймворки, способные работать с данными на различных уровнях иерархии (от индивидуального до общеорганизационного), а внутрь можно было внедрить какую угодно логику. Примером может служить разработанная Texas Instruments для United Airlines система GADS (Gate Assignment Display System), которая поощряла принятие решений в Field Operations — назначение гейтов, определение оптимального времени стоянки и т.п. [33]

В конце 80-х появились ПСППР (Продвинутые — Advanced), которые позволяли осуществлять «what-if» анализ и использовали более продвинутой инструментарий для моделирования.

Наконец, с середины 90-х на свет стали появляться ИСППР, в основе которых стали лежать инструменты статистики и машинного обучения, теории игр и прочего сложного моделирования.

## МНОГООБРАЗИЕ СППР

На данный момент существует несколько способов классификации СППР, опишем 3 популярных:

### По области применения

Бизнес и менеджмент	Прайсинг, рабочая сила, продукты, стратегия и т.П.
Инжиниринг	Дизайн продукта, контроль качества...
Финансы	Кредитование и займы
Медицина	Лекарства, виды лечения, диагностика
Окружающая среда	

### По соотношению данные\модели (методика Стивена Альтера)

FDS	File Drawer Systems	Системы предоставления доступа к нужным данным
DAS	Data Analysis Systems	Системы для быстрого манипулирования данными
AIS	Analysis Information Systems	Системы доступа к данным по типу необходимого решения
AFM(s)	Accounting & Financial models (systems)	Системы расчета финансовых последствий
RM(s)	Representation models (systems)	Системы симуляции, AnyLogic как пример
OM(s)	Optimization models (systems)	Системы, решающие задачи оптимизации
SM(s)	Suggestion models (systems)	Системы построения логических выводов на основе правил

### По типу используемого инструментария

Model Driven	в основе лежат классические модели (линейные модели, модели управления запасами, транспортные, финансовые и т.п.)
Data Driven	на основе исторических данных
Communication Driven	системы на основе группового принятия решений экспертами (системы фасилитации обмена мнениями и подсчета средних экспертных значений)
Document Driven	по сути проиндексированное (часто — многомерное) хранилище документов
Knowledge Driven	внезапно, на основе знаний. При чем знаний как экспертных, так и выводимых машинно

## ТРЕБОВАНИЯ К ИСППР

Несмотря на такое многообразие вариантов классификаций, требования и атрибуты СППР хорошо ложатся в 4 сегмента:

- Качество
- Организация
- Ограничения
- Модель

На рис. 1 покажем, какие именно требования и в какие сегменты ложатся:

Отдельно отметим такие важные атрибуты, как масштабируемость, способность обрабатывать плохие данные, юзабилити и user-friendly interface, нетребовательность к ресурсам.

Рисунок 1



**АРХИТЕКТУРА И ДИЗАЙН ИСППР**

Архитектура существующих ИСППР многообразна, единого рецепта нет, да и, пожалуй, быть не может, т.к. каждая организация определяет архитектуру исходя из задач, возможностей и имеющегося инструментария.

Несмотря на разнообразие подходов, осуществляются попытки создать некую унифицированную архитектуру, хотя бы на верхнем уровне.

Действительно, СППР вполне можно разделить на 4 больших слоя:

- Интерфейс
- Моделирование
- Data Mining
- Data collection

На схеме ниже представлено примерное видение архитектуры, которая была реализована в одном из проектов для Министерства Туризма одной из ближневосточных стран, с описанием функционала и примерами инструментов:

Рисунок 2



При построении ИСППР необходимо придерживаться следующих шагов:

- Анализ домена (собственно, где мы будем нашу ИСППР использовать)
- Сбор данных
- Анализ данных
- Выбор моделей
- Экспертный анализ\интерпретация моделей
- Внедрение моделей
- Оценка ИСППР
- Внедрение ИСППР
- Сбор обратной связи (на любом этапе, на самом деле)

На схеме это выглядит так — рис. 3 [34]:

Оценивать ИСППР можно двумя способами. Во-первых, по матрице атрибутов, которая представлена выше. Во-вторых, по критериальному чек-листу, который может быть любым и зависеть от вашей конкретной задачи. В качестве примера такого чек-листа можно привести следующее — рис. 4:

Рисунок 3

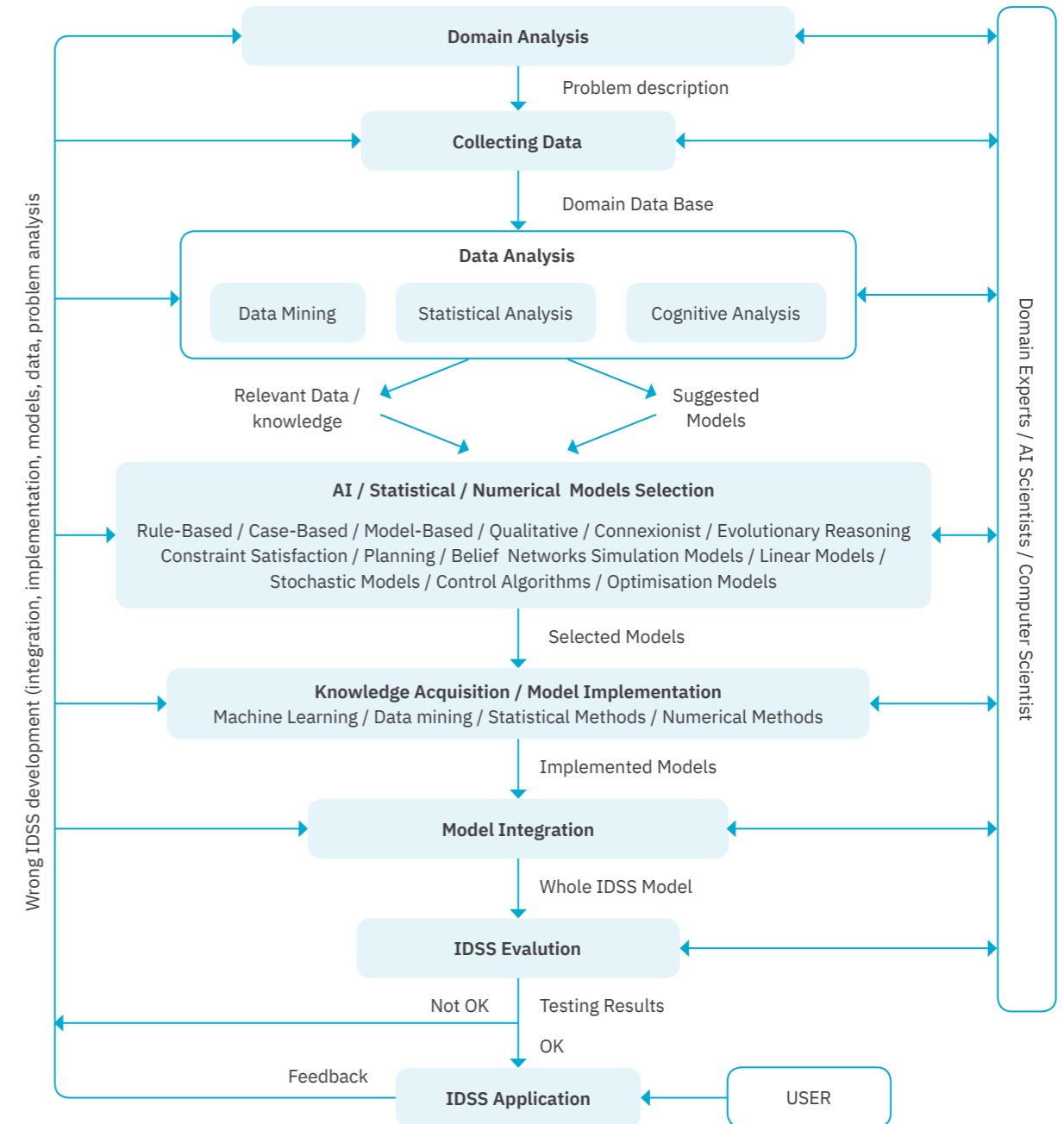


Рисунок 4



## ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЕ В ИСППР

Машинное обучение играет значительную роль в системах поддержки принятия решений, независимо от индустрии.

С одной стороны, есть классические домены, назовем их «тяжелыми», вроде управления цепями поставок, производства, запасов ТМЦ и проч. В тяжелых доменах алгоритмы машинного обучения могут привнести дополнительные инсайты для зарекомендовавших себя классических моделей. *Пример:* предиктивная аналитика по выходам из строя оборудования (машинное обучение) отлично сработается с каким-нибудь FMEA (Failure mode and effects analysis) анализом (классика).

С другой стороны, в «легких» доменах, вроде клиентской аналитики, предсказании churn, выплаты кредитов — алгоритмы машинного обучения также будут на первых ролях. А в скоринге, например, можно совмещать классику с NLP, когда решаем выдавать ли кредит на основе пакета документов (это будет document driven СППР).

## КЛАССИЧЕСКИЕ АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ

*Приведем пример:* менеджеру по продажам стальной продукции надо еще на этапе получения заявки от клиента понимать, какого качества готовая продукция поступит на склад и применить некое управляющее воздействие, если качество будет ниже требуемого.

**В данном случае проблема решается в несколько шагов:**

- Шаг 0.** Определяем целевую переменную (например, содержание оксида титана в готовой продукции).
- Шаг 1.** Определяемся с данными (выгружаем из SAP, Access и других источников данных).
- Шаг 2.** Собираем фичи\генерируем новые.
- Шаг 3.** Рисуем процесс data flow и запускаем его в продакшн.

Подобные «простые» алгоритмы также использует IBM в своей СППР Tivoli, которая позволяет определять состояние своих суперкомпьютеров (Watson в первую очередь): на основе логов выводится информация по перформансу Watson, прогнозируется доступность ресурсов, баланс cost vs profit, необходимость обслуживания и т.п. <sup>[35]</sup>

Компания ABB предлагает своим клиентам DSS800 для анализа работы электродвигателей той же ABB на бумагоделательной линии <sup>[36]</sup>.

Финская Vaisala, производитель сенсоров для минтранса Финляндии использует ИСППР для предсказания того, в какие периоды необходимо применять антиобледенитель на дорогах во избежание ДТП <sup>[35]</sup>.

Опять-таки финская Foredata предлагает ИСППР для HR, которая помогает принимать решения по годности кандидата на позицию еще на этапе отбора резюме <sup>[35]</sup>.

В аэропорту Дубай в грузовом терминале работает СППР, которая определяет подозрительность груза. Под капотом алгоритмы на основе сопроводительных документов и вводимых сотрудниками таможни данных выделяют подозрительные грузы: фичами при этом являются страна происхождения, информация на упаковке, конкретная информация в полях декларации и т.п.

- Шаг 4.** Выбираем и обучаем модель, подключаем ее к пайплайну.
- Шаг 5.** Определяем feature importances.
- Шаг 6.** Определяемся со вводом новых данных.
- Шаг 7.** Пишем простой web-based интерфейс, с которым менеджер взаимодействует как при вводе необходимых параметров, так и в анализе результатов моделирования.

## ОБЫЧНЫЕ НЕЙРОННЫЕ СЕТИ

Кроме простого ML, в СППР отлично ложится и Deep Learning.

Некоторые примеры можно найти в ВПК, например в американской TACDSS (Tactical Air Combat Decision Support System). Там внутри крутятся нейронные сети и эволюционные алгоритмы, помогающие в определении свой-чужой, в оценке вероятности попадания при залпе в данный конкретный момент и прочие задачи <sup>[36]</sup>.

В немного более реальном мире можно рассмотреть такой пример: в сегменте B2B необходимо определить, выдавать ли кредит организации на основе пакета документов.

ИСППР там может строиться так: потенциальный заемщик приносит заранее согласованный пакет документов в офис (или по email), документы попадают в OCR, затем передаются в NLP-алгоритм, который дальше уже делит слова на фичи и скармливает их в NN. В зависимости от качества модели и самих документов решение может быть принято достаточно быстро, пока клиент еще находится в офисе.

Подобными алгоритмами пользуются также в МИД: анкета на визу + прочие справки анализируются прямо в посольстве \ консульстве, после чего сотруднику на экране высвечивается один из 3 смайликов: зеленый (визу выдать), желтый (есть вопросы), красный (соискатель в стоп-листе). Если вы когда-нибудь получали визу в США, то то решение, которое озвучивает вам сотрудник консульства — это именно результат работы алгоритма в совокупности с правилами.

В тяжелых доменах известны также СППР на основе нейронных сетей, определяющие места накопления буфера на производственных линиях (см, например, Tsadiras AK, Papadopoulos CT, O’Kelly MEJ (2013) An artificial neural network based decision support system for solving the buffer allocation problem in reliable production lines. Comput Ind Eng 66(4):1150–1162), Общие Нечеткие Нейронные Сети на основе мин-макса (GFMMNN) для кластеризации потребителей воды (Arsene CTC, Gabrys B, Al-Dabass D (2012)

Decision support system for water distribution systems based on neural networks and graphs theory for leakage detection. Expert Syst Appl 39(18):13214–13224) и другие.

**Вообще стоит отметить, что NN как нельзя лучше подходят для принятия решений в условиях неопределенности, т.е. условиях, в которых и живет реальный бизнес.**

## БАЙЕСОВСКИЕ СЕТИ

Бывает и так, что данные неоднородны по видам появления. Приведем пример из медицины. Поступил к нам больной. Что-то мы про него знаем из анкеты (пол, возраст, вес, рост и т.п.) и анамнеза (перенесенные инфаркты, например). Назовем эти данные статическими. А что-то мы про него узнаем в процессе периодического обследования и лечения (несколько раз в день меряем температуру, состав крови и проч). Эти данные назовем динамическими. Понятно, что хорошая СППР должна уметь учитывать все эти данные и выдавать рекомендации, основываясь на всей полноте информации.

Динамические данные обновляются во времени, соответственно, паттерн работы модели будет такой:

| обучение-решение-обучение,

что похоже на работу врача: примерно определить диагноз, прокапать лекарство, посмотреть за реакцией. Таким образом, мы постоянно пребываем в состоянии неопределенности, поддается лечению или нет. И состояние пациента меняется динамически. Т.е. нам надо построить динамическую knowledge driven СППР.

В таких случаях нам отлично помогут Динамические Байесовские Сети (ДБС) — обобщение моделей на основе фильтров Калмана и Скрытой Марковской Модели.

Разделим данные по пациенту на статические и динамические [37].

Статические	
Код	Описание
Sex	Пол
age1	Возраст
Period_entr	Дата приема
Orig	Происхождение пациента
Detorig	Детальная биография
priseAnti	Реакция на антибиотики
Knaus	Предыдущий диагноз
Cissue	Пациент выжил или умер
Diag	Последний диагноз
Ant	Предпосылки болезни
Result	Предсказание болезни (статическая вероятность)

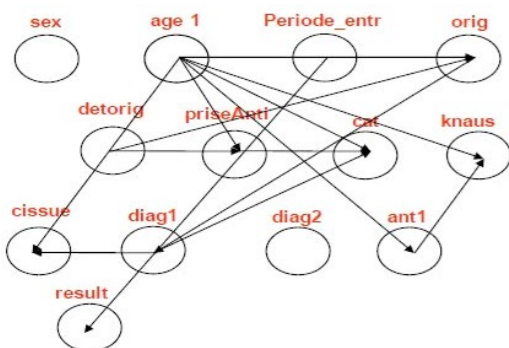
Динамические	
Код	Описание
dsj	Разница во времени между датой приема и текущей
acti	Лечение, примененное в день i
cissuei	Состояние пациента в день i
examinf i	Результаты обследования в день i
sensi	Чувствительность к антибиотику в день i
resulti	Предсказание болезни (динамическое) на день i

Если бы мы строили статическую байесовскую сетку, то нашей задачей было бы посчитать следующую вероятность:

$$p = P(V_i/C),$$

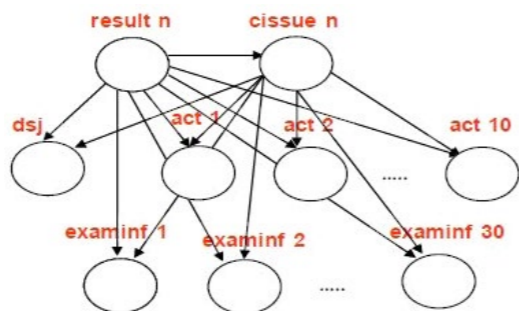
где  $V_i$  — узел сети (вершина графа, по сути), т.е. значение каждой переменной (пол, возраст...), а  $C$  — предсказываемый класс (болезнь).

Статическая сеть выглядит так:

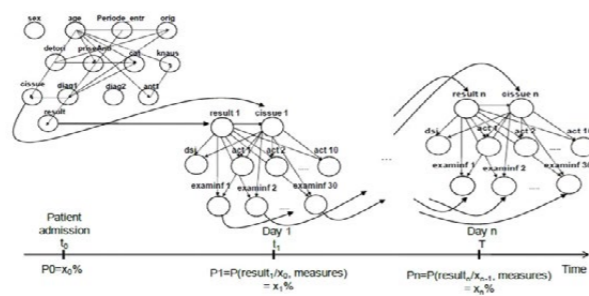


На следующем шаге переходим к динамике, для этого и применим ДБС.

Сначала, в день приема пациента, строим статическую сетку (как на картинке выше). Потом, в каждый день  $i$  строим сетку на основе динамически меняющихся данных:



Соответственно, совокупная модель примет следующий вид:



Таким образом, результат мы рассчитаем по следующей формуле:

$$P(\text{result}1:T) = P_t = 1 \Pi P_i = 1 N P(\text{result}i | P_a(\text{result}i)),$$

где  $T$  — совокупное время госпитализации,  $N$  — количество переменных на каждом из шагов ДБС.

Внедрить эту модель в СППР необходимо несколько иначе — скорее тут надо идти от обратного, сначала эту модель зафиксировать, а потом строить интерфейс вокруг. Т.е., по сути, мы сделали хард модель, внутри которой динамические элементы.

**ТЕОРИЯ ИГР**

Теория игр, в свою очередь, гораздо лучше подойдет для ИСППР, созданных для принятия стратегических решений. Приведем пример.

Допустим, на рынке существует олигополия (малое количество соперников), есть определенный лидер и это не наша компания. Нам необходимо помочь менеджменту принять решение об объемах выпускаемой нами продукции (или, в случае авиалинии, о количестве рейсов на маршруте): если мы будем выпускать продукцию в объеме  $q_2$ , а наш соперник —  $q_1$ , уйдем мы в минус или нет? Для упрощения возьмем частный слу-

чай олигополии — дуополию (2 игрока). В данном кейсе можно начать с классической модели Штакельберга. В этой модели поведение фирм описывается динамической игрой с полной совершенной информацией, при этом особенностью игры является наличие лидирующей фирмы, которая первой устанавливает объем выпуска товаров, а остальные фирмы ориентируются в своих расчетах на неё.

Для решения нашей задачи нам надо всего-то посчитать такое  $q_2$ , при котором решается задача оптимизации следующего вида:

$$Profit_2 = \max(Price(q_1 + q_2) * q_2 - Cost_2(q_2))$$

Для ее решения надо лишь приравнять первую производную по  $q_2$  к нулю.

При этом для такой модели нам понадобится знать только предложение на рынке и стоимость за товар от нашего конкурента, после чего построить модель и сравнить получившееся  $q$  с тем, которое хочет выкинуть на рынок наш менеджмент. Согласитесь, несколько проще и быстрее, чем NN.

Искать победителя в битве ML vs ToG бессмысленно. Слишком разные подходы к решению задачи, со своими плюсами и минусами.

Машинное обучение	Теория игр
Необходимость исторических данных	Нет необходимости в исторических данных
Обобщение на основе данных	Расчет равновесия модели
Предсказание	Расчётный результат

**ЧТО ДАЛЬШЕ?**

С современным состоянием ИСППР вроде бы разобрались, куда идти дальше?

В недавнем интервью Джуда Перл, создатель тех самых байесовских сетей, высказал любопытное мнение. Если слегка перефразировать, то

**«все, чем сейчас занимаются эксперты в машинном обучении, это подгонка кривой под данные. Подгонка нетривиальная, сложная и муторная, но все-таки подгонка».** [37]

Скорее всего, через лет 10 мы перестанем жестко хардкодить модели, и начнем вместо этого повсеместно обучать компьютеры в создаваемых симулируемых средах. Наверное, по этому пути и пойдет реализация ИСППР — по пути general AI. Уже появляются стартапы, такие, как например, лондонский causaLens, которые разрабатывают ИСППР с анализом причинностей для финансовых аналитиков и инвесторов.

Если же посмотреть на более близкую перспективу, то будущее ИСППР за гибкостью решений. Ни один из предложенных способов (классические модели, машинное обучение, DL, теория игр) не универсален с точки зрения эффективности для всех задач. В хорошей СППР должны сочетаться все эти инструменты, при этом разные модули должны использоваться под разные задачи и иметь разные интерфейсы вывода для разных пользователей.



**Варламов Олег Олегович**

Профессор кафедры «Системы обработки информации и управления» (ИУ-5) МГТУ им. Н.Э. Баумана, Президент НИИ МИВАР, доктор технических наук, доцент

## Обзор систем принятия решений, построенных на правилах

Причинно-следственные (каузальные) рассуждения являются основой научного познания человечества и принятия самых сложных решений с их обоснованием (принцип «белого ящика»). В самой простой форме это выражается в построении «цепочек рассуждений» в формате правил: «Если (условие), То (действие)» (или «Что будет, если»). Именно такие каузальные рассуждения и призваны автоматизировать «системы принятия решений, основанные на правилах» или просто «системы, основанные на правилах» (Rule-Based Systems) в области искусственного интеллекта (ИИ).

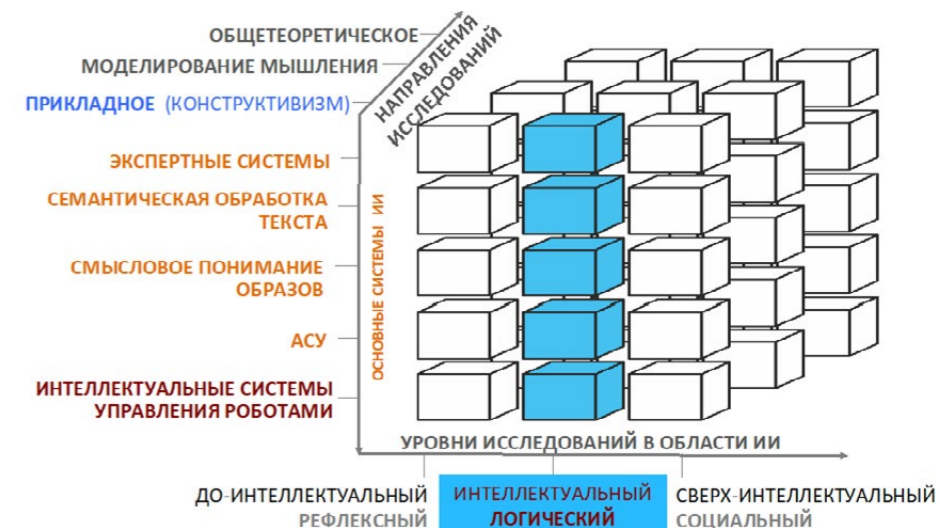
Необходимо сразу определить: машинное обучение (ML) и системы, основанные на правилах (СОП) (Rule-Based Systems — RB) **НЕ ПРОТИВОПОСТАВЛЯЮТСЯ друг другу, а взаимно дополняют**, как рефлекс и логическое мышление человека. Эти подходы относятся к разным уровням исследований в области ИИ, но и возникли они практически одновременно, и развиваются параллельно, и надо использовать их в одном комплексе для достижения наилучших результатов. Такой подход сейчас называют: «гибридные информационные интеллектуальные системы» (ГИИС). Создавать и использовать надо и ML, и RB одновременно!

### СИСТЕМНАЯ МОДЕЛЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В области искусственного интеллекта (ИИ) выделяют различные уровни и направления исследований по созданию систем ИИ, которые можно отобразить в трехмерном представлении — пространстве «Уровни-Направления-Системы» (рис. 1). В области ИИ предложено выделять 3 уровня научных исследований: рефлексный, логический и социальный [38-39]. Отметим, что основные общеизвестные и широко разрекламированные успехи достигнуты пока на рефлексном уровне, т.е. решены многие задачи, которые могут выполнять животные [38-40].

На логическом уровне (системы, основанные на правилах (СОП) (Rule-Based Systems) также есть прорывные открытия, например, автоматический конструктор алгоритмов и/или механизм продукционного логического вывода на миварных сетях с линейной вычислительной сложностью [41-45]. Этот механизм реализован в виде программного продукта КЭСМИ [46] и позволяет обрабатывать (на тестах) более 5 млн правил «Если, То» за сотые доли секунды на обычных компьютерах. Для справки, по оценкам ученых человек может одновременно обрабатывать

Рис. 1. Системная модель научной области Искусственный интеллект.



в своей голове около 1 тысячи правил и думает несколько минут при принятии решений. Следовательно, КЭСМИ в тысячи раз быстрее «думает» (выполняет построение логического вывода и/или конструирует алгоритм своих действий) и в тысячи раз больше информации одновременно обрабатывает на обычном компьютере, который можно разместить в автономных роботах.

Про достижения на социальном уровне ИИ известно гораздо меньше, т.к. эти исследования не финансируются в достаточной степени, поэтому им сложно перейти от фундаментальной науки к прикладным задачам. Однако, в последнее время многие общественные деятели, чиновники и ученые активно обсуждают тематику создания «этики для роботов», которая и находится на этом уровне научных исследований. Мы также считаем эти исследования актуальными и необходимыми для того, чтобы ИИ не стал последним изобретением человечества, а мы могли бы существовать и развиваться параллельно с роботами и искусственным интеллектом.

На рисунке показаны три направления исследований:

1. **общетеоретическое направление** занимается поиском ответа на философский вопрос «Могут ли машины мыслить?» и от него многое зависит в прикладных науках;
2. целью направления **«моделирование мышления»** является создание компьютерной и/или математической полноценной модели человеческого мозга, что важно и нужно, прежде всего, для медицины и оздоровления человека;

3. **прикладное направление** (Конструктивизм) — это фактически автоматизация, которая исходит от возможностей современного и перспективного оборудования и путем развития математики и создания все более совершенного программного обеспечения стремится к созданию умных и автономных компьютерных систем, которые смогут решать «творческие» задачи лучше и быстрее людей.

На третьей оси показаны основные системы ИИ, которые создаются для решения практических задач [47-58].

Прежде всего, это **«экспертные системы»** и научные проблемы представления знаний [39, 43, 46-58]. При классическом подходе ИИ это и есть — системы, основанные на правилах (Rule-Based Systems). Мы под знаниями понимаем факты и правила их обработки, т.е. данные и правила [39, 43, 46, 59-67]. Здесь необходимо не только накапливать огромные объемы знаний, но и уметь быстро с ними работать для принятия адекватных решений в окружающей среде.

**Вторым типом** являются системы понимания естественного языка, когда компьютер или робот должны понимать что им говорит человек [39, 43, 68]. Важно отметить, что на рефлексном уровне рассматриваются как раз системы коммуникаций и отдельных команд, которые умеют выполнять животные по командам своих хозяев. На логическом уровне все гораздо более сложно и робот должен иметь адекватную систему представления «картины мира», т.е. знаний, чтобы в разных контекстах понимать что хочет сказать человек.



**Третий тип** — это системы распознавания образов, к которым прежде всего относятся задачи распознавания картинок и видеоизображений [39, 43, 69, 70]. Учитывая, что в настоящее время системы технического зрения являются много-спектральными, то и картины и видеоизображения также будут различными. В этом направлении уже достигнуты большие успехи и это очень важно, чтобы компьютеры и роботы могли понимать где они находятся и как перемещаться в реальном мире.

**К четвертому типу** относятся «автоматизированные системы управления», которые решают задачи управления различными акторами роботов и сложными техническими системами [39, 43, 71, 72]. Отметим, что АСУ — это и целые заводы, электростанции и т.п. сложные устройства и производства. В последнее время стал популярен термин «киберфизические системы», который более расширяет термин АСУ и объединяет в себе все возможные технические системы с возможностью кибернетического управления: от заводов до умных вещей (чайников и пылесосов) и мультимедийных систем.

**Пятый тип** систем ИИ появился совсем недавно и он объединяет все 4 типа систем, т.к. мы хотим от роботов: наличия знаний, понимания наших указаний, ориентации в пространстве и управления самими собой и другой техникой [39, 43, 44, 73-79]. Следовательно, интеллектуальные системы управления роботами соединяют в себе все основные системы ИИ. Если говорить об автономных роботах, то все указанные системы ИИ должны функционировать на достаточно компактных вычислительных устройствах, что является важным «вызовом» как для математиков (разработка программного обеспечения), так и для физиков (создание оборудования для автономных роботов).

#### «БОЛЬШИЕ ЗНАНИЯ» И ОБУЧЕНИЕ СИСТЕМ, ОСНОВАННЫХ НА ПРАВИЛАХ

Основное направление в исследованиях СОП (RB) — это обработка уже готовых правил, которые составлены «кем-то» заранее. Но в процессе работы СОП новые правила можно добавлять, что является отдельной областью исследований. Таким образом, **вопрос создания правил выносятся за рамки СОП (RB)** и его можно решать разными способами. Классический способ создания правил — это люди «инженеры по знаниям» вручную изучают предметные области, формализуют

знания и создают эти правила. В нашей практике есть примеры, когда один человек за 3 месяца создал формализованную модель предметной области на 360 правил, а 3 человека за 4 месяца создавали модель «по учебнику» с более, чем 3 000 правил. Даже «новички» вполне способны за полгода работы создать 300-500 правил в любой предметной области.

Есть научное направление, когда с помощью машинного обучения ML «находят скрытые правила и зависимости», а потом такие правила также можно «загружать» в СОП (RB).

В настоящее время в области миварных технологий логического ИИ создан продукт для «понимания смысла текстов» ТЭЛМИ (Tel!Mi) [68], который обрабатывает тексты различных документов (инструкций, учебников и т.п.) и автоматически создает причинно-следственные правила в формате миварных сетей «Если, То». На макете, после настройки «парсера» на конкретный вид инструкций в банковской сфере, наш ТЭЛМИ за 8 часов обрабатывал более 1,5 тысяч страниц инструкций и создавал более 3 000 правил с точностью 90%, причем основные ошибки были найдены в текстах исходных инструкций.

Еще одно направление автоматического создания правил применяют в области систем принятия решений для автономных роботов, например, для классической игры в перестановку «кубиков на трех башнях». Этот пример относится к области STRIPS-планирования и MIPRA [78, 79]. Здесь у «робота с одной рукой» ограниченное количество действий в каждый момент времени, но в зависимости от количества кубиков и башен автоматически порождается в реальном времени более 10 000 правил для конкретной задачи, которые алгоритмом MIPRA решаются за секунды на обычном компьютере. Известны и другие способы и методы создания правил для СОП (RB). Но, как было отмечено выше, способы создания правил в «классических» вариантах выходят за область работы «систем, основанных на правилах», т.к. основное внимание уделяется алгоритмам и их сложности работы с правилами, которые собираются в базы знаний или базы правил. Сейчас уже можно говорить о «Больших Знаниях» для RB, по аналогии с «Большими данными» для ML.

#### ЭКСПЕРТНЫЕ СИСТЕМЫ КАК ПРИМЕР RULE-BASED SYSTEMS

Как известно, в области искусственного интеллекта [80-83] «вторая волна» (1980-2000 годы) [84-87] базировалась на «системах, основанных на правилах» (Rule-Based Systems) и была посвящена созданию машин логического вывода, конструированию алгоритмов действий, экспертных систем, систем поддержки принятия решений (СППР), интеллектуальных пакетов прикладных программ (ИППП), ситуационных центров, баз знаний и т.п. [88-95]. Существовали разные трактовки названий этих систем, но для выявления сути и однозначности описания, будем называть их наиболее распространенным [39, 41-67] обобщающим термином — **«экспертные системы» (ЭС)**.

Особенностью ЭС является то, что они не имеют процедур, способных построить в пространстве состояний сразу весь путь решения задачи. Траектория поиска решения полностью определяется данными, получаемыми от пользователя в процессе вывода, что называли «вывод, управляемый потоком входных данных». На научно-популярном языке такие системы должны помогать людям в ситуациях, когда есть некоторое описание предметной области в виде элементарных правил «Если, То» и в зависимости от начальных «исходных» данных человеку или искусственному интеллекту надо выстроить цепочку рассуждений для обоснования и принятия решения. Это близко к классическому школьному решению задач: «Дано; Найти», но когда нет одного алгоритма действий и надо его построить из имеющихся каузальных правил. Собственно, результатом решения задачи и должен быть построенный алгоритм и проведенные по нему вычисления. Важно подчеркнуть, что экспертная система должна обладать способностью объяснить полученное решение, доказать его обоснованность («белый ящик»). ЭС может взаимодействовать с разными пользователями. В некотором смысле, такие задачи относятся к классу «логического вывода», например, на производствах «Если, То». Опустим тонкости и детали, а сосредоточимся на сути ЭС, **в смысле «логических рассуждающих систем принятия решений»**.

Для примера в 1980-х годах была сформулирована проблема автоматического решения классических задач в предметной области «Треугольники» (8-й класс обычной советской школы) [59]. Некое продолжение исследований по созданию Универсального решателя задач (General Problem

Solver, GPS, 1959 г.). Напомним: есть описание треугольника в виде перечня параметров: стороны, углы, периметр, площадь, радиусы вписанных и описанных окружностей — всего более 70 параметров. Все эти параметры связаны между собой множеством зависимостей — уравнений и теорем. Самая простая: сумма углов равна 180 градусам, что порождает 3 «элементарных правила», когда каждый угол равен разнице между 180 градусами и суммой двух других углов. Всего таких «элементарных правил» в области «Геометрия» более 250 шт. Известно, что если треугольник правильно «зафиксирован», то по трем (или нескольким) известным параметрам можно найти любой другой параметр. Задача состоит в том, чтобы по известным «входным» параметрам построить такой алгоритм действий из известных «элементарных правил», который позволит вычислить все необходимые «выходные» параметры. Как говорится, от «Дано» построить алгоритм к «Найти». Казалось бы, простая задача и практически каждый советский школьник мог строить такие алгоритмы в 3-7 шагов. Более сложные алгоритмы тоже были, но не все уже с этим справлялись.

Практическая ценность этой задачи «Треугольники» в том, что она являлась модельной задачей для самых разных областей: от оценки военно-политической обстановки, управления компаниями и до принятия бытовых решений, например, для человека «какой чай и как заварить». С чаем правила тоже элементарные: «если в чайнике есть вода, то можно его кипятить», «если воды в чайнике нет или ее недостаточно, то надо налить воду» и т.д., учитывая выбор типа чая, сахара, лимона и т.п. И правил таких тоже более 250 шт.

Итак, задачи по геометрии — это и есть модельная задача для всех предметных областей [41, 59, 88-95], где также есть наборы взаимосвязанных в виде «Если, То» правил, называемых «база знаний», т.е. это факты и правила их обработки. В такой предметной области возникают задачи, когда по значениям одних известных переменных (фактов) «Дано» необходимо определить другие переменные «Найти», алгоритмов решения готовых нет, но есть правила перехода от одних переменных к другим и надо построить из этих правил алгоритм решения задачи, а потом по этому алгоритму вычислить нужные значения переменных «Найти». Как известно, до 2002 года [41] задача логического вывода на производствах — а именно эта задача и решается в «Треугольни-

ках» — считалась NP-полной и даже элементарные задачи треугольников порождали огромные в 250! (двести пятьдесят факториал) комбинации для перебора. Для справки: уже 20 правил — это 20! (факториал) комбинаций, общее количество которых превышает 10 в 18 степени, а полный перебор на самых быстрых процессорах [96-100] потребует более года работы (рис. 2). Это назвали «проклятие размерности», когда для простейших областей с 30 логическими правилами полный перебор всех вариантов даже с эвристиками и прочими сокращениями перебора (на «деревах решений») занимал часы и потенциальные годы вычислений. Такие задачи есть [101] и их необходимо решать (рис. 3).

Кроме того, задача логического вывода, как и близкая ей по существу, задача автоматического построения алгоритмов действий для роботов и/или киберфизических систем относится к классу «последовательных задач» и принципиально не распараллеливается, т.к. все зависит от входных данных и алгоритм решения должен строиться каждый раз от «Дано» к «Найти». Поэтому, создание многопроцессорных систем [96] и различных «видеокарт и спецпроцессоров» не помогает ускорить решение задачи логического вывода. Итого, задачи «треугольников» никто в мире на компьютерах решать не умел, т.к. надо было уйти от «проклятия размерности» и полного перебора. Собственно, поэтому «вторая волна ИИ» и закончилась «Зимой ИИ».

Рис. 2. Факториал и проблема полно-переборного последовательного логического вывода

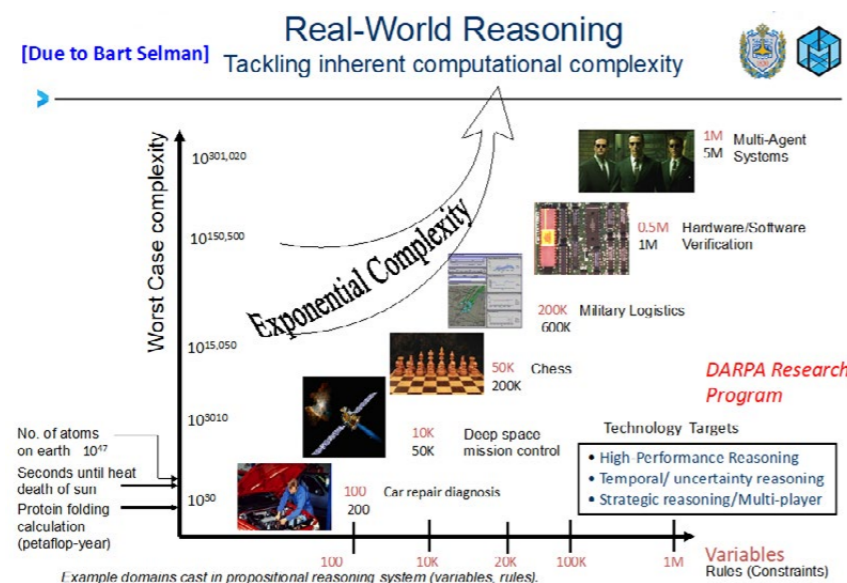
Время решения задач при переборной (N!) вычислительной сложности логического вывода на N правилах

Исходные данные: скорость процессора (операции в секунду) = 10x10<sup>11</sup>

Среднее количество секунд в году = 31 557 600

N	Количество комбинаций (N!)	Время решения (в секундах)	Время решения (в годах)
1	1	<1	<1
2	2	<1	<1
3	6	<1	<1
4	24	<1	<1
5	120	<1	<1
6	720	<1	<1
7	5 040	<1	<1
8	40 320	<1	<1
9	362 880	<1	<1
10	3 628 800	<1	<1
11	39 916 800	<1	<1
12	479 001 600	<1	<1
13	6 227 020 800	<1	<1
14	87 178 291 200	<1	<1
15	1 307 674 368 000	13	<1
16	20 922 789 888 000	209	<1
17	355 687 428 096 000	3 557	<1
18	6 402 373 705 728 000	64 024	<1
19	121 645 100 408 832 000	1 216 451	<1
20	2 432 902 008 176 640 000	24 329 020	<1
21	51 090 942 171 709 400 000	510 909 422	16
22	1 124 000 727 777 610 000 000	11 240 007 278	356
23	25 852 016 738 885 000 000 000	258 520 167 389	8 192
24	620 448 401 733 239 000 000 000	6 204 484 017 332	196 608
25	15 511 210 943 331 000 000 000 000	155 112 100 433 310	4 915 206
26	403 291 461 126 606 000 000 000 000	4 032 914 611 266 060	127 795 352
27	10 888 869 450 418 400 000 000 000 000	108 888 694 504 184 000	3 450 474 513
28	304 888 344 611 714 000 000 000 000 000	3 048 883 446 117 140 000	96 613 286 375
29	8 841 761 993 739 700 000 000 000 000 000	88 417 619 937 397 000 000	2 801 785 304 884
30	265 252 859 812 191 000 000 000 000 000 000	2 652 528 598 121 910 000 000	84 053 559 146 510
31	8 222 838 654 177 620 000 000 000 000 000 000	82 228 388 641 779 200 000 000	2 605 660 333 541 820
32	263 130 636 933 094 000 000 000 000 000 000 000	2 631 308 369 336 940 000 000 000	83 381 130 673 338 100
33	8 683 317 618 811 890 000 000 000 000 000 000 000	86 833 176 188 118 900 000 000 000	2 751 577 312 220 160 000
34	295 232 799 039 604 000 000 000 000 000 000 000	2 952 327 990 396 040 000 000 000 000	93 553 628 615 485 400 000
35	10 333 147 966 386 100 000 000 000 000 000 000 000	103 331 479 663 861 000 000 000 000 000	3 274 377 001 541 990 000 000

Рис. 3. Сложность решения задач в системах, основанных на правилах



СОВРЕМЕННЫЕ ЭКСПЕРТНЫЕ СИСТЕМЫ И RULE-BASED SYSTEMS

В настоящее время «бытует мнение», что Rule-Based Systems умерли и ими никто не занимается. Это не соответствует действительности, даже классические ЭС с N! вычислительной сложностью активно используются. В настоящее время на логическом уровне создания экспертных систем для принятия решений применяют три технологии:

1. Таблицы алгоритмов решений (базы данных).
2. Деревья принятия решений.
3. Продукционный подход.

1. ТАБЛИЦЫ АЛГОРИТМОВ РЕШЕНИЙ (БАЗЫ ДАННЫХ)

Фактически, это технология баз данных [41, 102-105] и она не относится к логическому выводу, но, как это не парадоксально, ее часто применяют для систем принятия решений. Многим хорошо знакома ситуация, когда в определенной ситуации надо «вытащить соответствующий конверт из сейфа» и прочитать в нем алгоритм своих действий. На производствах и в бизнесе стараются заранее подготовить инструкции для самых различных ситуаций, но, как правило, это сильно ограниченный набор ситуаций и количество таких инструкций не превышает десятка (иначе сам человек запутается, какой «конверт» ему открывать). Наш анализ и общение с пользователями показывает, что наибольшее количество заранее просчитанных ситуаций и написанных алгоритмов действий имеет место в управлении техникой и производством. Количество инструкций там может доходить до сотни и более. Фактически люди-эксперты заранее определяют само разнообразие ситуаций и потом по возможности формируют алгоритм действий, т.е. пишут инструкцию. Каждой ситуации присваивается некий код, а технология баз данных позволяет быстро по этому коду найти соответствующую инструкцию.

По нашим оценкам, в настоящее время таким образом можно создавать алгоритмы для нескольких тысяч ситуаций (до 5 тысяч ситуаций мы встречали в некоторых организациях). Главное ограничение: небольшое количество ситуаций, по сравнению с реальным возможным набором ситуаций, который определяется факториалом от количества элементарных действий. Достоинство — быстрый поиск.

2. ДЕРЕВЬЯ ПРИНЯТИЯ РЕШЕНИЙ

Дерево принятия решений — средство поддержки принятия решений. Структура дерева представляет собой «листья» и «ветки» [98]. На ребрах («ветках») дерева решения записаны атрибуты, от которых зависит целевая функция, в «листьях» записаны значения целевой функции, а в остальных узлах — атрибуты, по которым различаются случаи. Чтобы классифицировать новый случай, надо спуститься по дереву до листа и выдать соответствующее значение [88-94]. Подобные деревья решений широко используются в интеллектуальном анализе данных. Цель состоит в том, чтобы создать модель, которая предсказывает значение целевой переменной на основе нескольких переменных на входе. Каждый лист представляет собой значение целевой переменной, изменённой в ходе движения от корня по листу. Каждый внутренний узел соответствует одной из входных переменных. Дерево может быть также «изучено» разделением исходных наборов переменных на подмножества, основанные на тестировании значений атрибутов. Это процесс, который повторяется на каждом из полученных подмножеств. Рекурсия завершается тогда, когда подмножество в узле имеет те же значения целевой переменной, таким образом, оно не добавляет ценности для предсказаний.

Основное достоинство — реальное большое количество возможных ситуаций и алгоритмов действий, которое определяется факториалом от количества элементарных правил. Главным недостатком технологии «Дерево принятия решений» является то, что проблема получения оптимального дерева решений является NP-полной с точки зрения некоторых аспектов оптимальности даже для простых задач. Таким образом, практическое применение алгоритма деревьев решений основано на эвристических алгоритмах, таких как алгоритм «жадности», где единственно оптимальное решение выбирается локально в каждом узле. Такие алгоритмы не могут обеспечить оптимальность всего дерева в целом.

3. ПРОДУКЦИОННЫЙ ПОДХОД И ТЕХНОЛОГИЯ КАУЗАЛЬНЫХ ПРАВИЛ «ЕСЛИ, ТО»

Продукционная модель. Прежде всего, основная технология создания ЭС — это продукционная модель знаний, основанная на правилах, которая позволяет представить знание в виде предложений типа «Если (условие), то (действие)». В общем случае продукционную модель можно представить в следующем виде [41, 47-67]:

$i = \langle S; L; A \rightarrow B; Q \rangle$ , где:

S — описание класса ситуаций;

L — условие, при котором продукция активизируется;

$A \rightarrow B$  — ядро продукции;

Q — постусловие продукционного правила.

Классическая продукционная модель обладает тем недостатком, что при накоплении достаточно большого числа (порядка нескольких сотен) продукций они начинают противоречить друг другу. В этом случае разработчики начинают усложнять систему, включая в неё приоритеты правил, эвристики по выбору и сокращению перебора и т.п. Основное достоинство продукционного подхода — реальное большое количество возможных ситуаций и алгоритмов действий, которое определяется факториалом от количества элементарных правил. А ограничение на вычислительную сложность логического вывода будет разным и зависит от способа представления знаний и алгоритма поиска: от полного перебора для Пролога и Rete-сетей, до линейной для миварных сетей.

**Язык Пролог.** Логический подход может быть проиллюстрирован применением для этих целей языка и системы логического программирования Пролог. Программы, записанные на языке Пролог, представляют наборы фактов и правил логического вывода без жесткого задания алгоритма как последовательности действий, приводящих к необходимому результату. Для каждой предметной области пишется своя программа. **Главное ограничение — «обход дерева решений», полный перебор всех правил** и NP сложность.

**Алгоритм Rete** — эффективный алгоритм сопоставления с образцом для продукционных систем, экспертных систем и баз знаний, созданный Чарльзом Форги [106-108]. Rete стал основой многих популярных экспертных систем, включая CLIPS, Jess, Drools, BizTalk Rules Engine и Soar. При классической реализации экспертная система выполняет полный перебор и проверяет применимость каждого правила вывода к каждой переменной (факту базы знаний), при необходимости выполняет его и переходит к следующему правилу, возвращаясь в начало при исчерпании всех правил. NP полнота приводит к тому, что даже для небольшого набора правил (до 30 шт.) и фактов такой метод работает неприемлемо медлен-

но. Алгоритм Rete обеспечивает более высокую эффективность, чем «обход дерева решений». При использовании Rete экспертная система строит несколько префиксных деревьев, важную роль играет порядок записи правил и т.п.

Алгоритм Rete жертвует объемом памяти ради скорости. За счет этих эвристик сокращается перебор вариантов. Но в самом «плохом случае», а именно так и определяют вычислительную сложность, вычислительная сложность становится факториал от N (полный перебор).

Известны развития в виде Rete II и Rete-NT. Rete II улучшен по двум параметрам: повышена общая производительность сети включая хешированную память для больших массивов данных, добавлен алгоритм обратного вывода, работающий на той же сети. Скорость обратного вывода по сравнению с Rete I повышена значительно. Rete-NT разработан в 2010 году и это самое новое поколение, которое используется для машины вывода продукта SMARTS. Считается, что Rete-NT в 500 раз быстрее Rete и в 10 раз быстрее Rete II. Но **вычислительная сложность этих модификаций для худшего случая все равно осталась факториал от N (полный перебор — N!).**

**Миварные сети представления знаний и MOGAN** На основе объединения продукционного подхода «Если, То» и сетей Петри в 2002 г. [41] создан формализм «двудольных ориентированных сетей», названных «миварные сети». В явном виде из продукций выделили параметры «Если» и «То», которые стали называть «входные» и «выходные» параметры. Параметры стали одной «долей» графа, а второй «долей» — «правила» (рис. 4). Использованы идеи сетей Петри: если все входные параметры одного правила получают значение, то правило обязательно выполняется и все выходные параметры получают значение (рис. 4). Из общего списка правил автоматически выбирали правила с известными входными параметрами, запускали их и определяли их выходные параметры (рис. 5).

Миварные сети позволили **выполнять логический вывод без перебора**, что снизило вычислительную сложность логического вывода с N! до **линейной** относительно правил (рис. 6) [38, 39, 41-46, 59-79]. Базовый миварный алгоритм логического вывода бесплатный, т.к. он опубликован в монографии на русском языке в 2002 г. [41] и на английском — в 2011 г. [42].

Рис. 4. Переход от Продукций и сетей Петри к миварным сетям

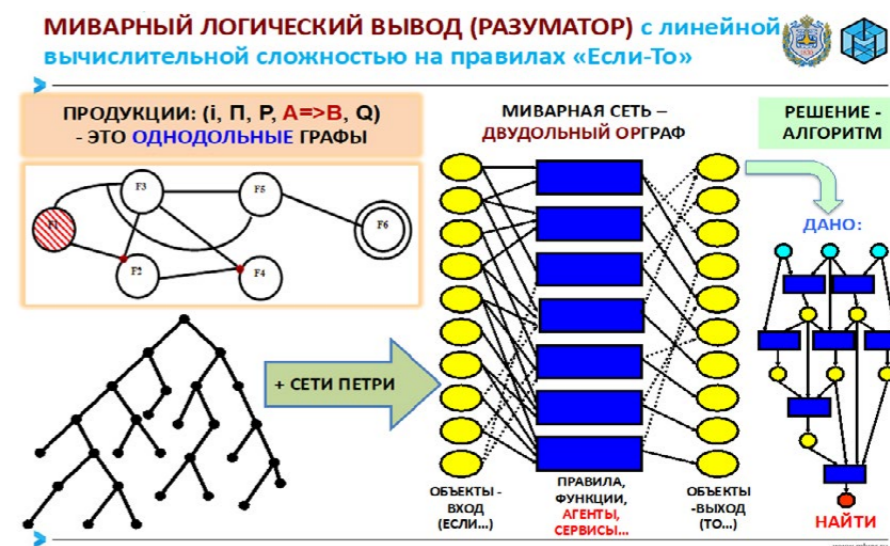


Рис. 5. Пример работы миварных сетей без перебора правил



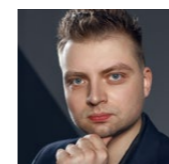
**Для справки: сети Петри** — математический аппарат для моделирования динамических дискретных систем, предложенный Карлом Петри еще в 1962 году [109-110]. Сеть Петри представляет собой двудольный ориентированный мультиграф, состоящий из вершин двух типов — позиций и переходов, соединённых между собой дугами. Вершины одного типа не могут быть соединены непосредственно. В позициях могут размещаться метки (маркеры), способные перемещаться по сети. Событием называют срабатывание перехода, при котором метки из входных позиций этого перехода перемещаются в выходные позиции. События происходят мгновенно либо одновременно, при выполнении некоторых условий. Сети Петри не используют при создании экспертных систем и для принятия решений.

Многолетние исследования миварных технологий позволили создать новый мощный, универсальный и быстрый инструмент — многомерная открытая гносеологическая активная сеть — multidimensional open gnoseological active net (MOGAN). Этот инструмент объединяет миварное информационное пространство эволюционного накопления данных, основанное на гносеологической модели «Вещь-Свойство-Отношение» [111-113], и миварные сети логико-вычислительной обработки информации. Все вместе это позволяет **MOGAN** быстро и просто конструировать алгоритмы и работать с логическими рассуждениями формата «Если, То». MOGAN используют для обработки причинно-следственных зависимостей и создания больших баз знаний прикладных систем ИИ нового поколения (Rule-Based Systems). Это и есть переход к «Большим знаниям».

В 2015 году НИИ МИВАР создал «Конструктор экспертных систем миварный» КЭСМИ Wi!Mi Разуматор [8-9, 22-42], обрабатывающий более 5 млн правил/мс. Практическое использование показало, что задачи с предметными базами знаний от 300 до 3000+ правил действительно решаются за доли секунд на обычном ноутбуке.

**Вывод.** «Автоматизация (информатизация) — это материализация человеческих идеальных мыслей в виде алгоритмов и программ компьютеров». Автоматизация позволяет отделить от человека его мысли и выполнять их автономно от своего создателя на компьютерах.

До появления компьютеров, по существу, таких возможностей у человечества не было. В этом смысле, под искусственным интеллектом принято было понимать такие компьютерные системы, которые превзойдут человека по мыслительным способностям — по полной аналогии с обычными механизмами и машинами.



**Константин Коточигов**

Руководитель отдела анализа данных Weborama

## Анатомия рекомендательных систем

### ВВЕДЕНИЕ

Задача любой рекомендательной системы — проинформировать пользователя о товаре, который ему может быть наиболее интересен в данный момент времени. В результате клиент получает ценную для него информацию, а сервис зарабатывает на повышении лояльности клиента или на прямой продаже.

Персонализация онлайн-маркетинга — очевидный тренд последнего десятилетия. Львиная доля выручки таких гигантов рынка как Amazon или Netflix приходится именно на рекомендованные товары и процент этот, вероятно, будет расти.

Рекомендательные системы получили широкое распространение в самых разных бизнес-доманах. Для некоторых компаний они являются ядром бизнеса, например в классических сервисах рекомендаций (TripAdvisor), для других — выполняют сервисную функцию, призванную улучшить customer experience и сделать поиск интересующего товара или контента более удобным.

Иногда термин «рекомендация» употребляют в более широком смысле, имея в виду любые системы, оптимизирующие операционные процессы внутри компании (такие как таргетирование рекламных кампаний, оптимизация ценообразования или персонализация коммуникаций). Однако в данном обзоре будем придерживаться классического определения рекомендаций — предложение наиболее интересного товара.

Проиллюстрировать многообразие рекомендательных систем можно перечислив несколько важных характеристик.

Во-первых, что рекомендуется. Это могут быть товары (Amazon, Ozon), статьи (Arxiv.org), новости (Surfingbird, Яндекс.Дзен), изображения (500px), видео (YouTube, Netflix), люди (LinkedIn, LonelyPlanet), музыка (Last.fm, Pandora) или любой другой контент.

По степени персонализации рекомендации ранжируются от полностью неперсональных, когда всем пользователям рекомендуется одно и то же (различия только по региону или времени) до персональных — когда рекомендации используют всю доступную информацию о клиенте, в том числе историю его покупок. Промежуточный вариант — когда используется история действий в рамках текущей сессии.

Шаблоны потребления у людей разные, и не обязательно должны рекомендоваться новые товары. Можно показывать повторные позиции, например, для пополнения запаса. По этому принципу выделяют две группы товаров:

- Повторяемые. Например, шампуни или бритвенные станки, которые нужны всегда.
- неповторяемые. Например, книги или фильмы, которые редко приобретают повторно.

Если продукт нельзя явно отнести к одному из классов, имеет смысл определять допустимость повторных рекомендаций индивидуально (кто-то ходит в магазин только ради арахисового масла определенной марки, а кому-то важно попробовать все, что есть в каталоге).

Понятие «интересности» тоже субъективное. Некоторым пользователям нужны вещи только из их любимой категории (conservative recommendations), а кто-то, наоборот, больше откликается на нестандартные товары или группы товаров (risky recommendations). Например, видеохостинг может рекомендовать пользователю только новые серии любимого сериала, а может периодически закидывать ему новые шоу или вообще новые жанры. В идеале стратегия показа должна определяться с помощью профилирования клиента.

Также в подобных системах крайне важен фактор доверия. Прозрачность механизма рекомендации гарантирует, что система не продвигает проплаченный или более дорогой товар. Кроме того, система должна уметь бороться с купленными отзывами или накрутками продавцов (завышения оценок бывают и непреднамеренными, например из-за раннего всплеска интереса к вышедшему фильму).

Отчасти этого позволяют добиться объяснения, почему была выведена та или иная товарная рекомендация. Например

- «Вам может понравиться фильм X поскольку там играют Y и Z»
- «Пользователи с похожими на ваш музыкальными вкусами оценили альбом XYZ на 4.5 из 5».

## ПОСТАНОВКА ЗАДАЧИ

В центре любой рекомендательной системы находится так называемая матрица предпочтений. Это матрица, по одной из осей которой отложены все клиенты сервиса (Users), а по другой — объекты рекомендации (Items). На пересечении некоторых пар (user, item) данная матрица заполнена оценками (Ratings) — это известный нам показатель заинтересованности пользователя в данном товаре, выраженный по заданной шкале (например от 1 до 5).

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	0	3	0	3	0	0
User 2	4	0	0	2	0	0
User 3	0	0	3	0	0	5
User 4	0	0	0	0	3	0
User 5	4	0	0	4	0	0

Пользователи обычно оценивают лишь небольшую часть товаров, что есть в каталоге, и задача рекомендательной системы — обобщить эту информацию и предсказать отношение клиента к другим товарам, про которые ничего не известно. Другими словами нужно заполнить все незаполненные ячейки на картинке выше.

Пользовательские оценки можно получить двумя способами:

- явно (explicit ratings)** — пользователь сам ставит оценку товару / оставляет отзыв / ставит лайк;
- неявно (implicit ratings)** — оценка восстанавливается исходя из действий пользователя: купил товар — значит он ему нравится (вес=0.8), долго читал описание — значит есть интерес (вес=0.4) и т.п.

Разумеется, явные оценки содержат больше информации о предпочтениях, однако явное оценивание возможно далеко не во всех сервисах. Часто оба типа оценок суммируются, в результате чего конструируется комбинированная оценка.

Поскольку все пользователи оценивают по-разному — кто-то всем подряд пятерки ставит, а от кого-то четверки не допросишься — перед расчетом рекомендаций матрицу предпочтений обычно нормализуют, приводя к единой шкале (с обратным преобразованием при выводе этих рекомендаций). Делается это одним из следующих способов:

- центрированием (mean-centering)** — из оценок просто вычитаем его среднее (по пользователю или по товару)
- стандартизацией (z-score)** — помимо центрирования делим оценку на стандартное отклонение
- двойной стандартизацией** — сначала нормируем по пользователю, затем по товару

Если у фильма «Самый лучший фильм» средняя оценка 2.5, а пользователь ей ставит 5.0, то это сильный фактор, говорящий о том, что такие фильмы ему действительно сильно нравятся.

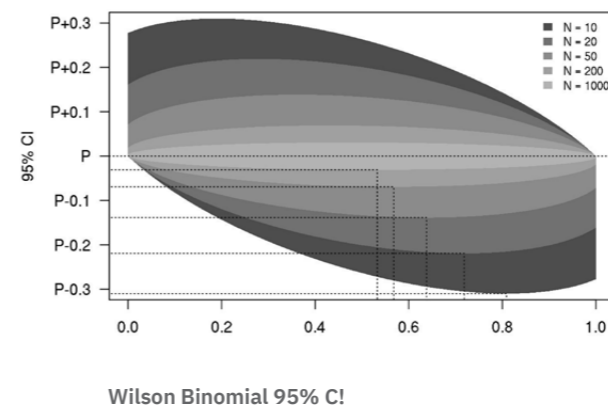
## НЕПЕРСОНАЛИЗИРОВАННЫЕ РЕКОМЕНДАЦИИ

Самый базовый подход — неперсонализированные рекомендации. Потенциальный интерес пользователя определяется просто средним рейтингом товара: «Всем нравится — значит понравится и вам». Это стандартная история, когда пользователь не авторизуется в системе (например, на том же Booking.com).

Холодный старт — это типичная ситуация, когда ещё не накоплено достаточное количество данных для корректной работы рекомендательной системы (например, когда товар новый или просто его очень редко покупают). В таких ситуациях рейтинги могут искусственно корректировать.

**Первый способ** — замена среднего рейтинга на сглаженное среднее (damped mean). Смысл таков: сначала в оценке рейтинга доминирует некое «безопасное» среднее значение, а после набора статистики вес постепенно переходит ко второму слагаемому — реальному набору оценок.

**Другой способ** — рассчитывать по каждому рейтингу интервалы достоверности (confidence Intervals). Консервативные системы могут вывести в качестве оценки, например, нижнюю границу интервала (Low CI Bound). Рейтинги, как правило, ограничены определенной шкалой (например от 1 до 5), поэтому при подсчете приходится использовать модифицированные интервалы. Один из вариантов — Wilson Confidence Interval, который гарантирует, что распределение вероятности не выходит за границы шкалы.



В системах с информационным контентом (новости / статьи / посты) также важно учитывать «временной» фактор — более «свежие» записи должны чаще попадать в топ рекомендаций. Для этого при расчете рейтинга в формулу добавляют время жизни записи (T). Ниже приведена пара примеров:

Ранжирование статей в журнале Hacker news:

$$Rank = \frac{(U - D - 1)^{0.8} * P}{T^{1.8}}$$

где U = upvotes, D = downvotes, а P (Penalty) — дополнительная корректировка для имплементации иных бизнес-правил

Расчет рейтинга в Reddit:

$$Rank = \log_{10} (\max(1, U - D)) - \frac{|U - D|T}{const}$$

где U = число голосов «за», D = число голосов «против», T = время записи. Первое слагаемое оценивает «качество записи», а второе делает поправку на время.

## CONTENT-BASED РЕКОМЕНДАЦИИ

Персональные рекомендации уже используют информацию о самом пользователе, в первую очередь о его предыдущих покупках.

Для подхода content-based filtering необходимо наличие описания товара. Это может быть, в том числе, текстовое описание в свободной форме. Суть подхода: описания товаров сопоставляются с интересами пользователя, полученными из его предыдущих оценок. Далее из наиболее релевантных товаров формируется список рекомендаций.

Каждый товар описывается вектором в некотором признаковом пространстве (Vector-Space model). Каждый пользователь описывается набором векторов в этом же пространстве (по одному вектору на ранее оцененный товар), а также нормализованной суммой этих векторов — вектором предпочтений.

По мере взаимодействия пользователя с системой (скажем, он покупает фильмы), вектор предпочтений обновляется. Имеет смысл рекомендо-

вать фильмы наиболее близкие к этому вектору, которые пользователь еще не видел. В качестве меры близости чаще всего используется косинусное антирасстояние (cosine similarity).

$$sim(A, B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

При подсчете близости старые оценки имеют смысл учитывать с меньшим весом, более актуальные — с большим, так как любые предпочтения обладают свойством меняться со временем.

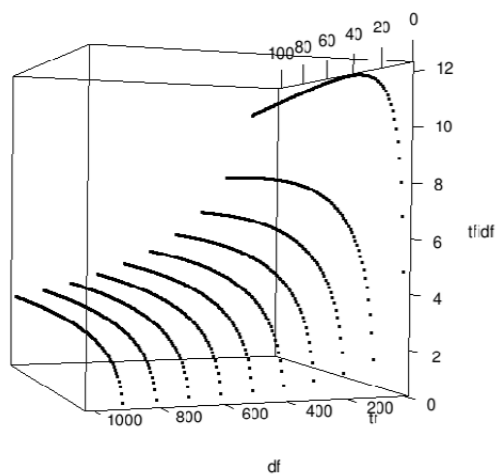
Кроме того, при расчете близости разные атрибуты должны учитываться с разным весом. Ведь совпадение редких интересов (например, атрибут «сербский артхаус») больше говорит о близости товара и пользователя, чем совпадение популярных (атрибут «полнометражные фильмы»). Для этого используется стандартное для NLP взвешивание TF-IDF (или его многочисленные модификации вроде BM25). При этом TF-IDF взвешивание применимо как к отдельным словам из текстового описания, так и к стандартным номинальным атрибутам (жанр, режиссер, язык).

$$W_{x,y} = tf_{x,y} \times \log\left(\frac{N}{df_x}\right)$$

**TF-IDF**

Вес слова  $x$  в описании товара  $y$

$tf_{x,y}$  = Частота слова  $x$  в описании товара  $y$   
 $df_x$  = Количество товаров, содержащих слово  $x$   
 $N$  = Общее количество товаров



Разумеется, возможно расширение модели любыми техниками из NLP, например, вместо отдельных слов можно использовать шинглы или  $n$ -граммы (последовательные пары слов, тройки и т.д.).

**КОЛЛАБОРАТИВНАЯ ФИЛЬТРАЦИЯ (USER-BASED ВАРИАНТ)** [114, 115]

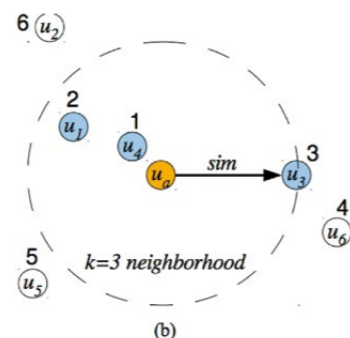
Данный класс систем начал активно развиваться в 90-е годы. В рамках подхода рекомендации генерируются на основании интересов других пользователей, близких по своим интересам к анализируемому. Такие рекомендации являются результатом «коллораации» множества пользователей, отсюда и название метода.

Классическая реализация алгоритма основана на принципе  $k$  ближайших соседей. Для каждого пользователя мы ищем  $k$  наиболее похожих на него (в терминах интересов) и дополняем его предпочтения основываясь на интересах соседей. Например, если известно, что близкие к вам пользователи в восторге от фильма «Кровь и бетон», а вы его по какой-то причине еще не смотрели, это отличный повод предложить данный фильм для субботнего просмотра.

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$	$i_6$	$i_7$	$i_8$
$u_1$	?	4.0	4.0	2.0	1.0	2.0	?	?
$u_2$	3.0	?	?	?	5.0	1.0	?	?
$u_3$	3.0	?	?	3.0	2.0	2.0	?	3.0
$u_4$	4.0	?	?	2.0	1.0	1.0	2.0	4.0
$u_5$	1.0	1.0	?	?	?	?	?	1.0
$u_6$	?	1.0	?	?	1.0	1.0	?	1.0

$u_a$	?	?	4.0	3.0	?	1.0	?	5.0
$\hat{r}_a$	3.5	4.0			1.3			2.0

$$\hat{r}_{ui} = \frac{1}{\sum_{j \in S(i)} s_{ij}} \sum_{j \in S(i)} s_{ij} r_{uj}$$



$$sim_{Cosine}(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

На картинке выше в матрице предпочтений желтым цветом выделен пользователь, для которого мы готовим рекомендации (знаки вопроса). Синим цветом выделены три его ближайших соседа.

Близость по интересам может считаться разными метриками. Некоторые типичные варианты:

Корреляция Пирсона

$$p = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

Корреляция Спирмена

$$p = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Косинусное расстояние

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Расстояние Жаккара

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Расстояние Хэмминга

$$D_H = \sum_{i=1}^k |x_i - y_i|$$

В качестве анализируемой окрестности пользователя можно выбирать:

- к ближайших соседей
- всех пользователей в базе
- всех пользователей с близостью выше пороговой
- случайную выборку

В качестве числа соседей ( $k$ ) исследователи из компании MovieLens рекомендуют брать 30-50 для фильмов и 25-100 для других товаров.

Одно из развитий коллаборативного подхода — так называемые Trust-based рекомендации, когда учитывается не только близость по интересам, но и «социальная» близость (число общих друзей). В каком-нибудь Вконтакте, например, вполне может рекомендоваться музыка из плейлистов друзей.

Для корректной работы коллаборативного подхода принимаются 2 допущения:

- а. вкусы людей не меняются со временем (или меняются равномерно)
- б. если вкусы людей совпадают, то они совпадают во всем. Это не всегда так, особенно когда рекомендуемые товары неоднородны — у пары клиентов могут совпадать предпочтения в еде, а политические взгляды быть прямо противоположными. Это важно учитывать.

У классической реализации алгоритма есть один явный минус — он плохо применим на практике из-за квадратичной сложности  $O(n^2m)$ , где  $n$  — число пользователей, а  $m$  — число товаров. То есть для миллиона пользователей нужно будет научиться эффективно хранить минимум 4ТБ данных.

На практике чаще прибегают к альтернативным методам — либо используют итеративные подходы (например, ALS), либо выбирают item-based вариант коллаборативной фильтрации.

**КОЛЛАБОРАТИВНАЯ ФИЛЬТРАЦИЯ (ITEM-BASED ВАРИАНТ)**

Подход Item-based является естественной альтернативой классическому подходу User-based, описанному в первой части, и почти полностью его повторяет, за исключением одного момента — применяется он к транспонированной матрице предпочтений. Т.е. алгоритм исследует «близкие» товары, а не пользователей — действительно, если продукты содержательно похожи, то скорее всего они либо одновременно нравятся, либо одновременно не нравятся.

**Преимущества Item-based перед User-based:**

- матрица расстояний, которую необходимо рассчитать и хранить, уменьшается до размеров  $n \times n$ , где  $n$  — число товаров (которое обычно кратно меньше числа клиентов)
- Так как пользователей больше, оценка близости товаров имеет гораздо меньшую дисперсию ошибки — оценка более устойчивая
- для выбранного товара можно сгенерировать гораздо больше рекомендаций, чем для выбранного пользователя, так как у столбцов матрицы предпочтений пересечение больше, чем у строк

Число соседних товаров  $N$  обычно рекомендуют выбирать в районе 20.

При использовании item-based подхода рекомендации имеют тенденцию быть более консервативными. Действительно, разброс рекомендаций получается меньше и следовательно меньше вероятность показать нестандартные товары.

Если в качестве рейтинга используется факт просмотра описания товара, то рекомендуемые товары скорее всего будут аналогами — товарами, которые часто смотрят вместе. Если же рейтинги в матрице предпочтений мы рассчитываем на основании покупок, то скорее всего рекомендуемые товары будут аксессуарами — товарами, которые часто покупают вместе.

Если рейтинг товара прогнозируется в рамках непрерывной шкалы (0-10), используют метрики класса Prediction Accuracy.

Название	Формула	Описание
MAE (Mean Absolute Error)	$E( P-R )$	Среднее абсолютное отклонение
MSE (Mean Squared Error)	$E( P-R ^2)$	Среднеквадратичная ошибка
RMSE (Root Mean Squared Error)	$E( P-R )$	Корень из среднеквадратичной ошибки

Если прогнозируется наличие интереса (купит / не купит), используют метрики класса Decision Support.

Название	Формула	Описание
Precision	$\frac{TP}{TP+FP}$	Доля рекомендаций, понравившихся пользователю
Recall	$\frac{TP}{TP+FN}$	Доля интересных пользователю товаров, которая показана
F1-Measure	$2\frac{PR}{P+R}$	Среднее гармоническое метрик Precision и Recall. Полезно, когда заранее невозможно сказать, какая из метрик важнее
ROC AUC		Насколько высока концентрация интересных товаров в начале списка рекомендаций
Precision@N		Метрика Precision, посчитанная на Top-N записях
Recall@N		Метрика Recall, посчитанная на Top-N записях
AverageP		Среднее значение Precision на всем списке рекомендаций

При этом любые непрерывные оценки всегда можно свести к бинарным {да / нет} просто делением шкалы на две половины.

**ТЕСТИРОВАНИЕ**

Генерацию рекомендаций можно рассматривать как стандартную предиктивную модель машинного обучения. Поэтому тестируется она тоже по стандартному протоколу:

1. offline тестирование модели на исторических данных с помощью ретро-тестов
2. тестирование в среде pre-production с помощью A/B тестов — запускаем несколько вариантов, смотрим какой дает лучший результат
3. мониторинг системы в production (логи + фидбек от пользователей)

При offline тестировании часть оценок маскируется и отдается под тест, а на остальных происходит обучение моделей. Как правило, оценок не так много, поэтому для получения более достоверных показателей качества используется кросс-валидация — многократное повторение теста с усреднением результатов. Обычно используется подход leave-one-out или leave-p-out — когда маскируется  $p$  оценок, а на остальных мы обучаемся.

Качество рекомендаций — величина субъективная, и может измеряться множеством метрик.

Если оценивается качество списка выводимых рекомендаций, то используются метрики класса Rank Accuracy, которые измеряют общую релевантность расположения товаров в этом списке.

Название	Формула	Описание
Mean Reciprocal Rank	$E(1pos)$	На какой позиции списка рекомендаций пользователь находит первую полезную
Spearman Correlation	$E( P-R )$	Корреляция (Спирмена) реального и прогнозируемого рангов рекомендаций
nDCG	$\sum R(i)\max(1, \log(i))$	Информативность выдачи с учетом ранжирования рекомендаций
Fraction of Concordance Pairs	$P(XR > XP)$	Насколько высока концентрация интересных товаров в начале списка рекомендаций

При A/B тестировании системы важно смотреть именно на инкрементальный эффект от рекомендаций — то есть сравнивать отклик с базовым режимом функционирования системы (без каких-либо рекомендаций). Здесь важной метрикой является показатель Lift, например, во сколько раз увеличилась конверсия в покупку. Различие, разумеется, должно быть статистически значимым.

Поведение пользователя — вещь плохо формализуемая и ни одна метрика в полной мере не опишет мыслительные процессы в его голове при выборе товара.

Здесь, **во-первых**, может помочь анализ поведения на сайте — какие рекомендации срабатывают, а какие нет, где их лучше размещать, какова доля пользователей дошедших до покупки и так далее.

**Во-вторых**, важно получать прямой фидбек от пользователей через опросы, которые позволяют лучше понять особенности своей аудитории: что важнее — релевантность или разнообразие, не раздражают ли повторяющиеся продукты, какова в целом полезность рекомендаций и прочее.

**АЛГОРИТМЫ ФАКТОРИЗАЦИИ** <sup>[116]</sup>

Было бы здорово описать интересы пользователя более «крупными мазками». Не в формате «он любит фильмы X, Y и Z», а в формате «он интересуется молодежными комедиями и боевиками 90-х». В машинном обучении в таких ситуациях рассчитывают сокращенное (латентное / скрытое) описание каждого объекта — компактный числовой вектор, который будет содержать только полезный сигнал и отфильтровать весь малозначимый шум. В случае с рекомендациями часто подобным образом описывают и пользователей, и товары.

В основе алгоритма SVD — разложение матрицы предпочтений  $R$  в произведение 3 матриц:

**$R=U \cdot D \cdot S$**

где размеры матриц  $(k, m) = (k, r) \cdot (r, r) \cdot (r, m)$

а  $r$  — ранг разложения — параметр, характеризующий степень детализации разложения.

В результате, получаем две матрицы сокращенных описаний (факторов):

**U** — компактное описание предпочтений пользователя

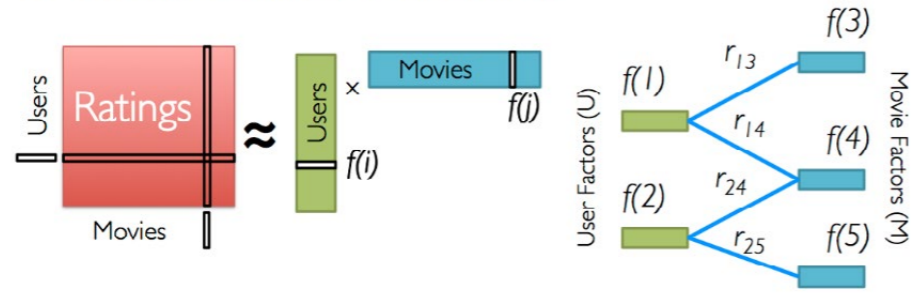
**S** — компактное описание характеристик продукта.

Перемножая обратно рассчитанные факторы, мы получаем очищенную от шума оценку матрицы предпочтений, которую дальше можно легко использовать для генерации новых рекомендаций.

Общее семейство подобных алгоритмов называется NMF (non-negative matrix factorization). Как правило, прямое вычисление таких разложений весьма трудоемко, поэтому на практике часто прибегают к их приближенным итеративным вариантам.

ALS (alternating least squares) — наиболее популярный итеративный алгоритм разложения матрицы предпочтений на произведение 2 матриц: факторов пользователей ( $U$ ) и факторов товаров ( $I$ ). Оптимизация происходит поочередно — сначала по факторам пользователей, потом по факторам товаров. Кроме того, к оптимизируемому функционалу добавляются регуляризационные коэффициенты для получения более устойчивой модели предпочтений.

Low-Rank Matrix Factorization:



Iterate:

$$f[i] = \arg \min_{w \in \mathbb{R}^d} \sum_{j \in Nbrs(i)} (r_{ij} - w^T f[j])^2 + \lambda ||w||_2^2$$

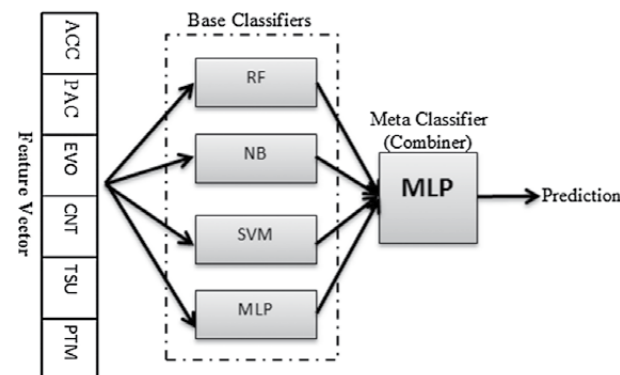
Taken from the BerkeleyX Course Big Data Analysis with Spark

Матрицу предпочтений вполне можно дополнить новыми измерениями (например, с информацией о пользователе или товаре), тогда раскладывать будем уже тензор предпочтений.

На практике редко используется только один подход. Как правило несколько алгоритмов комбинируются в один, чтобы достичь максимального эффекта. Например, с помощью ALS можем посчитать сокращенные представления товаров, и использовать их в качестве дополнительных фичей в content-based рекомендациях.

Часто используют следующие стратегии объединения:

- Weighting — считать средневзвешенный прогноз по нескольким оценкам
- Stacking — предсказания отдельных моделей являются входами другого (мета)классификатора, который обучается правильно взвешивать промежуточные оценки



- Switching — для разных продуктов/пользователей применять различные алгоритмы
- Mixing — вычисляются рекомендации по разным алгоритмам, а потом просто объединяются в один список.

Большим толчком к развитию алгоритмов рекомендаций был проводимый в 2009 году Netflix Prize — конкурс с миллионным призовым фондом по прогнозированию рейтингов фильмов.

Это была задача с явными рейтингами, оценки ставились по шкале от 1 до 5, а точность прогноза оценивалась по RMSE. Большинство первых мест заняли большие ансамбли классификаторов (200+ моделей).

Занявший в итоге первое место ансамбль классификаторов использовал модели следующих классов:

- basic model — регрессионная модель, основанная на средних оценках
- collaborative filtering — коллаборативная фильтрация
- RBM — ограниченные машины Больцмана
- random forests — классическая предиктивная модель

В качестве мета-алгоритма, объединявшего оценки отдельных алгоритмов, использовался стандартный в таких задачах градиентный бустинг.

ОБЗОР ДРУГИХ ПОДХОДОВ

Ассоциативные правила (Association Rules) обычно используются при анализе продуктовых корреляций (Market Basket Analysis) и выглядят примерно так «если в чеке клиента есть молоко, то в 80% случаев там будет и хлеб». То есть если мы видим, что молоко в корзину клиент уже положил, самое время напомнить о хлебе.

Анализ чека — не то же самое, что анализ разнесенных во времени покупок, но если мы будем считать всю историю одной большой корзиной, то вполне можем применить данный принцип и здесь. Это может быть оправдано, когда мы, например, продаем дорогие разовые товары (кредит / полет).

Задача прогнозирования интереса может формулироваться и в виде обычной модели машинного обучения (регрессии, если прогнозируется рейтинг товара или классификатора, если прогнозируется факт интереса):

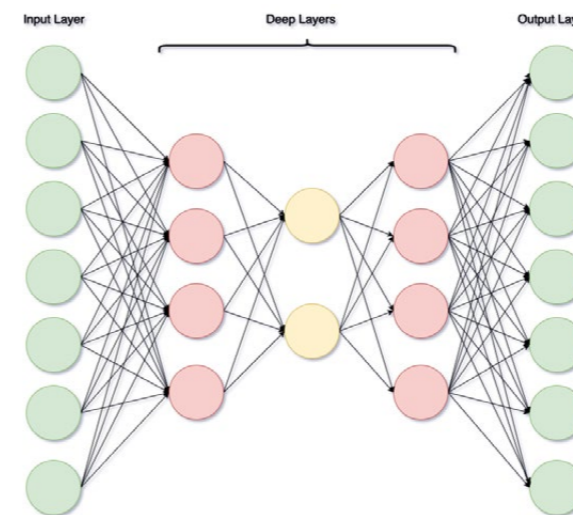
$$y = f(x) = f(\text{xuser}, \text{xitem}, \text{xcontext}).$$

В качестве фичей используется любая информация о пользователе, о товаре, о параметрах их взаимодействия. Обучается модель любым стандартным алгоритмом — линейными моделями, градиентным бустингом, случайным лесом или нейронными сетями.

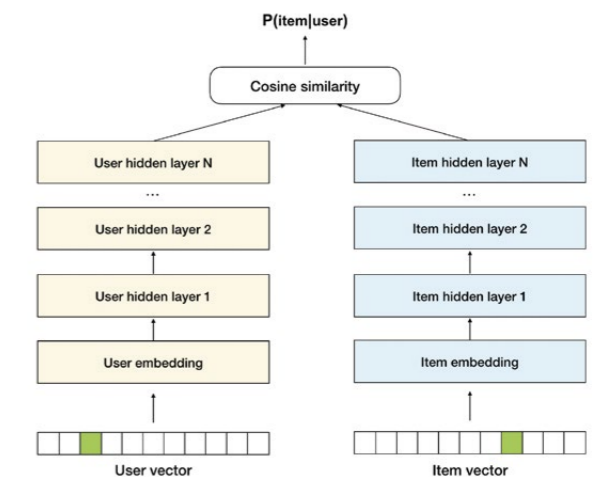
Нейросетевые методы <sup>[122-124]</sup>

**Автоэнкодер (denoising auto-encoder)** — простая сеть, реализующая концепцию бутылочного горлышка — декордер моделирует сокращенное описание интересов пользователя, а декордер восстанавливает их, чтобы минимизировать ошибку. В процессе модель учится убирать из входного сигнала весь незначимый шум (отсюда denoising). Сокращенное описание можно использовать как фичи в другой модели, а можно сразу сгенерировать вектор рекомендаций.

**RBM (restricted Boltzman Machines)** — довольно старый, но похожий подход — здесь также ищутся сокращенные описания интересов, но используется всего два слоя (входной и скрытый), а обучение происходит в результате последовательного сэмпирования по каждому из слоев.

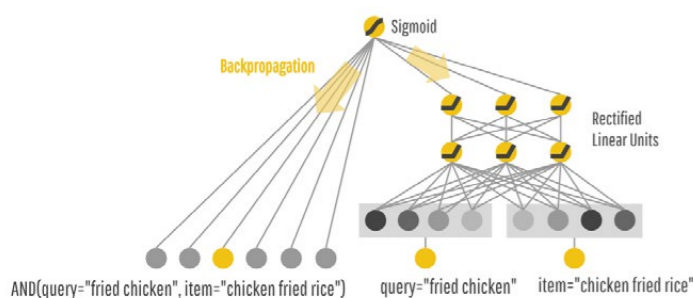


**DSSM <sup>[121]</sup> (deep semantic similarity models)** — один из современных подходов, основанный на нейронных сетях, в центре которого — параллельный расчет эмбедингов пользователя и товара, которые на внешнем слое через косинусное расстояние комбинируются в единую оценку релевантности. Истоки модели находятся в информационном поиске (там была задача матчинга запроса с документами), но принцип был перенесен в задачу сопоставления пользователей и товаров.

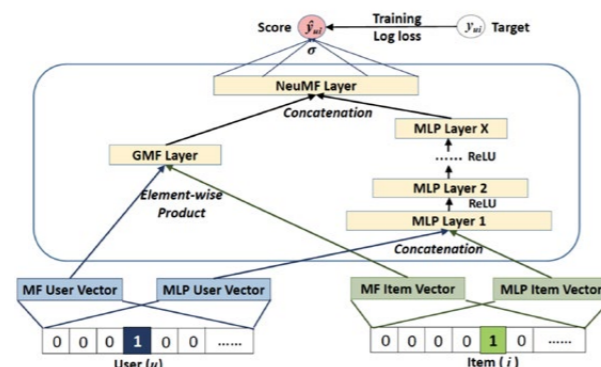




Среди более молодых подходов следует упомянуть модель **Wide-n-Deep от Google** (2016), которая является комбинацией двух сетей: «широкой», отвечающей за запоминание шаблонов встречаемости товаров и «глубокой», отвечающей за понимание структуры интересов пользователя. А также их усложненные модификации: DeepFM (2017) <sup>[117]</sup> и xDeepFM (2018) <sup>[118]</sup>.



Нейронное расширение коллаборативного подхода — алгоритм **Neural Collaborative Filtering** <sup>[119]</sup> (2017). Он также является комбинацией двух сетей — сети матричного разложения (GMF сеть) и глубокой сети получения эмбедингов (MLP сеть). Видна тенденция объединения компонентов в рамках одной сети.



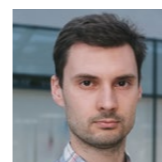
### Графовые подходы

Предметом рекомендаций могут быть, например, люди в социальных сетях. В этом случае применимы алгоритмы класса **Link Prediction**, когда наличие связи в графе предсказывается по характеристикам узлов (в простейшем случае это может быть просто число общих соседей).

В описанных выше методах фильтрации при определении близости между пользователями можно дополнительно использовать социальное расстояние (кратчайший путь в графе) между пользователями.

### РЕЗЮМЕ

Как видно из обзора, постановка задачи генерации рекомендаций очень проста — на основании некоторой выборки рейтингов, необходимо дать оценку релевантности каждого товара каждому пользователю. Вместе с тем, однако, было разработано множество подходов и их модификаций, решающих данную задачу, а современное развитие архитектур нейронных сетей предоставляет широкое поле для экспериментов.



### Артём Просветов

Кандидат физико-математических наук, Руководитель практики анализа данных Центра Компетенции «Большие Данные и Искусственный Интеллект» ГК Ланит, Руководитель практики анализа данных компании CleverDATA, Преподаватель ВШЭ, Ведущий математик ИКИ РАН



### Денис Афанасьев

ИТ-инженер, основатель и генеральный директор компании CleverDATA, Руководитель центра компетенции «Большие Данные и Искусственный Интеллект» ГК Ланит, преподаватель и автор курса «Дата-трансформация предприятия» курса EMBA МГИМО. Автор телеграмм канала CDO Club

# Выбор архитектуры рекомендательной системы

## ВВЕДЕНИЕ

Часто упоминается, что толчком к развитию рекомендательных систем стало соревнование от Netflix с призовым фондом в 1 млн \$. Тем не менее, сами рекомендательные системы на сегодняшний день применяются повсеместно с большим или меньшим успехом и приносят ежегодные прибыли компаниям на сумму, существенно выше изначального призового фонда.

Выработанные подходы для построения рекомендаций хорошо срабатывают в большинстве случаев и позволяют в сжатые сроки получить решение задачи в первом приближении. Чаще всего в джентльменский набор методов входят: коллаборативная фильтрация, ассоциативные правила, матричные разложения. Большинство из перечисленных методов работает с матрицей предпочтений, в которой в столбцах зафиксирована информация о продуктах, в строках — информация о клиентах, а в ячейке на пересечении строки-клиента и столбца-продукта находится информация о реакции пользователя в виде оценки фильма, либо факта покупки.

Существуют методы построения рекомендаций и с помощью нейронных сетей, однако с ними связан ряд сложностей и в 2019 году активно обсуждалась статья «Are we really making much progress? A worrying analysis of recent neural recommendation approaches» <sup>[125]</sup>, ставящая под сомнение преимущества рекомендательных систем с использованием нейронных сетей и показывающая, что хорошо адаптированные под задачу классические методы не только не уступают нейронным сетям по метрикам качества, но и в ряде случаев способны опережать их.

Типичные проблемы, возникающие при использовании классических методов заключаются в том, что для их применения приходится вводить дополнительные предположения и ограничения, например, что вкусы покупателя не меняются со временем. Кроме того, полученные с помощью классических методов рекомендации нередко отличаются тривиальностью и требуют дополнительных усилий для решения проблемы холодного старта (случая, когда приходит новый покупатель и информация о прошлых покупках отсутствует или сильно ограничена). Из-за недостатков классических подходов, продолжают развиваться новые направления в области рекомендательных систем, некоторые из которых будут далее рассмотрены.

**DEEP SEMANTIC SIMILARITY MODEL**

Задачу поиска товара наиболее релевантного прошлой истории покупок возможно сформулировать в форме задачи ранжирования, которая, например, решается в случае поиска релевантного документа по запросу пользователя. Одним из путей решения задачи ранжирования является тренировка сиамской нейронной сети, архитектура которой называется Deep Semantic Similarity Model и состоит из двух частей. Первая ветвь нейронной сети обрабатывает поисковой запрос, либо историю покупок клиента. Вторая ветвь нейронной сети обрабатывает информацию о каждом документе, участвующем в поиске, либо о каждом продукте, участвующем в рекомендациях. В финальном слое нейронная сеть находит косинусное расстояние между тензорами, полученными из двух веток, определяя дистанцию между запросом и документом, либо между историей покупок и следующей покупкой.

Рис. 1: Сиамская нейронная сеть, использующая 2 входа, X соответствует запросу, Y соответствует документу. На финальном слое используется функция, соответствующая нахождению косинусной дистанции между тензорами.

Преимущество DSSM архитектуры заключается не только в том, что нейронная сеть находит скрытые закономерности, или в том, что сеть способна учитывать последовательность купленных продуктов. Дополнительным преимуществом является возможность рассчитать раздельно тензоры из каждой ветви нейронной сети. В процессе поиска рекомендации возможно найти векторные представления и для клиента, и для продукта с помощью двух веток сиамской нейронной сети, а косинусное расстояние между векторами выполнить, не задействуя сеть, параллельным образом, существенно ускоряя процесс вычисления рекомендаций.

Использование архитектуры DSSM упоминается в докладах Яндекса и других компаний.

**ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ**

Обучение с подкреплением (Reinforcement Learning) завоевало повышенный интерес исследователей и энтузиастов. Чаще всего задача в обучении с подкреплением формализуется через понятия Среды, агента и Награды. Агент должен придерживаться оптимальной стратегии для того, чтобы получить максимум награды, а поиск оптимальной стратегии является оптимизационной задачей для модели. Примером может служить любая игра из сборника ATARI: игрок берет на себя роль агента, игра задает правила Среды, действия игрока приводят в оптимальном варианте к награде.

В рамках обучения с подкреплением на сегодняшний день разработано множество методов, однако чаще всего начинают свое знакомство с методов Q-learning, SARSA, DQN, Asynchronous Advantage Actor-Critic Algorithm, Soft Actor-Critic, Monte Carlo Tree Search.

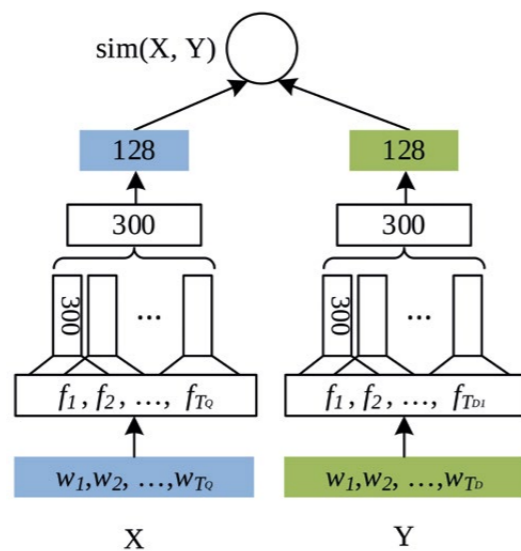
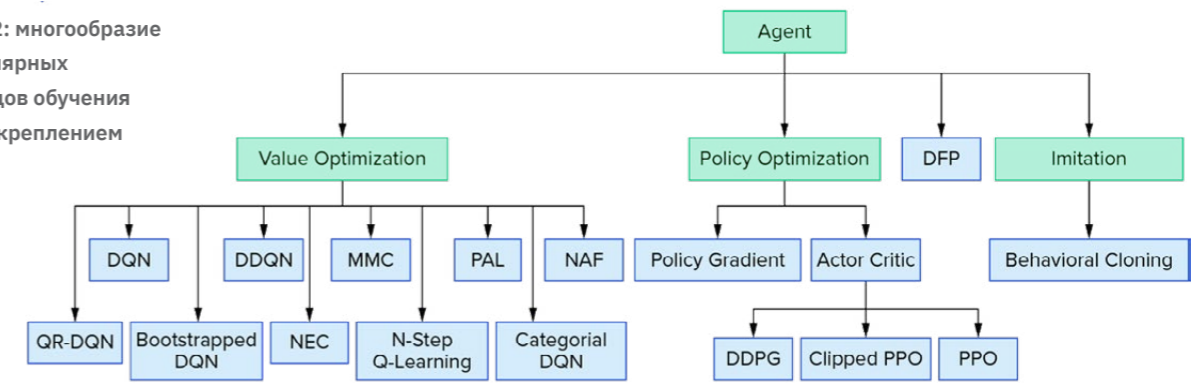


Рис. 2: многообразие популярных методов обучения с подкреплением



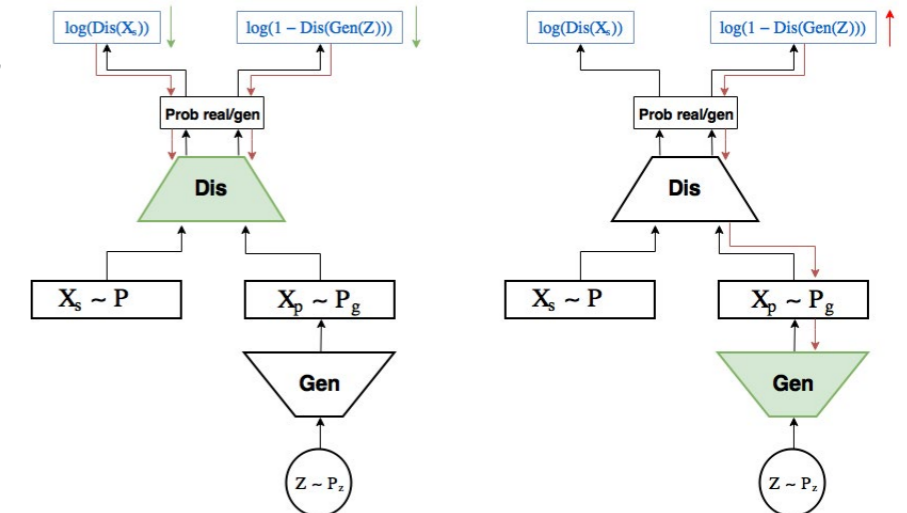
Задачу построения рекомендаций также возможно сформулировать в рамках задачи обучения с подкреплением: выбор продукта для рекомендации является задачей по оптимизации стратегии, а наградой является прибыль компании. Одной из сложностей в тренировке модели в рамках режима обучения с подкреплением является колоссальное количество проб и ошибок, необходимое для поиска оптимальной стратегии. В некоторых случаях и для рекомендательных систем есть возможность без существенных потерь провести большое количество экспериментов с предложением, например в случае показа баннерной рекламы.

По мере развития подходов обучения с подкреплением появляются эксперименты по их использованию в рекомендательных моделях. На сегодняшний день известны успешные случаи применения обучения с подкреплением в сервисах Сбербанка и группы компаний X5.

**СОРЕВНОВАТЕЛЬНЫЕ СЕТИ**

Соревновательные сети были предложены в 2017 году и получили активное развитие в области обработки изображений. В отличие от нейронных архитектур обычных, в соревновательной сети участвуют две нейронные сети, называемые Генератором и Дискриминатором. Процесс обучения происходит посредством последовательного объединения Генератора и Дискриминатора таким образом, чтобы Генератор создавал новые объекты, а Дискриминатор пытался распознать, являются ли поступающие на вход объекты реальными или синтетическими, созданными Генератором. Часто обучение происходит путем поочередного обновления весов. На конечном этапе обучения Генератор способен создавать максимально реалистичные объекты, которые Дискриминатор не способен отличить от реальных. После достижения неотличимости сгенерированных и реальных объектов, Генератор может быть использован для создания новых фотографий, либо модификаций уже имеющихся изображений с поразительной реалистичностью.

Рис. 3: процесс пошагового обучения соревновательной сети, на первом шаге обновляются параметры Дискриминатора, на втором шаге обновляются параметры Генератора.



Не смотря на активное использование в развлекательных целях, архитектура соревновательной сети дает возможность научить Генератор воспроизводить сложные и порой не формализуемые распределения, наблюдаемые в объектах реального мира. Например, нейронная сеть CosmoGAN была натренирована генерировать двумерное распределение темной материи, неотличимое от реального.

Интересным вариантом последующего использования Дискриминатора является его применение для поиска аномалий, что востребовано в задачах предиктивной аналитики, раннего обнаружения неисправностей и оптимизации профилактических работ.

Если же сформулировать задачу построения рекомендаций в рамках архитектуры соревновательной сети, то Генератор может тренироваться генерировать следующий выбор клиента на основе истории предыдущих решений, в то время, как Дискриминатор должен тренироваться отличать реальный выбор клиента от синтетического, предложенного Генератором. После тренировки такой соревновательной сети, возможно использование Генератора для создание новых рекомендаций, которые Дискриминатор не в силах отличить от реального последующего выбора пользователя.

Преимущество использования соревновательных сетей для построения рекомендаций заключается в том, что проблему Холодного старта можно изящно обойти, а для тренировки модели нет необходимости создавать тренировочную выборку с искусственными отрицательными примерами. На сегодняшний день известны успешные эксперименты с применением рекомендательной системы на базе GAN в одной из авиакомпаний.

**ГРАФЫ**

В ряде случаев структура связей объектов представляет собой граф, например общение пользователей в социальных сетях.

У графовой структуры есть ряд преимуществ: например, в ней возможно найти распределение узлов по степени важности, или вычислить минимальное остовное дерево, выделяющее основных распространителей информации.

Отдельной возможностью, представляющей интерес, является распространение рекомендаций по сети связей, основанной на идее, что если многим вашим друзьям и знакомым подходит определенная рекомендация, то есть высокая вероятность, что она подойдет и вам. К графовой структуре возможно применение специальных методов построения рекомендаций, использующих информацию о связях между элементами. Простым примером может быть использование случайных блужданий по графу (Random Walk): для каждого узла графа строится последовательность случайных переходов, вероятность которых определяется коэффициентами связи между узлами, а также учитывается константная вероятность продолжить путь из начального узла. Полученная таким образом последовательность блужданий для каждого узла позволяет учесть на только все напрямую связанные объекты, но и те объекты, с которыми есть опосредованная связь через цепочку других объектов. Дальнейшая обработка последовательности случайных блужданий может быть выполнена с помощью хорошо себя зарекомендовавших методов работы с текстами, ведь полученная последовательность узлов представляет из себя набор слов, последовательность которых определяется связями между элементами графа.

Другим путем учесть структуру графа модели является использование специальной архитектуры нейронных сетей под названием Graph Neural Network (GNN) или Graph Convolution Networks (GCN), предложенные в 2016 г. Интересно, что архитектура GNN в процессе обучения с математической точки зрения оказывается очень схожей с тренировкой архитектуры Transformer, широко применяемым в задачах обработки текста и дающей State Of The Art результаты на сегодняшний день. Таким образом, исследователи отмечают, что в ряде задач GNN дают более качественное решение по сравнению с использованием классических подходов.

Дальнейшее активное развитие графовых нейронных сетей дает дополнительный импульс к развитию рекомендательных систем в тех областях, доменная структура которых имеет графовое представление. Использование рекомендаций на графах практикуется в компании Одноклассники. Кроме того, графовые методы хорошо себя показывают в области фрод-анализа банковских переводов, также представляющих собой сложную сетевую структуру.

**ИНТЕРПРЕТАЦИЯ МОДЕЛЕЙ**

Как и для большинства моделей машинного обучения, для рекомендательной системы возможность интерпретации причин вычисленного результата является желательным условием при использовании на практике. В случае применения нетривиальных подходов к построению рекомендаций, требуется дополнительная работа по извлечению интерпретации каждой рекомендации. В направлении получения интерпретации результатов рекомендательной системы идут активные исследования и эксперименты.

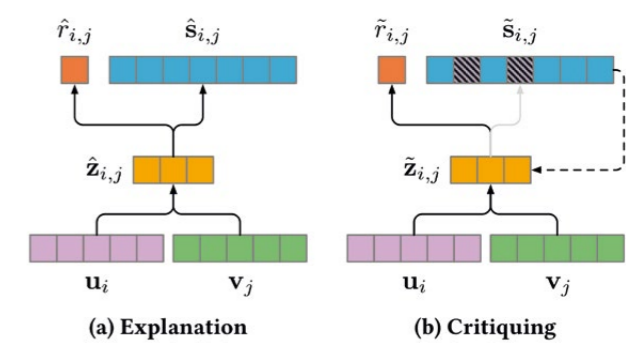


Рисунок 5: (из статьи Deep language-based critiquing for recommender systems, 2019): добавление возможности для интерпретации рекомендаций (а), а также коррекция рекомендации на основе критики (б)

Следующим шагом к улучшению рекомендаций является получение обратной связи от клиентов и использование её в автоматическом режиме для коррекции рекомендаций. Реализация подхода с обратной связью описывается в статье «Deep language-based critiquing for recommender systems» [126]: пользователи сервиса получали рекомендацию, а также получали возможность исключить лишнее ключевое слово в тэгах продукта, после чего рекомендация уточнялась. Для реализации описанной выше возможности исследователям пришлось усложнить архитектуру нейронной сети, введя дополнительный тензор, ответственный за интерпретируемые тэги. В результате, удалось получить следующие результаты для 3-х случайных клиентов — таблица 1.

Таким образом, удастся сделать рекомендательную систему более «человечной», а упоминание причин найденной рекомендации при ее показе несет дополнительную аргументацию и придает убедительности, что благотворно влияет на конечного получателя и повышает конверсию в целевое действие.

Рис. 4: (из статьи Thomas Kipf 2016) с иллюстрацией работы сверточной графовой нейронной сети

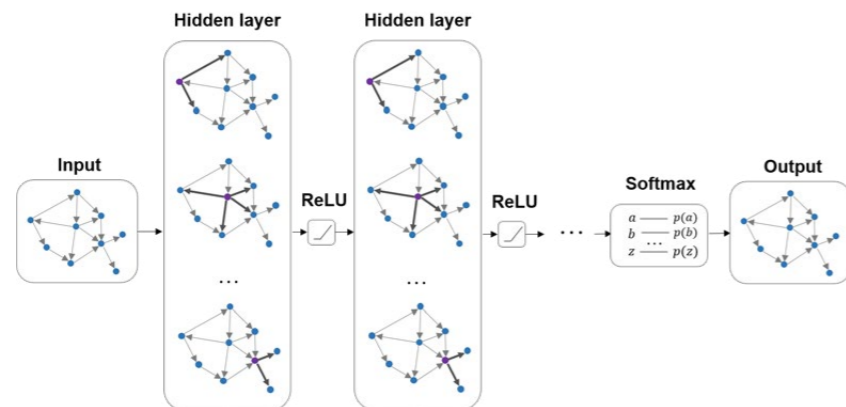


Таблица 1

Рекомендация напитка	Объяснение	Исключаемый тэг	Улучшенная рекомендация напитка	Объяснение
Aecht Schlenkerla Rauchbier Urbock	Smoke Bready Brown	Smoke	Piraat Ale	Sweet Fruit Gold
Ø lfabrikken Porter	Smooth Chocolate Black	Black	Brooklyn Brown Ale	Smooth Brown Sweet
10 Commandments	Honey Sugar Sweet	Sweet	Heady Topper	Fruit Grapefruit Orange

## МЕТОДЫ РАСПРЕДЕЛЕННОГО ОБУЧЕНИЯ РЕКОМЕНДАТЕЛЬНЫХ МОДЕЛЕЙ

По мере того, как все больше компаний осваивают методы разработки рекомендательных моделей, они все плотнее встраиваются в бизнес-процессы. Например, мы уже не можем представить себе ни один онлайн сервис, будь то интернет магазин, или площадка интернет обучения, без персонализации на основе рекомендаций продуктов, на основе опыта компаний по работе с нами, как с потребителем их продукта. Наши ожидания от уровня сервиса растут, мы, как пользователя, требуем все более качественного сервиса, основной которого по сути и является рекомендательная система. Поэтому компании все активнее вкладывают усилия и средства в совершенствование своих алгоритмов и подходов.

Но кроме самого алгоритма, в вопросах разработки эффективных рекомендательных моделей нельзя забывать и вклад данных. Сложно сказать однозначно, чей вклад в конечный результат больше — совершенного алгоритма или данных, на которых он выполняется. Однозначно можно только утверждать, что без наличия обоих компонентов результат невозможен.

Поэтому в последние годы, кроме совершенствования алгоритмов усиливается «гонка за данными» между разными компаниями. Как часто упоминаются в прессе, даже процессы приобретения одной компании другой, часто продиктованы получением доступа к новым данным, которые могут сделать рекомендательные системы лучше и получить рыночное конкурентное преимущество. Особенно часто такие упоминания были в 2015-17 годах и ярким примером можно назвать приобретение компанией Facebook компании WhatsApp за \$19 млрд, в то время как компания имела довольно сомнительные перспективы монетизации и выхода на прибыль [127]. Это приобретение позволило Facebook значительно увеличить качество модели рекомендации друзей.

Тем не менее, тренд «погони за данными», особенно так называемыми «пользовательскими данными», в последние годы начинает компенсироваться изменениями законодательства и вопросами этического характера. Не вдаваясь в детали, которые находятся за рамками темы данной статьи, можно кон-

статировать, что централизация данных разных компаний в едином хранилище становится все более и более сложной задачей в силу не столько технических, сколько организационных барьеров.

Как ответ индустрии на данные тенденции становится набирающие популярность методы построения децентрализованных систем для обеспечения возможности распределенного обучения рекомендательных моделей. Несмотря на то, что анализ данных с сохранением конфиденциальности изучается уже более 50 лет, только в последнее десятилетие решения начали получать широкое распространение.

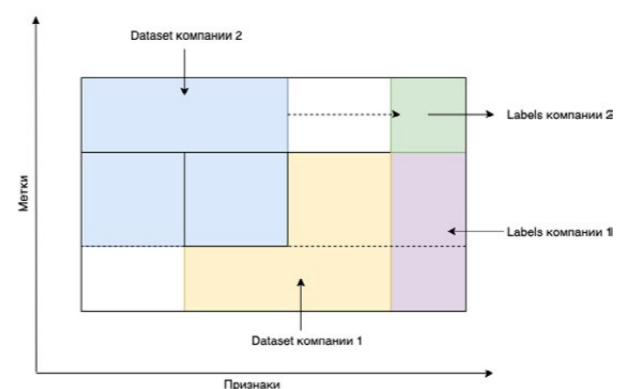
Одним из таких методов, который с 2017 года активно применяется такими компаниями как Google и Apple, является метод под названием Federated Learning и был описан в работе Н Brendan McMahan Communication-efficient learning of deep networks from decentralized data [128].

Эти компании применяют его для совершенствования рекомендательной модели, которая подсказывает слова при наборе текста, а так же своих голосовых помощников. Поскольку эти модели оперируют довольно «чувствительными» данными — текстами сообщения, которые набирают пользователи, и словами, которые произносятся в области слышимости микрофонов мобильных телефонов, то при их централизации возникали бы следующие сложности:

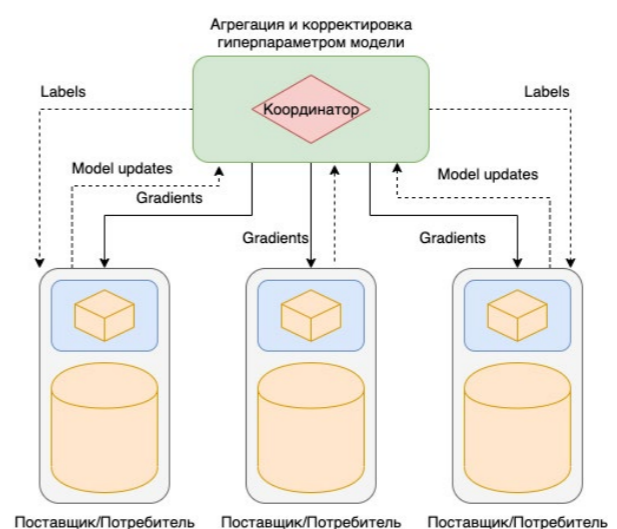
- мобильные телефоны используются по всему миру, в то время как в каждой отдельной стране сейчас разные требования по работе с данными и их защите. Многие страны (в том числе Россия) требуют хранения и обработки данных граждан на аппаратном обеспечении, которые находятся на территории страны;
- при локализации данные в едином хранилище создается единая точка взлома и утечки, что повышает риски сохранности пользовательских данных;
- технически, идея копирования больших массивов информации с миллионов устройств также является достаточно «дорогой» для реализации.

Метод распределенного обучения Federated Learning позволяет решить задачу обучения рекомендательной модели и избежать данных сложностей [129-130].

Суть метода достаточно проста: представим 2 компании, которые обладают 2-мя массивами данных. Данные можно представить в 2-х мерном пространстве как «набор переменных» и «набор меток» (см рисунок) ниже.



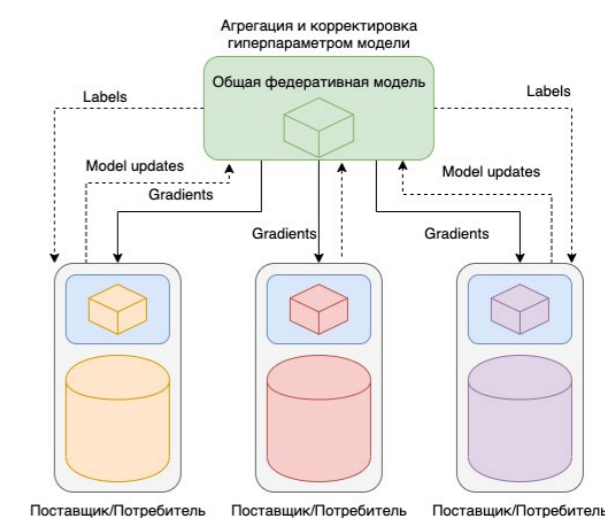
Данные хранятся на серверах каждой компании и задача состоит в том, чтобы натренировать две (и более, в случае большего количества участников) копии модели «в периметре» каждой компании. Для обеспечения процесса обучения в системе присутствует координатор, общий для всех участников, который отвечает за процесс обучения и занимается агрегированием значений гиперпараметров модели на каждой итерации обучения, вычислением градиента в передаче градиента в каждую копию модели.



Таким образом, в процесс обучения не происходит передача данных за периметр компании ни в каком виде, осуществляется только передача гиперпараметров, а в результате цикла обучения у каждой компании образуется идентичная остальным копия рекомендательной модели, которая натренирована на всей совокупности данных участников децентрализованной системы.

Различают 2 типа метода Federated Learning: в случае, когда пространство признаков у двух (и более) компаний совпадает (например в случае, если мы говорим о 2-х банках, или одном банке, который ведет свою деятельность в разных странах), но различается пространство меток (банки обладают непересекающимися клиентскими базами), то говорят о, так называемом, горизонтальном федеративном обучении. В случае, если две (и более) компании оперируют данными в различных пространствах переменных, но обладающих пересекающимися пространствами меток (например банк и ритейл компания), говорят о методе вертикального федеративного обучения.

В случае вертикального федеративного обучения схема работы несколько меняется, единая модель получается одна, на уровне координатора, тк на разных данных происходит обучение «частей» этой федеративной модели:



Отдельной темой в этом процессе стоит задача обеспечения безопасности самого процесса обучения модели. В то время, как применение данного метода нивелирует риск утечки исходных данных, он поднимает вопросы утечки самой модели, а также вопросы компрометации источника данных для обучения. Тут отдельно надо применять методы шифрования и доверительного обмена ключами, но их описание выходит за рамки данной статьи и с ними можно ознакомиться в первоисточниках.

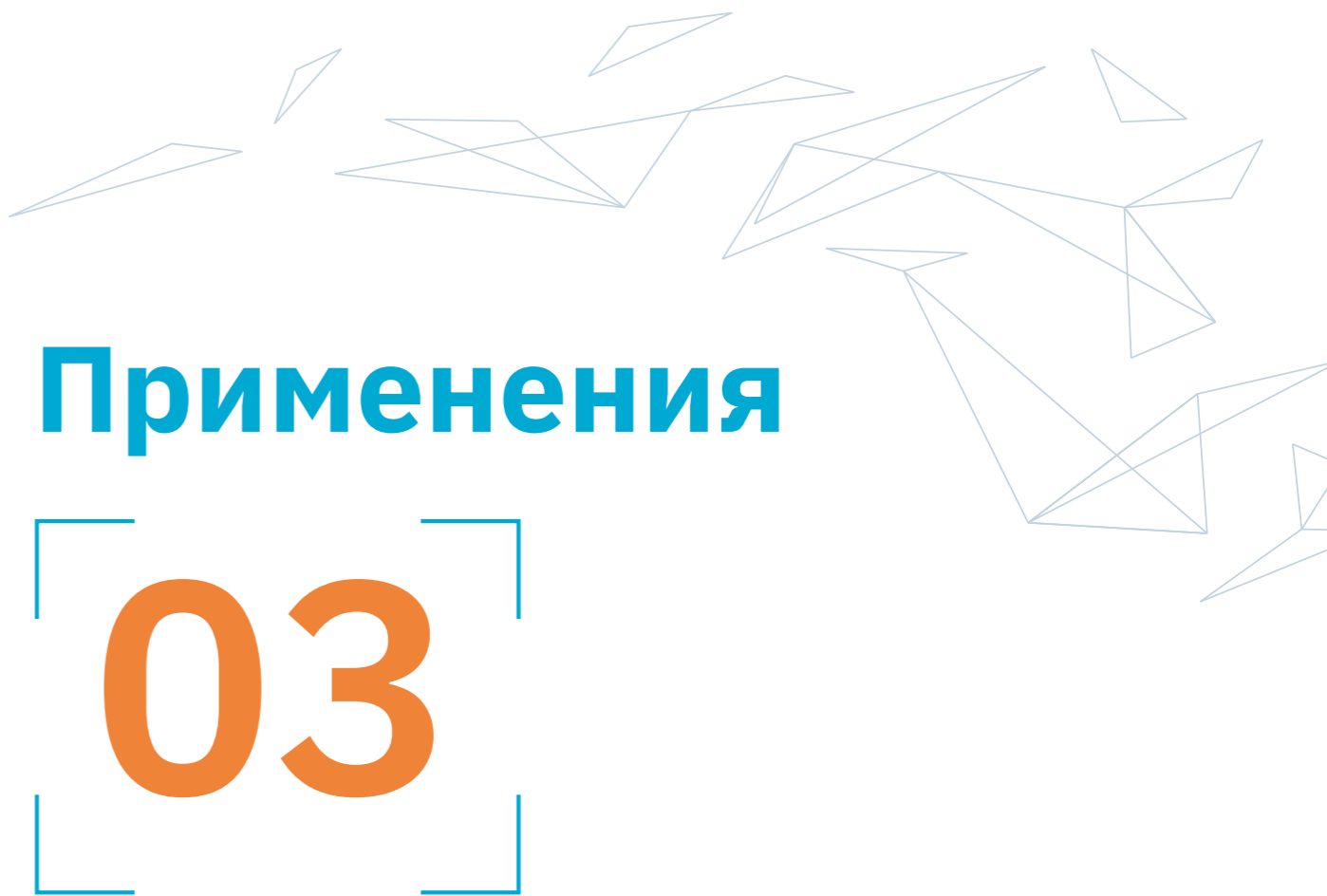
Так же нельзя не отметить и недостатки данного метода. Решая вопрос с передачей и централизацией данных, он поднимает вопросы необходимости синхронизации алгоритмов расчета переменных и синхронизации меток (в случае вертикального обучения). Так же необходимо решать вопросы надежности сети и доступности всех участвующих в обучении узлов.

Несмотря на эти недостатки данный метод активно развивается и поддерживается рядом популярных библиотек машинного обучения, таких как TensorFlow Federated, Federated AI Technology Enabler, PySyft, Leaf, PaddleFL and Clara Training Framework.

В заключении необходимо отметить, что Federated Learning подходит не только для обучения рекомендательных моделей, но и для решения задач определения мошенничества, систем постановки медицинских диагнозов и тд.

## ЗАКЛЮЧЕНИЕ

Большинство специалистов в области Data Science сходятся во мнении, что процесс разработки модели предиктивной аналитики лучше всего строить от простых моделей к более сложным. В таком эволюционном подходе есть возможность проследить, дают ли новые модели прирост качества и оправданы ли затраты на внедрение более сложных архитектур в промышленное производство. Описанные выше методы и подходы показывают впечатляющие результаты в частных случаях, однако No Free Lunch теорема утверждает, что нет универсальной модели, доминирующей во всех задачах. Отсутствие универсальной модели приводит к тому, что архитектуру решения необходимо подбирать в каждом частном случае в зависимости от качества данных, ограничений, объемов данных и нагрузки на сервис. Тем не менее, как показывает опыт множества успешных компаний, внедривших методы предиктивной аналитики, игра стоит свеч.





# Big Data в современном ритейле: предиктивные технологии для роста Retention и LTV

Как меняется ландшафт современного ритейла и почему магазины, до сих пор не использующие возможности больших данных и умных алгоритмов вскоре могут лишиться конкурентоспособности? [189-192]

Все понимают, что технологии Big Data внедрять необходимо, но зачастую представление об их применении довольно расплывчатое. Мы подготовили подробный материал на основе исследований и лучших российских и зарубежных кейсов.

## ИЗМЕНЕНИЕ ЛАНДШАФТА ECOMMERCE И ТРАДИЦИОННОГО РИТЕЙЛА

До появления технологий персонализации, маркетологи определяли потребности клиентов исходя из опросов и анализа продаж. Однако, как выяснилось, такой подход даёт результаты мало сопоставимые с реальностью.

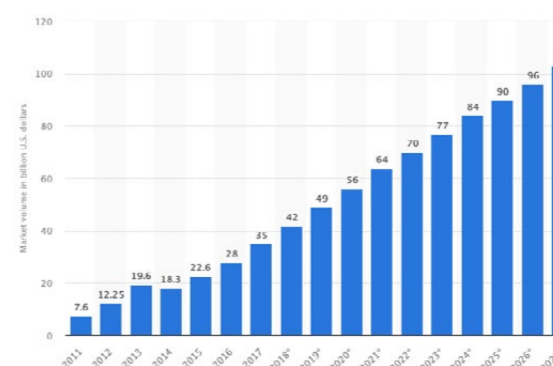
В 2018 году H&M терпит падение прибыли на протяжении 10 кварталов подряд, что стало угрозой для закрытия компании. Для стабилизации положения были использованы умные алгоритмы, позволившие убрать 40% ассортимента магазина не снизив продажи. Этот эксперимент показал, что ритейлеры не всегда знают, что действительно нужно их клиентам.

Ритейлеры обладают огромным количеством данных, которые можно анализировать и использовать как для коммуникации с клиентами, так и для оптимизации различных процессов внутри компании.

Раньше в сети Walmart работали около тысячи специалистов для анализа 24 тысяч запросов в час. Сейчас для этого используются технологии Big Data, и таким образом современный ритейл переходит от CRM-маркетинга к предиктивным технологиям.

### Немного цифр:

- Стоимость индустрии Big Data достигнет \$77 миллиардов к 2023 году.
- Компании, использующие Big Data отметили увеличение выручки на 8% при снижении общих затрат на 10%.
- Около 50% компаний говорят о том, что использование Big Data коренным образом изменили курс действий отделов маркетинга и продаж.



Прогноз роста рынка Big Data, основанный на выручке с 2011 до 2027 гг. (в млн долларов):

- 79% ритейлеров считают, что отказ от Big Data приведет к потере конкурентоспособности и закрытию бизнеса.

## ГЛАВНЫЕ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ BIG DATA В РИТЕЙЛЕ

Одна из важнейших возможностей Big Data — возвращение коммуникации 1:1. До появления масс-маркета и интернет-магазинов продавцы знали своих клиентов в лицо и могли советовать товары, исходя из их предпочтений. Впервые за долгое время, персонализированное обслуживание позволяет индивидуально настраивать коммуникации с покупателями. Чем больше данных удастся собрать, тем точнее рекомендации, а значит выше конверсия и прибыль.

### Главные возможности применения Big Data в ритейле:

- формирование ассортиментной матрицы и оптимизация товарных остатков
- распределение товаров между и внутри торговых точек (мерчендайзинг)
- предсказание спроса
- анализ данных программ лояльности
- ценообразование, в том числе динамическое
- оптимизация маркетинговых и рекламных кампаний
- персонализация коммуникации с пользователями во всех каналах

## КАК УВЕЛИЧИТЬ RETENTION, LTV И ЛОЯЛЬНОСТЬ КЛИЕНТОВ С ПОМОЩЬЮ ПРЕДИКТИВНЫХ ТЕХНОЛОГИЙ

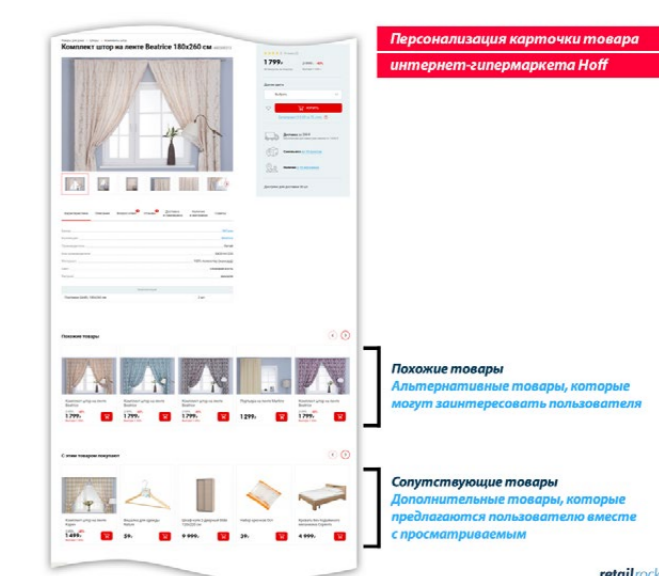
А теперь о самом главном: как использовать предиктивные технологии на практике? Мы хотим поделиться собственными разработками, успешно показавшими себя на российском рынке. Вы можете адаптировать их под свою компанию или брать как шаблон для создания собственных уникальных стратегий.

## ПЕРСОНАЛЬНЫЕ РЕКОМЕНДАЦИИ В РЕЖИМЕ REAL-TIME

Современные системы предиктивного маркетинга оценивают поведение пользователя, историю покупок и его интересы в режиме реального времени без непосредственного участия со стороны технических специалистов. Это открывает для ритейлеров широкий спектр возможностей персонализации сервиса.

Вы можете персонализировать обслуживание на любом этапе: начиная от сайта, заканчивая регулярными и триггерными рассылками. Чтобы убедиться в эффективности принятых решений, всегда проводите тщательное тестирование.

Для гипермаркета Hoff мы искали наилучшую конфигурацию блоков в карточке товара. Среди 4-х сегментов 2 показали почти нулевые приросты. В то же время выигравший сегмент дал значительный прирост среднего чека и входящего оборота на 5,8%:



## ПРЕДИКТИВНЫЙ МАРКЕТИНГ В ТРИГГЕРНОЙ КОММУНИКАЦИИ

Интернет-магазины получают огромный массив данных о своих покупателях, на основе которого можно строить цепочки потребления. Каждая транзакция пользователя генерирует несколько цепочек, и появляется расчетная вероятность, какие покупки он совершит.

Механизм предсказания следующей покупки состоит из нескольких этапов:

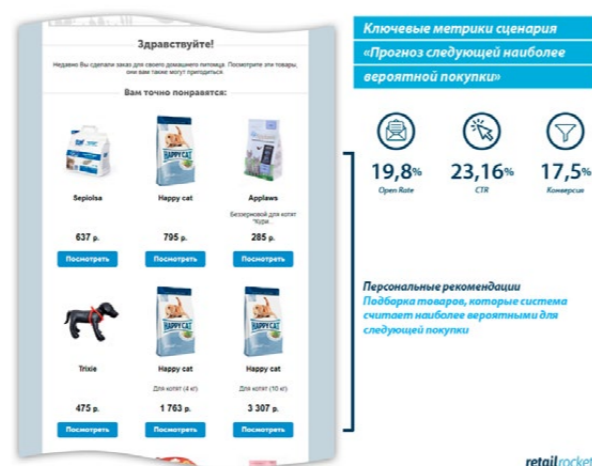
- Анализ последовательностей покупок всех клиентов
- Выявление статистически значимых цепочек потребления
- Прогнозирование совершения покупки в следующем «звене» цепочки потребления, после оформления заказа

Цепочки строятся для всех товарных категорий. Например, вот реальная цепочка потребления одного из магазинов товаров для детей:

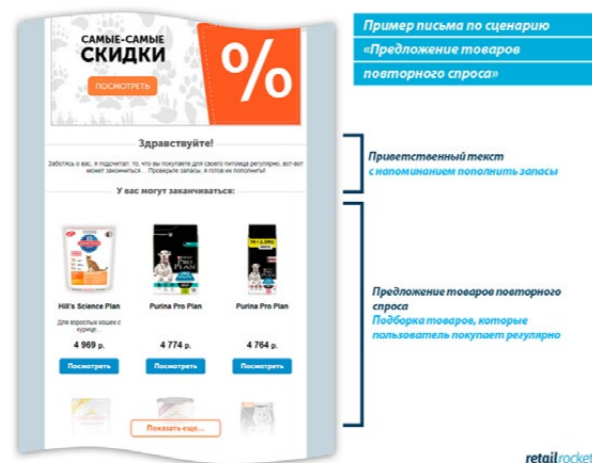


Предиктивный анализ потребностей клиентов.  
Реальный пример цепочек потребления

Пользователь может попадать сразу в несколько цепочек потребления, поэтому мы используем сложный механизм группировки предложений, выявляющий, что именно необходимо человеку. На основе этой информации можно составить несколько интересных кампаний. Например, персонализированная подборка в письме «Прогноз следующей наиболее вероятной покупки» (Next Best Offer) предлагает товары по интересам и предыдущим заказам:



Похожий прием используется в сценарии «Предложение товаров повторного спроса». Алгоритм учитывает срок потребления купленных товаров и, когда необходимо пополнить запасы, отправляет клиенту письмо:



## ИСПОЛЬЗОВАНИЕ ПРЕДПОЧТЕНИЙ В РЕГУЛЯРНЫХ РАССЫЛКАХ И АВТОМАТИЗАЦИЯ ЭТОГО ПРОЦЕССА

Покупатели оставляют множество информации, которую можно использовать в оригинальных кампаниях. К примеру, если клиент когда-либо совершал покупку одежды, то магазин знает его размер. Это можно использовать уже на этапе формирования предложения товара. В нашем случае — в email-рассылках. Причем можно как исключать товары, которых нет в нужном размере, так и просто выделять подходящие размеры.

Чтобы акцентировать внимание на товаре, можно выделить подходящий размер. Также можно указать смежные размеры (на один больше и меньше):



Если у вас нет данных для персонализации, то указывайте размеры, которые есть в наличии. У большинства пользователей средние параметры, поэтому информация в письме будет с наибольшей вероятностью будет актуальна.



Сейчас или никогда — это главный девиз компаний, которые ещё не внедрили высокие технологии в обслуживание. Можно надеяться на исключительность и лояльных клиентов, но, к сожалению, даже в случае с такими гигантами как H&M это может не сработать. Big Data работает на вас и подчеркивает уникальность магазина.

## КАК СТАНДАРТНЫЕ РЕКОМЕНДАЦИИ МОГУТ СНИЖАТЬ ПРОДАЖИ МАГАЗИНА

Рекомендательные блоки уже долгое время помогают интернет-магазинам повышать выручку и конверсию. «Новинки», «Хиты продаж», «Сопутствующие товары» — эти механики своего рода классика e-commerce. Но в последние годы почти все исследователи говорят о стремлении пользователей к персонализированному обслуживанию.

По информации исследовательской компании Gartner, интернет-магазины, игнорирующие эту тенденцию, к 2020 году будут зарабатывать на 15% меньше тех, кто внедряет персонализированный сервис.

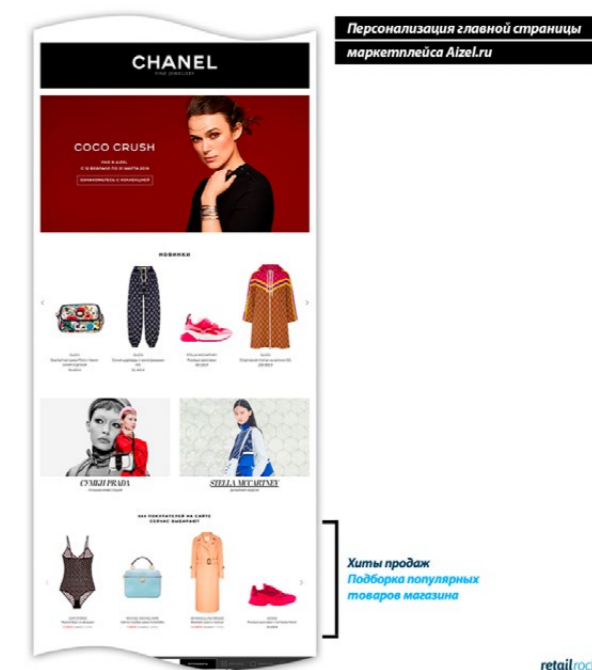
## ГЛАВНАЯ СТРАНИЦА AIZEL.RU: СНИЖЕНИЕ ВЫРУЧКИ НА 22% ИЗ-ЗА БЛОКА «ХИТЫ ПРОДАЖ»

Именно сотрудничество с маркетплейсом Aizel.ru натолкнуло нас на написание этой статьи. Раньше персональные рекомендации выступали скорее как альтернатива стандартным механикам, но теперь они необходимы многим интернет-магазинам.

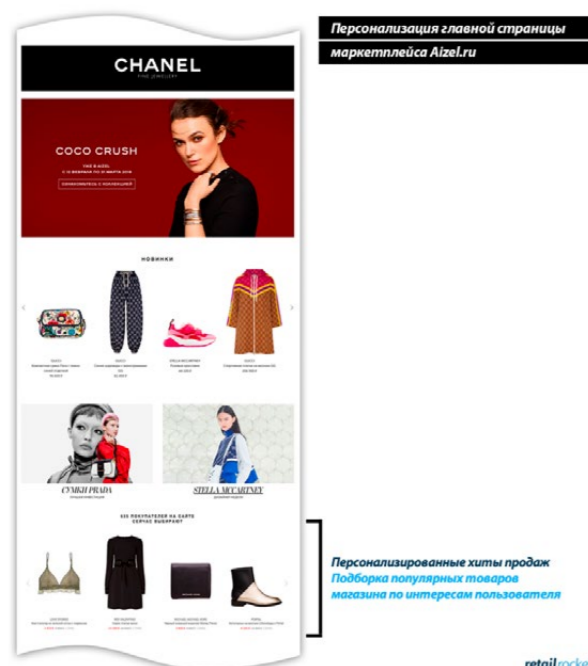
Например, на главной странице Aizel.ru мы тестировали два блока рекомендаций — персонализированные и стандартные «Хиты продаж». Первый блок показал рост выручки на 3,8%, а второй — падение на 22,4%!

Исследование эффективности проводилось с использованием механики A/B-тестирования, при которой все посетители сайта случайным образом делились на равных 3 сегмента.

Первому сегменту показывались хиты продаж:



Второму сегменту показывались персонализированные хиты продаж:



Третьему сегменту рекомендации не показывались. Он выступал в качестве контрольной группы.

## Результаты

По итогам тестирования были получены следующие результаты:

Сегмент	Прирост конверсии	Изменение среднего чека	Оценка увеличения выручки
Хиты продаж	-7,71%	-15,93%	-22,42%
Персонализированные хиты продаж	+10,07%	-5,91%	+3,57%
Контрольная группа	—	—	—

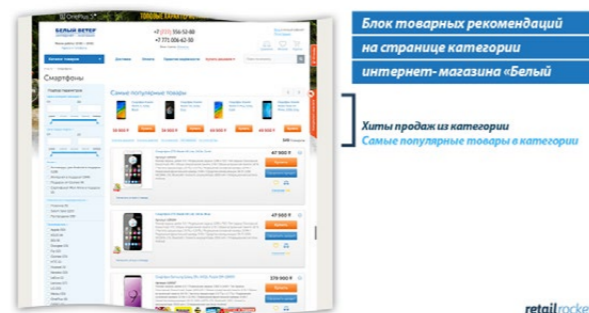
## СТРАНИЦА КАТЕГОРИИ КАЗАХСТАНСКОГО ИНТЕРНЕТ-МАГАЗИНА БЕЛЫЙ ВЕТЕР: СНИЖЕНИЕ СРЕДНЕГО ЧЕКА НА 2,6% ИЗ-ЗА БЛОКА «ХИТЫ ПРОДАЖ»

Страница категории очень важна и от нее во многом зависит как поиск и выбор товара пойдет дальше. Например, неудобная навигация или отсутствие релевантных позиций могут стать причиной ухода клиента без шанса на возврат. Вероятнее всего, не найдя ничего интересного, пользователь просто решит уйти к конкуренту, надеясь на более подходящий ассортимент.

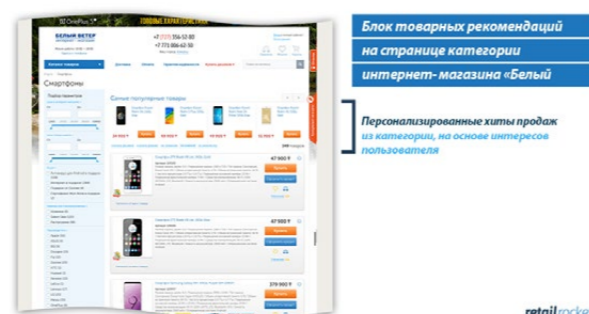
Возможно, так и происходило в казахстанском интернет-магазине электроники Белый Ветер. Здесь, также как и в кейсе выше, стандартные «Хиты продаж» показали снижение выручки и среднего чека на 1,4% и 2,6% соответственно, в то время, как механика «Персонализированные хиты продаж» повысила эти метрики на 16,5% и 9,6%.

Для оптимизации рекомендательной системы на сайте интернет-магазина «Белый Ветер» было проведено А/В-тестирование на странице категории, где все посетители сайта случайным образом делились на 4 сегмента:

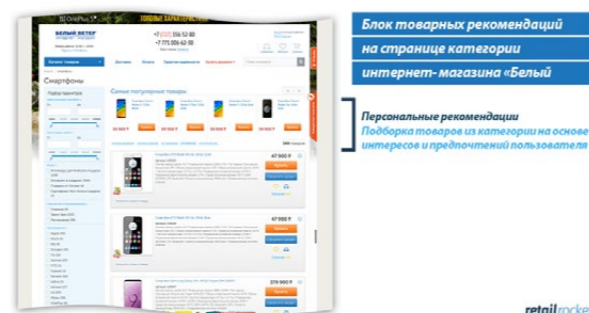
Первому сегменту показывались хиты продаж из категории:



Второму сегменту показывались хиты продаж из категории, персонализированные с учетом интереса пользователя:



Третьему сегменту показывались персональные рекомендации товаров из категории:



Четвертый сегмент был контрольной группой: рекомендации пользователям не демонстрировались

## Результаты

Сегмент	Прирост конверсии	Изменение среднего чека	Оценка увеличения выручки
Хиты продаж из категории	+1,23%	-2,61%	-1,44%
Персонализированные хиты продаж из категории	+6,33%	+9,61%	+16,54%
Персональные рекомендации из категории	+0,62%	+7,95%	+8,61%
Контрольная группа	—	—	—

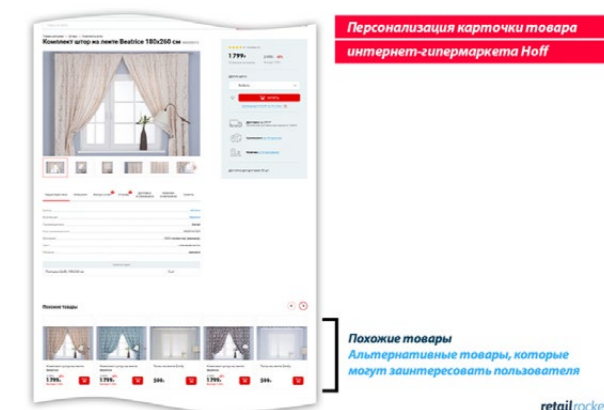
## КАРТОЧКА ТОВАРА ГИПЕРМАРКЕТА HOFF: СОПУТСТВУЮЩИЕ ТОВАРЫ В ОДИНОЧКУ ДАЮТ ОТРИЦАТЕЛЬНЫЙ ПРИРОСТ ВЫРУЧКИ НА 4,3%

Сопутствующие товары — тема, к которой нет равнодушных. Существует много стереотипов о «правильных» рекомендациях, например, что в карточке товара нужны исключительно блоки сопутствующих товаров. Но не бывает двух одинаковых магазинов и таких же одинаковых решений, подходящих для всех.

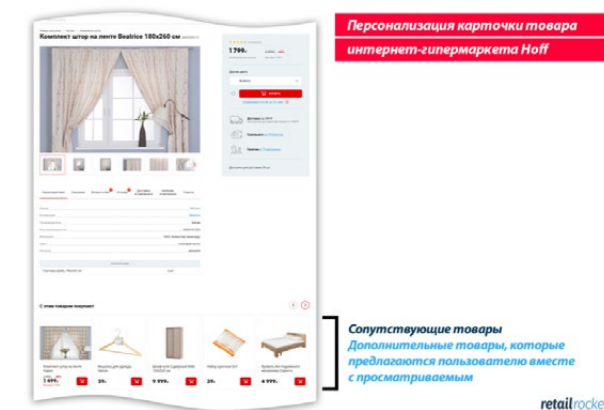
Среди последних кейсов с неоднозначными результатами в карточке товара был интернет-гипермаркет Hoff. Здесь блок сопутствующих товаров показал отрицательный результат. Но в сочетании с похожими товарами картина совершенно другая. Стоит обратить внимание на то, что в одиночку блок похожих товаров дал прирост примерно в два раза ниже (за исключением конверсии), чем вместе с сопутствующими товарами. Однако, если поменять блоки местами, мы снова получаем результаты на грани фола.

Как всегда для проверки эффективности использовалась методика А/В-тестирования, при которой пользователи случайным образом делились на 5 сегментов:

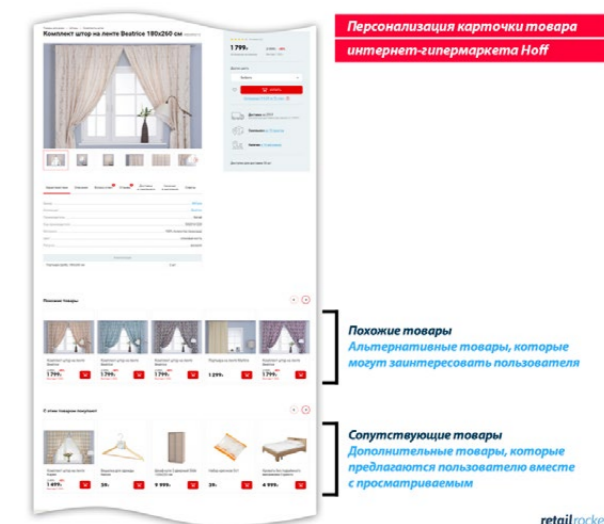
Первому сегменту показывались похожие товары:



Второму сегменту показывались сопутствующие товары:

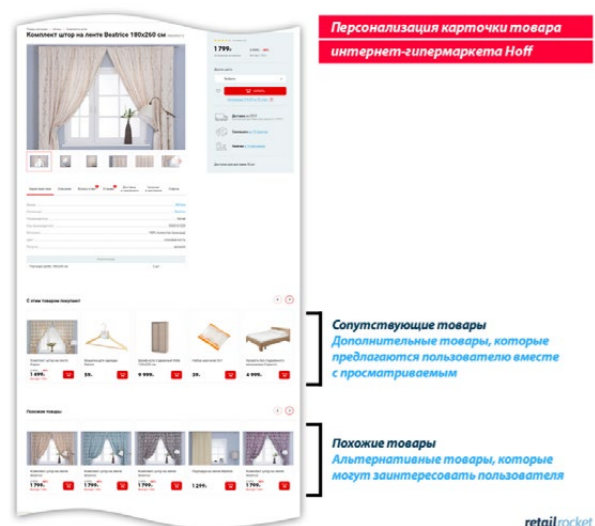


Третьему сегменту показывались два блока одновременно: похожие товары (сверху) и сопутствующие товары (снизу):





Четвертому сегменту показывались также два блока одновременно, но в обратном порядке. Сопутствующие товары располагались сверху, а похожие снизу:



Пятый сегмент был контрольной группой, рекомендации не показывались.

## Результаты

По итогам тестирования были получены следующие результаты:

Сегмент	Прирост конверсии	Изменение среднего чека	Рост входящего оборота
Похожие товары	+1,82%	+2,05%	+3,91%
Сопутствующие товары	-2,04%	-2,29%	-4,29%
Два блока: похожие товары и сопутствующие товары	-0,40%	*6,26%	+5,84%
Два блока: сопутствующие товары и похожие товары	-1,56%	1,55%	-0,03%
Контрольная группа	—	-	—

## ПОИСКОВАЯ СТРАНИЦА BOOK24: СТАНДАРТНЫЕ РЕКОМЕНДАЦИИ СНИЖАЮТ ВЫРУЧКУ НА 7%

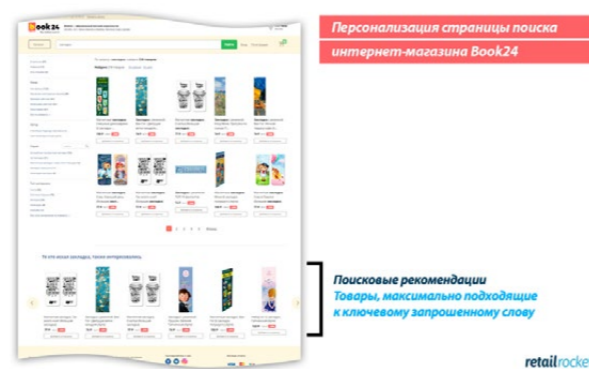
В этом кейсе мы поговорим о рекомендациях на странице поиска, о которых часто забывают. А ведь именно здесь многие посетители сайта начинают путь по сайту. Стандартные поисковые рекомендации основываются на ключевых словах, а приоритет в выдаче остаётся за самими

популярными позициями. Но так ли это эффективно, как кажется?

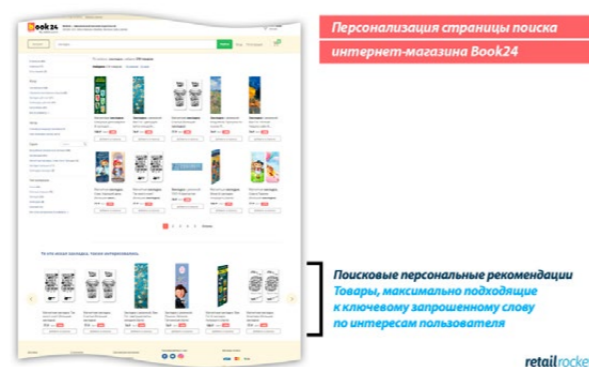
Возможно, пользователь уже видел блок «Хиты продаж» на главной, поэтому те же товары в поиске могут его не устроить, вплоть до ухода с сайта. Так вся поисковая страница начинает работать в минус для ритейлера. Эту проблему исправят персонализированные товары, которые будут учитывать и текстовый запрос пользователя, и его интересы.

Здесь мы также использовали механику A/B-тестирования. Для того, чтобы найти оптимальный алгоритм для страницы поиска, пользователи случайным образом делились на 3 сегмента:

Первому показывались стандартные поисковые рекомендации:



Второму сегменту показывались персональные поисковые рекомендации:



Третий сегмент был контрольной группой, рекомендации пользователям в этом сегменте не показывались.

## Результаты

Сегмент	Прирост конверсии	Изменение среднего чека	Оценка увеличения выручки
Стандартные поисковые рекомендации	+1,65%	-7,52%	-6,00%
Персональные поисковые рекомендации	+6,46%	*0,77%	+7,28%
Контрольная группа	—	-	—

## ЗАКЛЮЧЕНИЕ

Каждый интернет-магазин уникален и требует особого подхода. Поэтому всегда стоит проводить тесты эффективности и не надеяться на общепринятые решения. Особенно сейчас, когда пользователи нуждаются в персональном подходе.

Для того, чтобы правильно оценить эффективность того или иного решения, лучше всего использовать методику A/B-тестирования — она дает возможность получить наиболее честные результаты. Любую гипотезу можно проверить с помощью цифр, но важно делать это по определенным правилам, поскольку проведение A/B-теста — это очень сложный процесс с множеством подводных камней. Если A/B-тестирование проводится с ошибкой, то бизнес гарантированно принимает неверное решение и получает скрытые убытки.

Какая из конфигураций окажется выигрышной в вашем интернет-магазине? Знают только ваши пользователи. Тестируйте самые разные алгоритмы, даже те, которые на первый взгляд кажутся проигрышными. Иногда неочевидное решение оказывается лучшим.



**Волкова Наталия Борисовна,**

MSc in Operational Communications/ICT,  
Руководитель проектов Аналитического управления  
ДИТ г. Москвы

# Предсказательная аналитика и системы поддержки принятия решений: применения для умного города

## ИНИЦИАТИВЫ ООН В ОБЛАСТИ «УМНЫХ» ГОРОДОВ

Повестка «умных» городов далеко не нова — она возникла более сорока лет назад в Калифорнии. В 1974 году Бюро анализа населения (департамент мэрии) выпустило отчёт «Состояние города: кластерный анализ Лос-Анджелеса», содержащий предложения по созданию общегородской информационно-аналитической системы для управления мегаполисом «на научной основе»<sup>[131]</sup>. А ранее, ещё в конце 60-х годов, Бюро предприняло первую и довольно успешную попытку применения математических и программных моделей для решения насущных задач мегаполиса: борьбы с преступностью, снижения аварийности на дорогах, помощи малообеспеченным слоям населения, вывода из эксплуатации ветхого жилья — всего было выделено 16 проблемных областей<sup>[132]</sup>.

Однако на международном уровне тема «умных» устойчивых городов утвердилась значительно позже, когда в 2012 году Европейская экономическая комиссия ООН включила это направле-

ние в программу работы. Два года спустя стартовал проект ООН «Объединённые умные города» (United Smart Cities, USC), целью которого была поддержка устойчивого развития городов с фокусом на обеспечение доступа к инновационным технологиям<sup>[133]</sup>.

При выполнении проекта USC были проведены многочисленные консультации с технологическими компаниями, академическим сообществом, правительственными организациями, опрошено свыше 300 отраслевых экспертов, проанализировано 116 определений «умного» города<sup>[134]</sup>. В результате ООН и МСЭ\* подготовили ряд руководящих документов и сформировали единое определение: «Умный устойчивый город это инновационный город, использующий информационно-коммуникационные технологии (ИКТ) и другие средства для повышения уровня жизни, эффективности деятельности и услуг в городах, а также конкурентоспособности при обеспечении удовлетворения потребностей настоящего и будущих поколений в экономических, социальных, культурных и природоохранных аспектах»<sup>[133]</sup>.

Среди наиболее значимых технологий для «умных» городов в отчётах ООН-МСЭ названы большие данные, открытые данные и средства их анонимизации, технологии обеспечения безопасности данных, а также предсказательная аналитика и системы поддержки принятия решений<sup>[134]</sup>.

В 2019 году проект USC был интегрирован в инициативу «Объединение усилий в целях построения “умных” устойчивых городов» (United for Smart Sustainable Cities, U4SSC), в рамках которой продолжена поддержка проектов с применением передовых технологий обработки и анализа данных для достижения целей ООН в области устойчивого развития (ЦУР).

## ЛУЧШИЕ ПРАКТИКИ «УМНЫХ» МЕГАПОЛИСОВ МИРА

Какие же мировые мегаполисы считаются самыми «умными»?

Многочисленные рейтинги — Smart City Index, Cities in Motion, Innovation Cities Index, Sustainable Cities Index — не дают однозначного ответа на этот вопрос<sup>[135-138]</sup>. Зачастую удивительным образом на первых местах оказываются города, в которых размещаются организации-составители рейтингов.

Изучение методологии рейтингов показывает, что наличие платформ больших городских данных и аналитических систем, как правило, не является критерием отбора. Вместе с тем именно управление на основе данных (data-driven) позволяет городам занимать топовые позиции. Применение инструментов анализа и прогнозирования учитывается неявным образом, через влияние на экологические, социальные, транспортные и другие аспекты городской среды.

Рассмотрим наиболее интересные примеры, как старожилы рейтингов и новые игроки решают городские задачи с помощью предсказательной аналитики и машинного обучения. Мониторинг подобных проектов проводит ООН в рамках инициативы U4SSC «City Science Application Framework»<sup>[139]</sup>.

## Южная Калифорния: платформа EnviroSuite для контроля качества воздуха

В некоторых округах Калифорнии наблюдается высокий уровень загрязнения воздуха, вызванный автомобильным трафиком, промышленными выбросами, лесными пожарами и другими причинами. Контроль качества воздуха регулируется на региональном уровне различными службами по 35 локальным зонам. В области залива Сан-Франциско работает агентство BAAQMD (Bay Area Air Quality Management District), в ряде округов Южной Калифорнии — SCAQMD (South Coast Air Quality Management District). Оба агентства используют цифровую платформу EnviroSuite<sup>[140]</sup>.

До внедрения платформы данные с сенсоров проходили длительную обработку до передачи ответственным службам. Такой подход препятствовал оперативному реагированию на инциденты, связанные с ухудшением качества воздуха. В ряде случаев к тому времени, когда причины и источники выбросов были найдены, время уже было упущено.

Для решения этой проблемы был развёрнут комплекс EnviroSuite, который в режиме реального времени строит базовый сценарий распространения вредных веществ в атмосфере на основе набора данных о качестве воздуха, скорости и направлении ветра, уровне промышленных выбросов и с учётом прогноза погоды. Набор данных формируется по показаниям датчиков метеорологических станций и станций контроля загрязнения атмосферы, подключённых к EnviroSuite по каналам беспроводной связи. Платформа позволяет выявлять потенциальные источники загрязнения воздуха в течение нескольких секунд после возникновения инцидента.



Рис. 1. Платформа EnviroSuite прогнозирования загрязнения воздуха

\* Международный союз электросвязи (МСЭ) является специализированным учреждением Организации Объединённых Наций, отвечающим за информационно-коммуникационные технологии (ИКТ), и организацией по разработке международных стандартов.

\* Лос - Анджелес, Ориндж, Риверсайд, Сан - Бернардино

В системе применяются алгоритмы\*, основанные на математических моделях процессов рассеивания вредных веществ в атмосфере, и методы машинного обучения. EnviroSuite развёрнута в облачной инфраструктуре Amazon Web Services.

В результате внедрения системы органы контроля качества воздуха Калифорнии получили возможность оперативно реагировать на инциденты и жалобы жителей, разрабатывать эффективные планы по улучшению экологической ситуации и перехода к более устойчивой модели регионального развития. Прогнозируется, что улучшение качества воздуха окажет положительное влияние на здоровье граждан и увеличит продолжительность жизни. В настоящее время проект находится на стадии реализации, поэтому итоговые результаты будут получены в ближайшие годы.

#### Дубай: система Happiness Meter для измерения уровня счастья

Дубай стремится стать самым счастливым городом на планете. И это не утопические мечты, не маниловщина, а конкретная цель стратегии построения «умного» города [141].

Стоит отметить, что ОАЭ — единственное государство в мире, в котором есть министерство по вопросам счастья.

При этом есть также и понимание того, что счастье — понятие субъективное, а повышение уровня счастья в масштабах города является задачей, сложноконвертируемой в традиционные мероприятия государственного планирования.

Нестандартные цели требуют нестандартных подходов. И власти Дубая решили провести целевые замеры уровня удовлетворённости населения и бизнеса по всем основным аспектам городской жизни, включая образование, здравоохранение, транспорт, экономику, энергетику, экологию.

Опросы проводились с помощью так называемого «Измерителя счастья» (Happiness Meter) — мобильного приложения для смартфонов, в котором можно оценить услуги государственного или частного сектора по простой шкале: счастлив/несчастлив/нейтрален. В дополнение жите-



Рис. 2. Результаты измерения уровня счастья в 2015-2018 гг

лям и гостям Дубая предлагалось оставить отзыв, идеи по совершенствованию работы и критические замечания. Happiness Meter был установлен также на информационные киоски и добавлен на городские веб-ресурсы.

В течение нескольких лет было собрано множество данных, касающихся повседневной жизни людей: поездок на работу, получения медицинских рецептов, зачисления в школу, покупки и аренды недвижимости, открытия бизнеса. Для понимания корреляции оценок с параметрами оказанной услуги результаты были обработаны с применением статистического анализа и методов машинного обучения.

С момента запуска в 2015 году и до конца 2018 года\*\* через Happiness Meter было собрано свыше 22 млн оценок уровня счастья и более 650 тыс. отзывов, что достаточно много для эмирата с населением 4 млн человек [142].

По сравнению с социологическими исследованиями и анализом тональности (sentiment analysis) социальных сетей, которые также используются в городской аналитике, платформа и мобильное приложение «измерителя счастья» стали для властей Дубая уникальным инструментом предоставления оперативной информации в различных разрезах. Аналитическая платформа помогает властям и частному сектору проводить политику непрерывного повышения качества оказываемых услуг, концентрироваться на проблемных областях, улучшать социальное обслуживание населения, вовлекать жителей в управление городом, повышать прозрачность деятельности властей. Отмечается, что использование данного инструментария способствует

формированию здоровой конкурентной среды (компании соперничают за высокие оценки со стороны потребителей) и более быстрому распространению «лучших практик».

#### По итогам 2018 года индекс счастья в Дубаи достиг 90%

(хотя по смыслу этот показатель всё же отражает уровень удовлетворённости жизнью в городе).

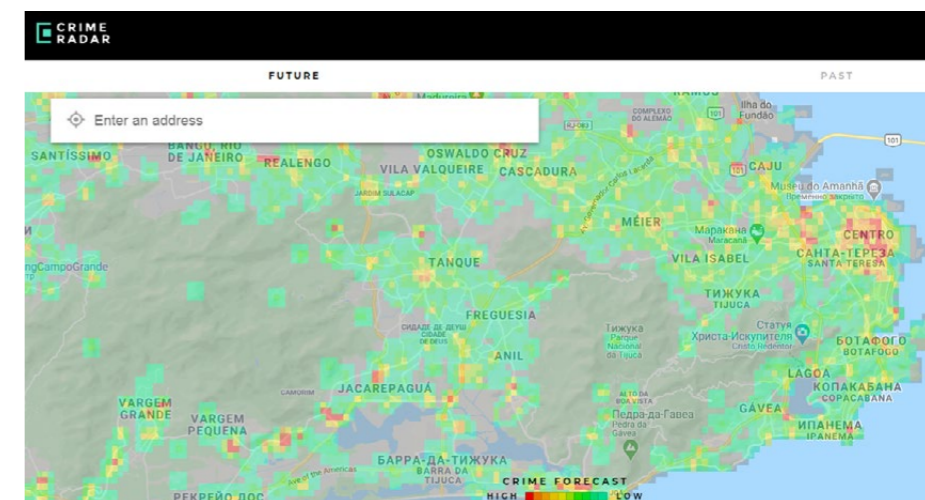
#### Рио-де-Жанейро: CrimeRadar для прогнозирования криминогенной обстановки [143, 144]

В преддверии летних Олимпийских Игр в Рио-де-Жанейро мэрия запустила цифровую платформу и приложение CrimeRadar. Местные жители и туристы смогли на тепловой карте города увидеть прогноз криминогенной обстановки по районам.

Возможность предсказания места и времени преступлений звучит фантастически и напоминает сюжет блокбастера Спилберга «Особое мнение». А между тем это именно то направление предиктивной аналитики, которое развивают правительства государств Северной Америки, Западной Европы и некоторых азиатских стран.

Высокий уровень преступности всегда был острой проблемой Рио. Правоохранительные органы апробируют новые методы выявления очагов криминогенности, повышения эффективности распределения ресурсов и составления оптимальных маршрутов полицейского патрулирования.

Рис. 3. Радар криминогенности



CrimeRadar — цифровая платформа ситуационного анализа и прогнозирования вероятности совершения преступления — использует методы машинного обучения и байесовские статистические методы. Обучение модели проводилось на ретроспективных полицейских данных за пять лет, предоставленных государственным институтом общественной безопасности Бразилии.

В качестве идейного вдохновителя проекта выступил аналитический центр Рио-де-Жанейро по вопросам безопасности Instituto Igarapé, который занимался подбором компаний-разработчиков платформы и приложения, подготовкой наборов данных для обучения ML-моделей и тестированием моделей.

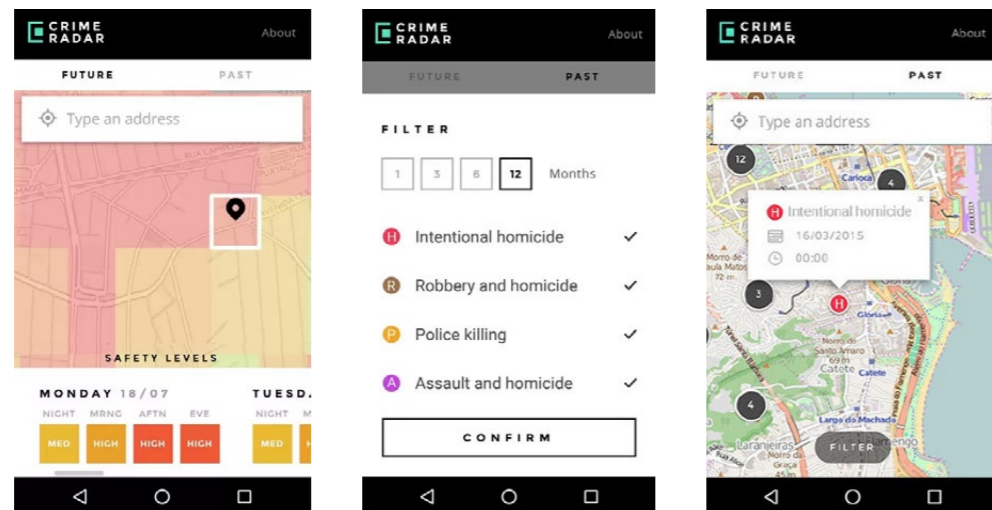
CrimeRadar позволяет визуализировать на картографической основе криминогенную обстановку в различных районах города Рио-де-Жанейро. Вероятность совершения преступления отображается по секторам 250x250 метров по десятибалльной шкале, где 1-самая безопасная зона (зелёная), а 10 — самая опасная (красная). Прогноз на будущее рассчитывается по дням недели и времени суток. В отличие от аналогичных систем в других странах, тепловая карта криминогенности Рио доступна не только силам правопорядка, но и жителям города.

Интересной особенностью проекта стало включение в лицензию на ПО платформы положения об ответственности, в том числе возмещение гражданам ущерба, вызванного неверным прогнозом. Лицензия содержала отсылку к этическому кодексу по использованию искусственного интеллекта, подготовленному организацией

\* CALPUFF — California Puff Model, CALMET — Computer Aid ed Learning in Meteorology, WRF - Weather Research and Forecasting.

\*\* За 2019–2020 гг. статистика отсутствует.

Рис. 4. Мобильное приложение CrimeRadar



FAT/ML (Fairness, Accountability, and Transparency in Machine Learning) — сообщество исследователей, занимающихся вопросами справедливости, подотчетности и прозрачности в машинном обучении [145]. Однако впоследствии на карте криминогенности был размещён дисклеймер с рекомендацией не полагаться полностью на предиктивные алгоритмы при принятии решений по вопросам персональной безопасности. В тексте дисклеймера указано, что модель в настоящее время не обновляется и точность её снижается.

По информации института Igararé, внедрение платформы CrimeRadar способствует сдерживанию роста преступности, повышает эффективность работы полиции и улучшает туристический имидж Рио-де-Жанейро. Вместе с тем статические исследования о результатах использования этой модели отсутствуют, а рандомизированные контролируемые испытания аналогичных моделей прогнозирования преступности дают неоднозначные результаты. Критики системы также указывают на то, что информация о «зелёных» зонах на карте даёт жителям ложное ощущение безопасности и может использоваться криминальными элементами для совершения преступлений.

## МОСКВА: РЕАЛИЗОВАННЫЕ И ПЕРСПЕКТИВНЫЕ ПРОЕКТЫ

В общей сложности в Москве в промышленной эксплуатации и в разработке находится свыше двух десятков проектов в области предсказательной аналитики и поддержки принятия решений на основе машинного обучения. Рассмотрим некоторые из них.

### 1. Поддержка принятия врачебных решений в ЕРИС

В январе 2020 года на XVIII Ассамблее «Здоровая Москва» Департамент здравоохранения города анонсировал начало эксперимента по внедрению компьютерного зрения для анализа медицинских изображений. Цель эксперимента — исследование возможности использования в системе здравоохранения Москвы методов поддержки принятия врачебных решений с применением инновационных технологий [146-149].

Технологической платформой эксперимента является единый радиологический информационный сервис (ЕРИС), на котором размещены наборы данных для обучения моделей распознавания рака лёгкого, рака молочной железы, туберкулёза и других патологий [150]. Разработчикам предоставляется доступ к базе изображений по различным видам медицинской визуализации (Рис. 5).

Рис. 5. Портал эксперимента mosmed.ai

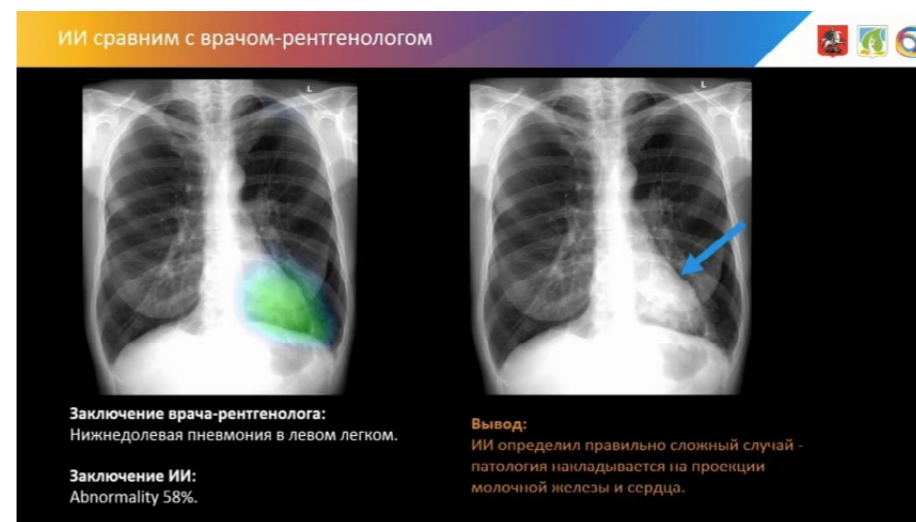
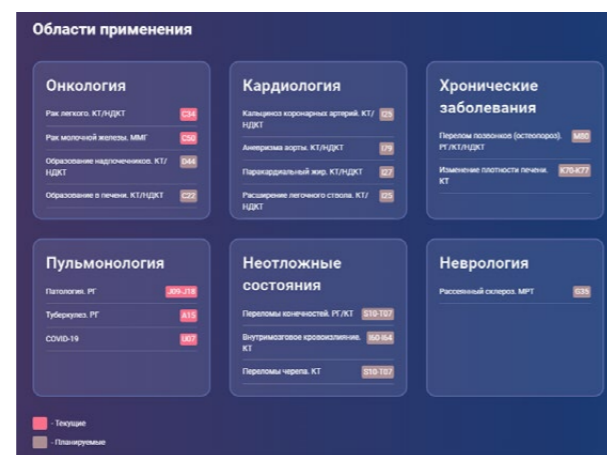


Рис. 6. Презентация к докладу проф. Морозова С.П. на круглом столе «Искусственный интеллект в здравоохранении» [147]

Регистрация участников осуществляется на официальном портале эксперимента mosmed.ai, там же размещены документация и регламент проведения конкурса.

Клинические испытания систем компьютерного зрения проводятся в соответствии с методическими рекомендациями, разработанными Научно-практическим клиническим центром диагностики и телемедицинских технологий (НПКЦ ДиТ ДЗМ) [150].

Участникам эксперимента выдаются гранты, размер которых зависит от количества проанализированных медицинских изображений [148].

Во время проведения эксперимента Москва столкнулась с пандемией COVID-19. Встала первоочередная задача создать сервис, позволяющий автоматически находить на КТ-снимках признаки пневмонии коронавирусной этиологии. В рекордные сроки команда НПКЦ ДиТ ДЗМ под руководством Сергея Морозова сформировала и опубликовала уникальный набор данных «MosMedData: результаты исследований компьютерной томографии органов грудной клетки с признаками COVID-19», который стал крупнейшим в мире. В настоящее время датасет содержит 1100 исследований, которые были проведены в более 80 лечебных учреждениях Москвы, подключённых к ЕРИС. Исследования разделены на 5 категорий в зависимости от объёма патологических изменений лёгких [151, 152].

Уже сейчас эксперимент показал позитивные результаты. С начала эпидемии в московских цен-

трах компьютерной томографии было проведено более 200 тысяч \* исследований лёгких, и при анализе большинства из них использовались технологии машинного обучения.

Сервис на основе искусственного интеллекта помечает на КТ-снимках зоны возможного поражения лёгких и формирует предварительное заключение о вероятности наличия патологии, что позволяет сократить время проведения исследования, повысить производительность труда врачей-экспертов, снизить объём рутинных действий и в конечном итоге улучшить качество диагностики. При этом постановка окончательного диагноза всегда остаётся за врачом.

### 2. Рекомендательная система в общегородском контакт-центре [153]

Традиционно при рассмотрении рекомендательных систем в «умных» городах в первую очередь подразумеваются инструменты в ГИС, направленные на поддержку принятия решений государственными служащими. Между тем городские рекомендательные системы нужны в первую очередь жителям, в дополнение к персональным ассистентам на их смартфонах.

С 2016 года в московском общегородском контакт-центре работает виртуальный оператор — система распознавания и синтеза речи на основе партнёрских решений. Оператор консультирует по 30 темам: оплата коммунальных услуг, готовность документов, местонахождение эвакуированного автомобиля и процедура его возврата, адрес ближайшего центра «Мои документы»

\* По состоянию на 01.09.2020 <https://www.mos.ru/news/item/76266073/>.

и другие. За 2019 год робот обработал 8,5 млн звонков. Обучающая база знаний системы распознавания и синтеза речи в контакт-центре в настоящее время содержит более 8,5 тыс. статей.

Виртуальный оператор снизил нагрузку на операторов контакт-центра, взяв на себя ответы на типовые обращения граждан, сократил время ожидания на линии при массовых сезонных обращениях. Целями и задачами на перспективу является обучение робота ведению полноценного диалога, возможности понимать эмоциональное состояние собеседника и «подстраиваться» под него, а также прогнозировать вопросы горожан.

В 2019 году на конкурсе «Хрустальная Гарнитура» виртуальный оператор Москвы одержал победу в номинациях «Лучшая практика обслуживания клиентов без участия оператора» и стал лауреатом в номинации «Лучшее применение автоматизации, роботизации и искусственного интеллекта».

### 3. Прогнозные модели для школ

В Москве также проходят апробацию системы с применением предиктивной аналитики в образовании. Амбициозной общей целью этих проектов является улучшение результативности обучения школьников и повышение вероятности поступления в вузы с помощью сервиса персональных рекомендаций по дополнительной подготовке.

«Цифровой след» школьника представляет собой наборы структурированных и неструктурированных данных, в первую очередь это оценки по различным предметам, посещаемость уроков и дополнительных занятий, результаты экзаменов, выбор кружков, участие в олимпиадах и занятые места, рейтинг школы — всего более 100 признаков. На ретроспективных данных 800 учебных заведений Москвы происходит обучение моделей\*, которые позволяют прогнозировать результаты выпускных экзаменов и автоматически выработать рекомендации по улучшению успеваемости школьника, в том числе с помощью материалов Московской электронной школы. В помощь как учащимся, так и преподавателям реализованы средства визуализации относительной успеваемости каждого ученика на лепестковых диаграммах. Средняя абсолютная ошибка (MAE) прогноза балла ЕГЭ столичных учеников по русскому языку составила 4.40, по математике — 7.64.

\* Использовался ансамбль моделей — LightGBM, XGBoost, CatBoost.

На основе результатов ОГЭ и ЕГЭ, количество участников и победителей олимпиад, интерес к проектной деятельности по предмету учителя определяются лучшие преподавательские кадры.

В планах на будущее — повышение точности моделей за счёт обогащения наборов данных дополнительной информацией о психологических характеристиках личности учащихся по результатам анкетирования, обучении у репетиторов, других факторах вне системы государственного образования. Планируется также разработка инструментов помощи в профессиональной ориентации школьников, выборе профильной школы и университетской программы. Ведутся также работы по изучению корреляционной зависимости экзаменационного балла ЕГЭ и учебного пособия. В будущем система сможет формировать рекомендации по выбору учебных материалов для каждого из предметов.

### 4. Рекомендательная система по капитальному ремонту многоквартирных домов

В сфере ЖКХ многообещающим пилотным проектом Москвы является прогнозирование аварийности и технического состояния жилых многоквартирных домов.

Традиционно очередность проведения работ в конкретных домах определяется региональными программами капитального ремонта на основе продолжительности эксплуатации дома и его инженерных систем (с момента ввода в эксплуатацию или предыдущего ремонта) и оценки технического состояния дома по результатам мониторинга. Межремонтные сроки и другие параметры установлены в нормативно-технических документах (СНиП, ТСН, ТУ и др). Однако в ряде случаев возникает необходимость проведения ремонта раньше регламентированного срока.

Новым подходом к этой задаче является применение методов машинного обучения. Набор данных формировался на основе паспортов домов с указанием местоположения, физико-техническими характеристиками, информацией по фактически проведенным ремонтам обогащается показаниями датчиков и обращениями жителей по поводу ремонта в Единый диспетчерский центр. Обучающая выборка включает данные 8,5 тыс. домов, около 90 тыс. договоров на проведение ремонта в более 8 тыс. домов, 6 млн

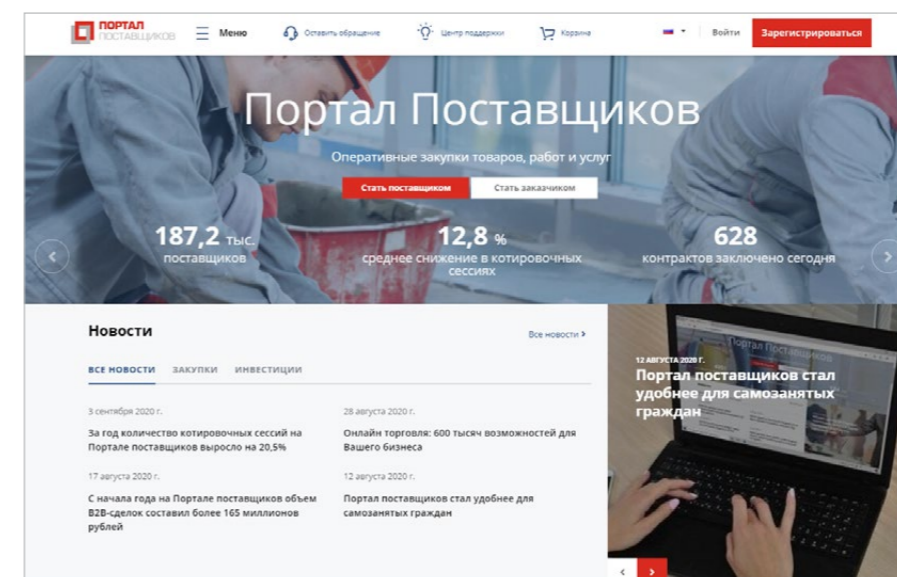


Рис. 7. Портал поставщиков

обращений жителей 30 тыс домов по 27 видам требуемого ремонта: лифты, сантехника, канализация, кровля и др. В модели «Предсказание поломок лифтов» (LightGBM) использовалось 20 признаков.

В результате пилотного проекта был сформирован рейтинг зданий, рекомендуемых для включения в программу капитального ремонта, с указанием статей\*.

### 5. Персональные рекомендации на портале поставщиков и платформе Инновационного кластера Москвы<sup>[154]</sup>

Портал поставщиков Москвы является цифровой платформой автоматизации закупочной деятельности малого объема, своего рода онлайн-магазином. На ресурсе зарегистрировано 187 тыс. поставщиков из 36 регионов России, в каталоге хранится больше полумиллиона стандартных товарных единиц.

Отслеживать «в ручном режиме» закупки — а их ежемесячно проводится около 40 тыс. — довольно ресурсоёмкий процесс, несмотря на функцию расширенной фильтрации. Поэтому в 2017 году на портале был запущен рекомендательный сервис, который подбирает для предпринимателей потенциально интересные для них закупки и направляет их в персональной рассылке. Исходными данными для модели являются опубликованные планы закупок, объявленные закупки и история работы поставщика на платформе.

\* Для прогнозирования использовался градиентный бустинг (LightGBM, Catboost). Точность предсказания модели «Лифты» по ROC\_AUC составила 0.95, модели «Кровля» 0.6.

В прошлом году более 12 тыс. предпринимателей воспользовались сервисом, то есть изучили информацию о рекомендованной закупке, что составило 7% аудитории целевой рассылки.

Похожая служба реализуется и на портале Московского Инновационного кластера. На основе семантического анализа запросов пользователей на сайте i.Moscow система рекомендует участникам кластера бизнес-партнёров. Тестируются также другие рекомендательные сервисы, включающие поиск инфраструктуры, инвесторов, мер поддержки, технологий и результатов НИР.

### Другие перспективные проекты Москвы

В городе апробируются также другие модели прогнозирования и поддержки принятия решений: инструменты поддержки планирования строительства объектов городской инфраструктуры, предсказание расхода ресурсов ЖКХ, выявление очагов преступности и анализ оперативной информации, построения карты криминогенности. С помощью технологий видеонаблюдения и видеоаналитики в Москве раскрывается две трети всех зарегистрированных преступлений. Автоматизируется разбор обращений граждан, их классификация и направление в профильные органы исполнительной власти.

## ЗАКЛЮЧЕНИЕ

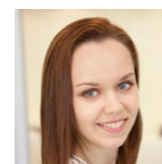
Природа человека не меняется — с древних времён люди стремятся заглянуть в будущее. Но если раньше страждущие обращались к дельфийским оракулам и пророчествам Нострадамуса, то сейчас — к датасайентистам и методам предсказательной аналитики.

В статье «Предсказывать — глупейшее занятие» основоположник советской космонавтики академик Борис Раушенбах писал, что «Человеческая жизнь развивается по очень сложным законам, нелинейным, как выражаются математики, а предсказывать мы умеем только линейно; прямое продолжение годится на пару лет, ну лет на 10, не больше [...] Поэтому ничего предсказывать нельзя, даже профессионалу в научной сфере» <sup>[155]</sup>.

Вместе с тем технологические гиганты нанимают в штат известных футурологов, и работа в Google Рэя Курцвейла (Raymond Kurzweil) — ярчайший

тому пример. А в Группу ООН по цифровому сотрудничеству были приглашены разработчики прогностических технологий в здравоохранении, эпидемиологии, социальной и финансовой сферах: доктор Кира Радинская (Kira Radinsky), Софи Суовон Ом (Sophie Soowon Eom) и другие <sup>[156]</sup>.

Конечно, стремительные социальные и технологические трансформации во всех сферах жизни ещё сильнее сужают горизонты прогнозирования. И одновременно порождает новые прогностические методы и данные, позволяющие заглянуть в будущее. Современные «умные» города обладают значительным количеством источников больших данных: интеллектуальные транспортные системы, устройства Интернета Вещей, ГИС и платформы видеонаблюдения. В этом контексте данные выступают в качестве стратегического актива мегаполисов, с помощью которого они могут решать федеральные и региональные задачи, улучшать городскую среду и способствовать достижению глобальных целей устойчивого развития ООН <sup>[157]</sup>.

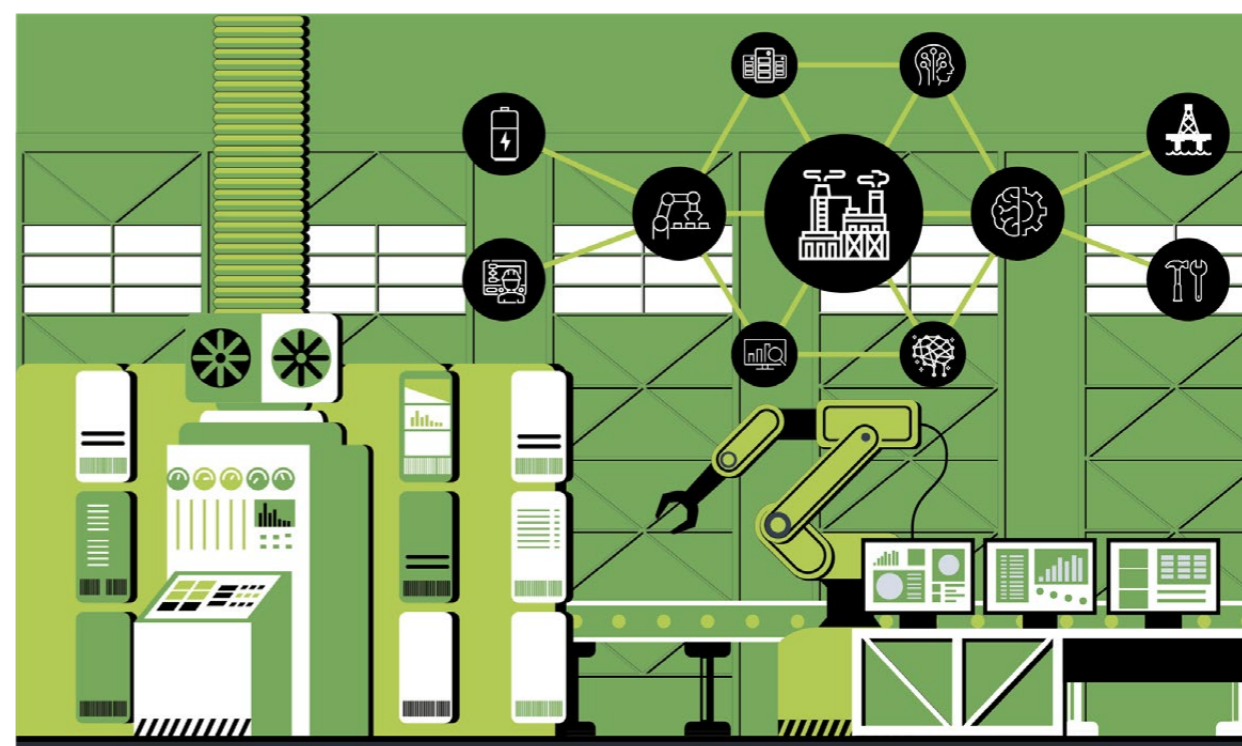


**Елена Ляпина**

Архитектор решений, ГК Цифра

# Предсказательная аналитика и системы принятия решений в промышленности

Цифровизация — один из главных приоритетов для промышленных компаний. Крупнейшие российские игроки ведут системную работу в этом направлении, много лет непрерывно работая над повышением эффективности производства и запустив масштабные оптимизационные программы. Частью усилий по цифровой трансформации является использование различного рода систем на основе машинного обучения и анализа больших данных.



Традиционно изменения в процессах и инвестициях в капиталоемкие и гиперконкурентные отрасли требуют больших денежных вложений. Инвестиции относятся к капитальным расходам и являются долгими деньгами. Поэтому текущее финансовое положение может не позволять провести полный физический ремонт заводов или сооружений. Высокие затраты подобного рода приводят к поиску более дешевых альтернатив. Инвестиции в цифровые решения и решения на базе искусственного интеллекта могут быть на порядок меньше.

При создании стратегии трансформации можно найти место для использования ИИ во всех операциях. Некоторые из сценариев обеспечивают возможность быстрого возврата инвестиций, как, например, системы поддержки принятия решений, сопровождающие технологические процессы, а другие требуют длительной апробации. В этой статье мы рассмотрим популярные и перспективные кейсы использования методов машинного обучения в промышленности.

Независимо от предметной области, в последние три года наблюдается расцвет методов искусственного интеллекта, особенно в области глубокого обучения. Чаще стало происходить сотрудничество вендоров и стартапов по машинному обучению. Фокус исследователей сместился анализа больших данных к периферийным вычислениям (edge computing). Однако, промышленные компании все еще с осторожностью внедряют прогрессивные подходы.

Отражение преимуществ внедрения искусственного интеллекта можно найти в общих тенденциях рынка. В новом отчете Meticulous Research [158] отмечает, что рынок искусственного интеллекта в производстве будет расти примерно на 40% ежегодно и достигнет оценки \$27 млрд в 2027 году. Объем рынка AI в индустрии нефти и газа Mordor Intelligence оценила [159] в \$2 млрд в 2019 году. На 2025 год оценка составляет \$3,81 млрд с ежегодным приростом в 11%.

#### AI в индустрии нефти и газа



#### ПРЕДСКАЗАНИЕ СПРОСА

Big River Steel, сталелитейный стартап из Арканзаса (США) объединился с консалтинговой фирмой по AI, чтобы улучшить свои процессы. В качестве приоритета они выбрали [160] задачу предсказания спроса с использованием моделей машинного обучения, обученных на данных об историческом спросе на сталь, макроэкономических данных и активности клиентов. В планировании активностей система отталкивается от прогнозируемого спроса, что позволяет улучшить управление предложением и уменьшить складские запасы. Из-за волатильности спроса в сталелитейной промышленности, новые подходы к аналитике имеют большой потенциал.

Больше инсайтов стало доступно из автомобильной промышленности [161]. Один из примеров объясняет подход мягких вычислений при прогнозировании спроса. Используемыми моделями являются нечеткая логика и метод Дельфи, а также применение нейронных сетей для анализа временных рядов. Общий подход объединяет мнения экспертов с искусственным интеллектом.

Автомобильные компании поднатерели больше остальных в прогнозировании спроса, что показывает пример Volkswagen AG Data: Lab Munich [162]. Они инициировали более 100 специализированных проектов для прогнозирования продаж в разрезе продуктов и регионов. По оценкам консалтинговой компании Capgemini, крупные производители [163-165] структурно сложных автокомпонентов (узлов, агрегатов) могут увеличить операционную прибыль до 16% за счет масштабного внедрения искусственного интеллекта.

Другой вендор, Predictive Layer [166] предлагает решение для прогнозирования потребления электроэнергии. Они поставляют движок для динамического ценообразования, который включает анализ спроса и эластичности предложения компании-клиента. По утверждению компании, удалось сформировать достаточно точный прогноз потребления на следующий день [167]. Заявленная экономия годовой закупки электроэнергии на национальных рынках электроэнергии ЕС составляет более \$45 млн.

Нефтегазовая отрасль также использует традиционные и продвинутое модели для прогнозирования энергопотребления [168] на различных горизонтах. Однако, обзор 2019 года [169] пока-

зал, что традиционные интерпретируемые модели остаются предпочтительными для ежегодного прогнозирования энергопотребления на национальных уровнях. Так как это модели не “черного ящика”, человеку более понятна взаимосвязь между данными о потреблении и влияющими факторами. Модели глубокого обучения до сих пор используются редко, и их эффективность и надежность требуют дальнейшей проверки несмотря на то, что их эффективность может значительно превосходить классические подходы на задаче прогнозирования спроса.

#### ОЦЕНКА РИСКА И ПРЕДИКТИВНОЕ ОБСЛУЖИВАНИЕ

Возросло количество публикаций по применению ИИ в управлении рисками за последние десять лет, удвоившись в период с 2016 по 2018 годы. Благоприятной почвой для применения однозначно являются автомобилестроение и строительство. Нефтегазовая, горнодобывающая и энергетическая отрасли идут следом. Научный обзор [170] по обеспечению безопасности, проведенный в 2020 году, дает представление об использовании искусственных нейронных сетей для оценки рисков.

Двумя основными источниками данных для применения методов машинного обучения являются текстовые и табличные данные. Текстовые данные в управлении рисками представляют собой отчеты об инцидентах и авариях. Табличные данные поступают в виде частоты аварий или других временных рядов.

Практики оценивают оценку риска [171] нефтяных компаний как доступную услугу от технологической компании, которая обогащает данные разведки. Большинство кейсов сосредоточены на совершенствовании существующих методов моделирования при создании карты рисков по результатам разведки и совмещения их оценки с данными, поступающими в режиме реального времени при бурении.

Сейчас модная тема — предиктивное обслуживание, но для развития этого направления нужно высокое качество данных и системная перестройка процессов технического обслуживания и ремонта. В нефтегазовой промышленности системы обнаружения аномалий и предиктивного обслуживания являются важными [172] областями применения машинного обучения. Широко рас-

пространное применение сенсоров способствует использованию этого вида аналитики. Дефекты турбомашин, насосов [173] и двигателей могут быть обнаружены на ранней стадии, и, таким образом, можно предотвратить дальнейшие потери за счет перевода внеплановых ремонтов в плановые.

Выгоды от имплементации подходов анализа данных были наглядно продемонстрированы в проекте BCG [174] по созданию системы управления нефтегазовой платформы. Получение оперативных данных помогает в предупреждении остановов динамического оборудования, и оценке достижимого потенциала его работы. Помимо этого, контроль процесса открывает новые возможности по работе с узкими местами производства.

#### РАННЕЕ ОБНАРУЖЕНИЕ АНОМАЛИЙ В ТЕХНОЛОГИЧЕСКОМ ПРОЦЕССЕ

Безусловно, один из ключевых приоритетов для всей индустрии — это безопасность, исключение человека из опасных процессов, онлайн-мониторинг режимов работы и нарушений режимов эксплуатации оборудования, а также оперативное реагирование на ключевые риски.

Одной из возможностей систем ИИ является обнаружение аномалий. Этот кейс можно рассматривать как одну из основных целей промышленного интернета вещей. Известные шаблоны процессов могут быть прерваны редкими событиями, которые обычно не обнаруживаются специалистом. При худшем варианте развития событий аномалия может привести к остановке всей производственной линии. Учитывая огромные массивы данных с сенсоров промышленных предприятий, поиск аномалий с помощью ручного осмотра кажется неразумным.

Украинский поставщик технологий Sciforce рассказывает о своем клиенте, который хотел ускорить регулярные алгоритмы обработки и повысить стабильность системы. Они создали коммерческий процесс обнаружения аномалий, который включает само обнаружение аномалий и прогнозирование будущих аномалий. В качестве моделей для обнаружения они использовали автоэнкодеры, для прогнозирующей части использовались рекуррентные нейронные сети. Эти модели давали достаточно точные прогнозы на срок до 10 минут.





## ЦЕНООБРАЗОВАНИЕ

Связанный с эпидемией кризис еще больше обострил давно назревшую потребность в цифровизации промышленности. Из-за отрицательных цен на нефть, компании нефтедобывающей промышленности оказались в крайне нестабильной среде, заставившей их переосмысливать стратегии и операционную деятельность.

Когда мы смотрим на управление цепями поставок нефтегазовой отрасли, есть несколько моментов, в которых можно применять искусственный интеллект. ИИ применяется в прогнозировании рыночных цен на нефть и готовой продукции. Это помогает в принятии решений по ценовым контрактам. Исследование моделей ценообразования на основе ИИ является постоянной точкой притяжения исследователей.

Классические модели машинного обучения для прогнозирования цен на нефть уже используются более пяти лет: можно найти оценку перспективности кейса в обзоре 2015 года [175]. Исследователи пришли к выводу, что из-за сложности факторов, влияющих на цену нефти, модели очень ограничены. Необходимые входные данные еще предстоит выяснить. Несмотря на непредсказуемость цен на нефть, малейшее преимущество необходимо использовать в условиях жесткой конкуренции.

## ЛОГИСТИКА

Управление сырьем, логистикой и транспортом тесно связано с концепцией умной фабрики. Если посмотреть на отрасли, которые начали процесс цифровой трансформации в направлении полного цифрового производства, автомобильная промышленность снова становится цифровым лидером. В логистике, целевой концепцией являются автономные цифровые заводы и фабрики. В масштабах международных корпораций стало возможным создание единой цифровой логистической системы.

Одним из примеров интеллектуального решения, уже находящегося в эксплуатации, является программа General Electric [176]. GE управляет 20 фабриками по всему миру и связывает данные в режиме реального времени от проектирования, производства, цепочки поставок до распределения и обслуживания клиентов. Они вместе составляют единую взаимосвязанную интеллектуальную систему [177]. DHL и IBM описывают функцию ИИ в логистике в связке с роботизацией процессов (RPA), где ИИ учится копировать и улучшать процессы на основе данных, предоставленных RPA. Системы искусственного интеллекта становятся помощниками в логистических процессах, основанных на человеческих решениях. Помимо RPA, системы искусственного интеллекта помогают логистике перевести свою операционную модель с реактивного поведения на прогнозирование и упреждающие операции с помощью различных моделей прогнозирования.



## ПРОИЗВОДСТВЕННЫЕ ПРОЦЕССЫ

Умная автоматизация меняет процессы на фабриках. Несмотря на то, что большинство технических процессов прошло многолетний традиционный процесс оптимизации, машинное обучение может выжать из оптимизации еще больше [178]. Сложные процессы характеризуются большой степенью вариативности. Для создания стандартизированного подхода к управлению техпроцессом создаются экспертные системы на основе анализа данных, помогающие операторам качественнее управлять производственным процессом. Использование различного рода цифровых советчиков является приоритетом в первую очередь для процессных

производств. На обогатительных и металлургических предприятиях управление технологическими процессами основывается на умениях и знаниях операторов, но можно увидеть первые проекты в этой области.

Примером применения таких систем является оптимизация расхода добавок на многих производственных процессах, например использования ферросплавов в металлургии. Показательным является кейс НЛМК [179], раскрывающий возможности машинного обучения по прогнозированию химического состава, при добавлении тех или иных материалов. Люди часто перестраховываются и используют дорогие добавки с запасом, чтобы гарантировать продукции определенные свойства. Применение методов машинного обучения помогает осуществлять более глубокий анализ исходного состава материала, качества сырья, потоковые данные сотен датчиков, прогнозировать свойства выходной продукции и рекомендовать необходимые действия.

В 2019 году фирма Accenture [180] представила финансовую модель от внедрения для типичного предприятия химической промышленности. В этом примере предприятие с годовым доходом в \$11,3 млрд может увеличить свою прибыль на 10%. Поставщик технологий bitrefine [181] предлагает общее решение для широкого спектра процессов. Благодаря подготовленному набору инструментов машинного обучения, потенциально становится возможна оптимизация процесса обогащения руды, газопереработки, нефтепереработки, химической промышленности, пластмасс, стекла и производства полупроводников. Когда с помощью цифровизации мы можем увеличить скорость разлива стали или снизить количество брака — это прямое увеличение дохода.

## КОНТРОЛЬ КАЧЕСТВА

К системам поддержки принятия решений можно также отнести контроль качества, помогающий визуально контролировать свойства выпускаемой продукции и выстраивать логику работы с аномалиями после. Здесь ИИ снова является ядром систем. Используя методы глубокого обучения, ИИ может обнаружить дефекты лучше и точнее человека.

Согласно исследованию McKinsey [182], в области контроля качества все еще есть большое поле для улучшений, особенно с использованием ИИ



для автоматического тестирования и определения типов дефектов. Оценивается, что производительность может увеличиться на 50%. Усовершенствованные методы распознавания изображений и обнаружения неисправностей могут повысить уровень обнаружения дефектов до уровня в 90%. Это увеличение является значительным улучшением по сравнению с ручным осмотром.

Визуальный контроль качества может быть использован для широкого спектра продуктов. Диапазон варьируется от осмотра деталей, солнечных панелей, окрашенных кузовов до текстурированных металлических поверхностей. Практическими примерами автоматизированного визуального контроля являются подходы DevisionX [183], предлагающих решения для автомобильной промышленности. Его можно использовать для проверки соответствия сварки, штамповки, сборки или проверки геометрии продукции. Захваченное изображение анализируется с помощью алгоритмов машинного обучения, и система может принимать меры и предоставлять настраиваемые отчеты в режиме реального времени.

К системам поддержки принятия решений можно также отнести контроль качества, помогающий визуально контролировать свойства выпускаемой продукции и выстраивать логику работы с аномалиями после. Здесь ИИ снова является ядром систем. Используя методы глубокого обучения, ИИ может обнаружить дефекты лучше и точнее человека. Например, на заводе BMW [184] в Дингольфинге ИИ сравнивает данные заказа автомобиля с фотографией автомобиля в зоне окончательной проверки, чтобы подтвердить обозначение модели. Audi установила систему анализа изображений с глубоким обучением в своем пресс-цехе в Ингольштадте для анализа изображений в режиме реального времени и проведения проверок качества продукции.



Пресс-центр Audi <sup>[185]</sup> в Ингольштадте утверждает, что устраняет 100% потерь, вызванных плохими проверками.

### НОВЫЕ ПОДХОДЫ ЗАВОЕВЫВАЮТ ПОПУЛЯРНОСТЬ

В последние годы популярность набрали глубокие нейронные сети и другие концепции машинного обучения, и их применение процветает. Пока они широко распространены во многих отраслях, но не в тяжелой промышленности. Здесь, в нефтегазовой отрасли, горнодобывающей промышленности, обрабатывающей промышленности, металлургии и металлообработке все еще преобладают традиционные подходы контроля процессов. Легко добывать ценность из высококачественных данных, если бизнес-модель компании изначально строилась вокруг них. Компании тяжелой промышленности имели дело с физическими активами и технологиями разной степени автоматизации, поэтому долго присма-

тривались к продвинутому анализу данных.

Большинство проектов все еще находится на экспериментальной стадии, или же их продвигают поставщики технологий. Крупные компании, такие как Royal Dutch Shell, Tata Steel, General Electric, Volkswagen Group и Bayerische Motoren Werke, проходят период цифровой трансформации. Их стратегии включают в себя множество небольших ИИ-проектов, направленных на получение быстрого эффекта от внедрения. В следующие пять лет многие компании, сделавшие первые шаги в цифровой трансформации, столкнутся с вызовом цифровизации в масштабе.

Академические исследования показывают зрелость технологий ИИ и возможности отрасли к адаптации технологии в своих технологических процессах. Некоторые стартапы демонстрируют успешные реализации передовых технологий и являются пионерами отрасли.



**Игорь Пивоваров**

Центр Науки и Технологий  
Искусственного Интеллекта МФТИ



**Наталья Гутенева**

Центр Науки и Технологий  
Искусственного Интеллекта МФТИ

## Большие данные в финансах

Индустрия финансовых услуг претерпела серьезные преобразования. Инновации в технологиях передачи данных стимулировали рост методов прогнозной аналитики и интеллектуального анализа данных, которые за последние несколько лет кардинально трансформировали банковское дело.

«Данные! данные! данные!»  
воскликнул он нетерпеливо.  
«Я не могу делать кирпичи  
без глины!»

Артур Конан Дойл,  
«Приключения Шерлока Холмса»

Платформы больших данных, такие как Hadoop, не только экономно хранят большие объемы структурированных и неструктурированных данных, но и эффективно обрабатывают эти большие данные. Результатом является постоянный рост качественных вычислительных моделей, работающих в реальном времени. С появлением технологий Hadoop и Big Data, финансовые сервисы теперь могут получать и анализировать огромные объемы данных по гораздо более низкой цене.

Пользователи могут не только генерировать информацию, используя традиционную модель специальных запросов или описательные модели, но и создавать расширенные статистические модели на основе этих данных. Обычно для вы-

полнения высокоточного прогнозного моделирования используются классификация, кластеризация, регрессионный анализ и нейронные сети. Современные модели включают в себя как традиционные данные из СУБД, так и потоковые данные в режиме реального времени.

Сфера финансовых услуг известна проблемой большого количества разрозненных данных. Вот лишь некоторые традиционные источники внутренних данных:

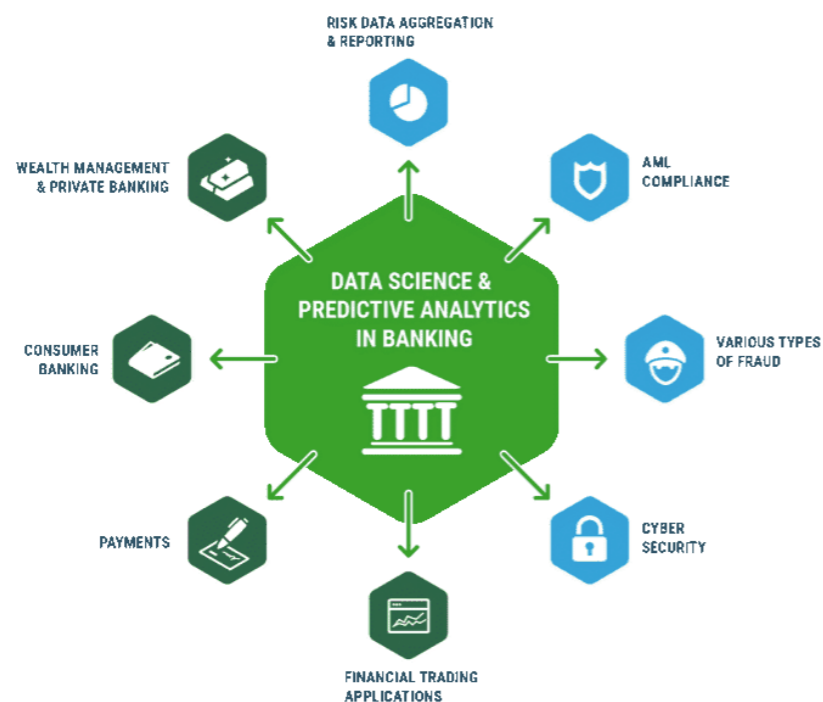
- Данные учетной записи пользователя
- Демографические данные
- Транзакционные данные на разных уровнях детализации
- Платежные данные
- Данные о биржевой торговле и открытых позициях
- Данные бухгалтерского учета, включая кредиторскую задолженность, дебиторскую задолженность, управление денежными средствами, информацию о закупках
- Данные из банковской отчетности

## PREDICTIVE ANALYTICS ACROSS BANKING <sup>[188]</sup>

### COLOR KEY

**DEFENSIVE**  
SAVE THE BANK

**OFFENSIVE**  
DRIVE PROFITS & COMPETITIVE ADVANTAGE



Все эти данные создавались человеком, но сегодня, благодаря достижениям в области интеллектуальных датчиков и устройств на основе телеметрии и POS-терминалов, данные автоматически генерируются в беспрецедентном масштабе. А все растущая сеть мобильных приложений, включая приложения для социальных сетей, будет продолжать генерировать новые огромные объемы данных — а также и связанные с ними проблемы.

### ПРОГНОЗИРОВАНИЕ И АНАЛИЗ РИСКОВ

Отделы внутренней безопасности все чаще обращаются к методам обработки данных для создания и запуска моделей выявляющих потенциальные риски. Множество типов моделей и алгоритмов используются для поиска признаков мошенничества и аномалий в данных для прогнозирования поведения клиентов. Типично используются модели:

- Байесовские фильтры
- Кластеризация
- Регрессионный анализ
- Нейронные сети

Исследователи данных и бизнес-аналитики вооружены такими инструментами, как MapReduce, Spark (через Java, Python, R), Storm, SAS, Hadoop

и другими для создания этих моделей. Это делает разработку моделей для обнаружения мошенничества очень простой для реализации.

Работа с рисками стала самым частым применением Data Science. Она включает в себя прогнозирование и измерение различных показателей риска на рынке. В потребительском банкинге такие сектора, как ипотечное кредитование, кредитные карты и другие финансовые продукты, все в значительной степени используют Data Science для классификации продуктов и клиентов по различным категориям риска.

Для обнаружения в реальном времени мошеннических действий в аккаунте пользователя часто применяются прогнозные модели для постоянного анализа расходов клиентов, его местоположения и командировок, сведений о занятости и даже постов в социальных сетях.

### ТОРГОВЛЯ НА БИРЖЕ

Торговля на бирже — это еще один пример использования прогнозной аналитики на данных. Можно уверенно утверждать, что предсказательная аналитика и системы принятия решений стабильно работают в биржевой торговле. Информации по этим применениям крайне мало, но это понятно: компании, у которых предсказательная

аналитика работает и приносит доход предпочитают о ней не распространяться. Но целая индустрия торговых биржевых роботов показывает, что в этой области ПА успешно справляется со своими задачами.

Сегодня можно агрегировать множество данных, включая данные о торговле, открытых позициях, корпоративных новостях и даже неструктурированные данные постов в социальных сетях. И все эти данные объединяются, чтобы предугадать возможные движения рынка, доходность и прибыльность трейдеров. Data Science повторно использует уже существующую инфраструктуру, помогая подключать сложные торговые стратегии, основанные на алгоритмах, к новым классам активов, включая акции, FOREX, ETF и товары.

### РОЗНИЧНЫЙ БАНКИНГ

Эффективность методов прогнозной аналитики была доказана в нескольких применениях для потребительского банкинга. Такие методы, как персонализированный маркетинг и персонализированные рекомендации, становятся все более распространенными. Можно наблюдать четкую тенденцию перехода финансовых учреждений в направлении подхода «сначала аналитика» к созданию новых бизнес-приложений.

Сегодня большинство розничных и потребительских банков еще не имеют полного представления о своих клиентах. Однако, используя возможности сбора данных и построения прогнозных моделей на них, банки могут обеспечить взаимодействие с пользователем, которое конкурирует с наиболее популярными платформами социальных сетей, и в то же время получить полное представление о своих клиентах во всех точках контакта.

Четыре способа как использовать прогнозную аналитику в финансовом секторе:

1. Используйте данные клиентов для создания цифровых платформ, которые лучше привлекают клиентов, партнеров и сотрудников.
2. Соберите и проанализируйте все доступные потоки данных, чтобы получить полное представление о розничном клиенте.
3. Связывайте разные данные из разных хранилищ, разрушайте внутренние организационные барьеры препятствующие этому.
4. Используйте подходы и модели, основанные на данных, чтобы внедрять и поддерживать культуру непрерывных инноваций и экспериментов.



С. Серый

Алгоритмист, Ноосфера

# Прогностика динамических процессов социума

## Смогут ли AI стать Индустрией?

За последние десятилетия человечество сместило акцент прогнозирования с «**астрологического**» (гороскопы, шаманы с бубнами, карты Таро) на «**инженерный**» подход (Data Science, Machine Learning). Удивительно, но «инженерный» подход, по сути, так же малообъясним для обычных людей — точнее, объясним неким набором непонятных слов — что и «астрологический»: нейронные сети, глубокое обучение, искусственный интеллект...

Другой вопросный аспект, на который «инженерный» подход даёт даже меньше ответов, чем шаманство — класс и масштаб решаемых проблем людей. В гаданиях по «Книге жизни» люди могли задавать ЛЮБЫЕ вопросы и искать ответы на ЛЮБЫЕ проблемы. В основе гадания находился человек (и его проблемы). В основе «научного» подхода может находиться что угодно, но не проблемы человека или человечества: продвижение (реклама), доставка (логистика), контроль (распознавание), общение (NLP), торговля (биржа) и т.д., и т.п.

«Инженерный» подход прекрасно решает «не-человеческие» задачи (в которых люди не являются акторами) для массового применения, например, прогноз погоды. Сложно представить какое количество гадалок и шаманов понадобилось

бы сейчас, чтобы отвечать на вопросы про погоду от интернет-пользователей. И это при том, что прогнозы погоды — неотъемлемая часть любого новостного ТВ-канала и половины сайтов федерального или локального уровня! Количество установок виджетов систем прогноза погоды всё равно лидирует.

Но как только предстоит решать задачу конкретного человека («любит/не любит»), группы людей (покупать или продавать валюту), этноса (выборы власти) или всего человечества (летим на Марс или остаемся на Земле) — с «инженерным» подходом у нас возникает проблема. «**Проблема экспертов**».

Не погружаясь глубоко (статья всё-таки про другое), «проблема экспертов» — когда человек/система выдает прогнозы на основе своего/накопленного прошлого опыта и живет только до прихода «**Черного лебедя**» — приводит к глобальной проблеме для всего Человечества: мы не знаем, что делать завтра. Эксперты не справились ни с одним экономическим кризисом, ни с одной большой войной, не предугадали победы Трампа или настрой англичан на Brexit, и даже самую возможность существования невозможной «отрицательной цены» на нефть.

## Грубо говоря, если ИИИ (индустрия искусственного интеллекта) когда и сможет стать ИНДУСТРИЕЙ, то именно тогда, когда (И) ИИ сможет заменить Астрологию.

Точно так же, как в прошлые века Физика полностью заменила Метафизику, а Химия для пользы человечества поглотила Алхимию.

Не секрет, что большинство властителей (царей, падишахов, королей и президентов) имеют «придворных» астрологов. Даже в XXI веке. Почему? — Потому что никто другой не рискует дать ответ на самый «простой» вопрос: «**Что делать завтра?**»

Чем проще вопрос — тем «человечнее» задача. Именно люди ставят самые «простые» вопросы, без всяких Big Data, уровней рисков и прочих доверительных интервалов («враки детям» Пратчетта).

**Чем проще вопрос — тем выше уровень принятия решения:** президент, руководитель компании, глава семьи, личность-«Я». Закрывать страну на карантин или не закрывать? Гасить волну BLM силой или «само рассосется»? Открывать кафе сразу после карантина или подождать? Едем этим летом семьей отдыхать за границу или к бабушке на дачу? Поехать на работу на метро или побыть на удалёнке?

**Парадигма «простого вопроса»** (ППВ: что делать завтра?) позволяет поставить еще несостоявшуюся, но потенциально перспективную, ИИИ с головы на ноги: **заняться поиском ответов на реально важные для людей вопросы**. Какой смысл прогнозировать отдачу от рекламы ресторана, если власти завтра введут режим карантина? Какой смысл прогнозировать доходы от туризма, если завтра закроют границы до конца года? Какой смысл брать ипотеку, если тебя завтра уволят, а твои «индустриальные» навыки стенографистки никому не нужны?

Применяя ППВ, получается, что во главе — проблемы всего Человечества, ниже — этноса/страны, еще ниже — регион/город, и т.д. до семьи и личности. **Человечество-народ-общество-группа-человек**. Тогда, на нужном уровне принятия решения, для ответа на «Что делать завтра?» нам становятся доступными (прозрачными) взаимозависимости акторов и прогностика «**сверху-вниз**».

Собрались переезжать в другую «лучшую» страну? Не надо начинать с поиска дома у пляжа и вакансий на рынке труда. Куда лучше посмотреть прогнозы на «**Что сделают с приезжими чужаками местные жители, когда наступит экономический кризис?**» Собрались открыть крутое арт-кафе? Не надо начинать с выбора декора и стульев. Куда лучше посмотреть прогнозы «**Какое количество покупателей перейдет на онлайн-доставку в эпоху самоизоляции?**».

Нет таких прогнозов? Вот это открытие! А почему? Потому что (И) ИИ бросило все силы (и ум) на «инженерный» подход:

замена рутинной работы — того, что раньше делал человек.

Высвобождая человека, освобождая, заменяя. Но не выступая «сердцем Данко», не освещая будущее, не ведя за собой, а лишь заменяя «вчерашние» функции. Индустрия — создает новое, а не подменяет прошлое. Пришла пора переакцентировать усилия, ставя правильные вопросы (задачи) и давая правильные (индустриальные) ответы.

У нас нет (пока) «полных» ответов за всё Человечество — полных, **в разрезе данных** всего мира — но мы уже научились давать некоторые «правильные» ответы на уровне **этноса-социума-города** (иногда района). Появляются тематические системы — дающие ответы на уровне развития индустрий (космос, финансы, умный город). Но тематические системы так же зависимы от глобальных (человеческих) прогнозов, например: сколько ни вкладывай денег в «умные очки», но люди пока к ним «не готовы».

Какой подход мы считаем достаточно хорошим (правильный-неправильный покажет очередной Черный лебедь)? **Социум** (этнос, народ, общество) — достаточно инертная, но вполне себе динамичная и «живая» структура. Как в океане, в социуме постоянно возникают цунами и штормы, но при этом есть свои Гольфстримы, веками несущие «нечто глубинное», определяющие «температуру» народов и взаимоотношений с окружающими странами.

**P.S.** Несколько лет назад возникло движение «Аффтар жжот», не закрепившееся в социуме. Нынешние поп-звезды поют-«э»кают (заменяют при пении(?) «е» на «э» в текстах песен), дикторши прогнозов погоды на ТВ дружно понижают тон на последнем слове предложения, и т.п. — да, какие-то «групповые» изменения в социуме закрепятся, но большинство — будут отринуты.

Огромные потоки публичных данных соцмедиа (десятки миллиардов сообщений, комментариев, фото и видео в сутки, триллионы пакетов данных от датчиков) — формируют текущий тезаурус этносов, ежесекундно гонят по всем линиям связи между людьми эмоции коммуникаций, наслаивают Антарктиды никому не нужных (но с удовольствием поглощаемых системами) мнений миллиардов людей (пока на Земле и чуток в космосе). Анализируя тренды в социуме (потоки на разных глубинах в океане) — мы можем прогнозировать «погоду» в обществе.

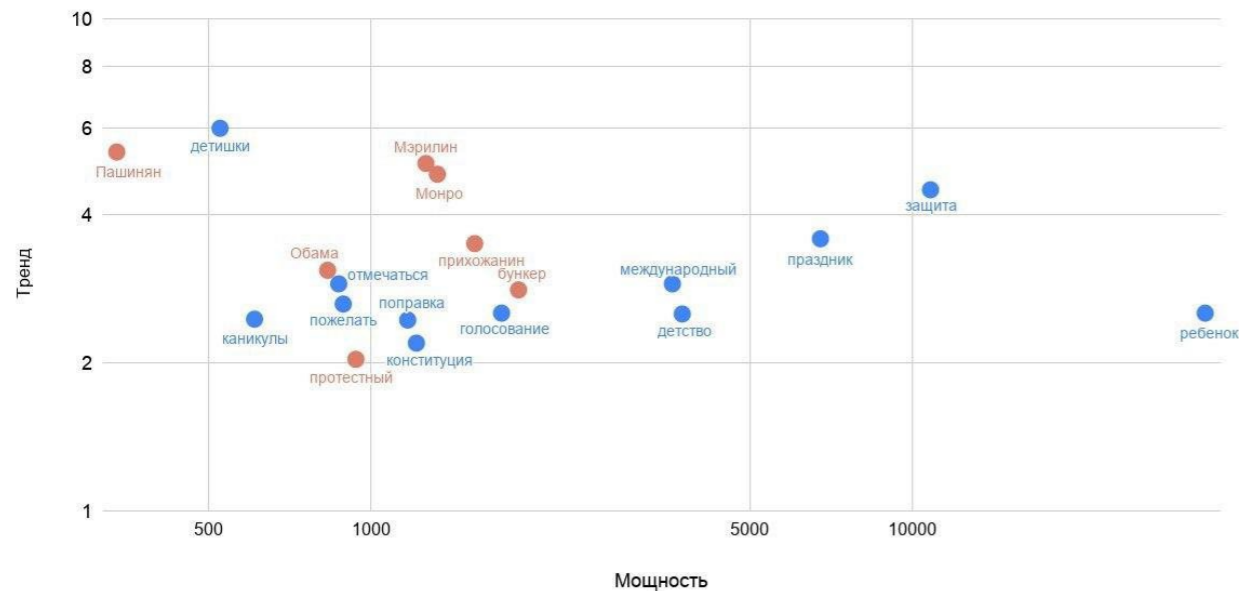
**P.P.S.** Только не надо «опускаться» до скучного уровня «прогнозов выборов» — ИИИИ это «умела» делать уже 10 лет назад (от первых выборов Обамы и выборов в Венесуэле после смерти Чавеса, до «экспертного шока» от результатов выборов Трампа и Brexit), на еще небольших, зато «чистых» данных (без ботов и прочих методик, «загрязняющих» оке-

ан данных соцмедиа). На президентских выборах 2017 года ЦИК Франции специальным решением (!) запретил на территории Франции публикацию регулярных прогнозов выборов одной из российских систем (ИИ), ссылаясь на декрет от 1977 (!!!) года.

Итак, тренды — опускаем научную беллетристику («враки детям») — как и в океане, в социуме постоянно возникают, затухают, существуют. Для ответа на ППВ («простой вопрос», помните?) можно выделить: зарождающиеся тренды, временные (затухающие, незакрепившиеся) тренды и динамику существующих (длинных) трендов. Подобное деление условно, индустрия ждет своих Менделеевых для классификации потоков данных, нам сейчас важнее «ехать, а не шашечки».

**Примеры зарождающихся трендов: гео-локального (Москва) и странового (Россия).**

Тезаурус социума. Самые «трендовые» слова, которые выросли за сутки в общественном медиа-поле. За 01 июня 2020 г. обработано 803 млн слов (из которых 115,4 тыс. уникальных) из 44,3 млн публичных сообщений с гео-привязкой «РФ». По «Москве»: обработано 44,1 млн слов (23,5 тыс. уникальных) в 2,1 млн публичных сообщениях:



**РОССИЯ В ЦЕЛОМ**

		упоминаний на 1 млн сообщений		
1.	«Ребенок»	34677	Международный День защиты детей отмечается 1 июня	
			связанные слова:	
			«защита» 10785	«пожелать» 888
			«праздник» 6755	«отмечаться» 871
			«детство» 3753	«детушки» 526
			«международный» 3599	
2.	«Голосование»	1741	Путин назначил 1 июля датой голосования по поправкам к Конституции	
			связанные слова:	
			«конституция» 1212	«поправка» 1168
3.	«Каникулы»	609	Начало летних каниул в школах	
4.	«Мурманский»	267	В Мурманской области обрушился железнодорожный мост	

**МОСКВА**

		упоминаний на 1 млн сообщений	
1.	«Бункер»	1871	во время протестов у Белого дома Трамп находился в подземном бункере; связанные слова: «протестный» (937);
2.	«Прихожанин»	1553	столичные храмы откроют для прихожан с 6 июня;
3.	«Монро»	1325	1 июня родилась американская киноактриса, секс-символ 50-х годов Мэрилин Монро; связанные слова: «Мэрилин» (1262);
4.	«Обама»	831	экс-советник Обамы обвинила Россию в причастности к беспорядкам в США;
5.	«Пашинян»	339	Премьер Армении заразился коронавирусом.

**ВРЕМЕННЫЕ ТРЕНДЫ (НЕДЕЛЬНЫЙ ПЕРИОД):**





ТРЕНДЫ ЗА МЕСЯЦ НА ПРИМЕРЕ ОДНОЙ СОЦСЕТИ

Битва трендов

В отличие от жизни сотни и тысячи лет тому назад **сейчас человечество живет в мощнейшей турбулентности**: на длинные тренды накладываются среднесрочные, ежечасно, ежеминутно взрывающиеся всполохами краткосрочных и зарождающимися трендами. Технократический XX век «захватил внимание» всех людей и создал иллюзию, что можно легко «управлять» массами: СМИ и ТВ, реклама, маркетинг, SMM, мемы, персонализация — всё легко считается и можно быстро оценить результат. Настала великая эпоха единого, объединенного, человечества под управлением интернет-гигантов!.. И тут пришел облом — классический долгосрочный, практически **вечный тренд «моя стая», переформатированный в «моя страна», «наш этнос».**

Помощь в победе Трампа оказала и компания Cambridge Analytica, с относительно небольшим бюджетом использовавшая всю **мощь долгосрочного тренда и персонализированных каналов доставки сообщения.** В 2020-м году противники Трампа «исправились» (убрали ненужные локальной «стае» призывы и внешние вызовы) и задействовали в предвыборной борьбе долгосрочные тренды (самоуправление/анархия штатов, вековые притеснения этнических меньшинств).

Технологический уклад, ввиду кратковременности своих целей (квартальная прибыль, плановый KPI, скоростная смена моделей авто или одежды) **акцентировался на коротких трендах.** Компании, существующие больше 10-20 лет — изысканная редкость. Да и то, как IBM или HP, вынужденные регулярно и кардинально менять свои стратегии (десятилетие — уже стратегия!). Даже управление государством, «рассчитанное», при фараонах и монархиях, на столетия, укоротилось до нескольких лет, из которых до половины срока уходит на подготовку следующих выборов.

Китайский кейс «прекрасен» (в алгоритмическом, «не-человеческом» взгляде) в рамках **цикличности**, столь приятной уху китайского этноса: использование энергии нового технологического уклада для закрепления и вывода долгосрочного тренда («*моя стая*»/»*мой этнос*») на новый виток. Система социального кредита — система оценки как отдельных граждан, так и организаций, по набору параметров, значения которых зависят от оценки других (членов стаи). Антиутопия «Мы» Замятина в действии.

**«Мстя» наступила быстро:** долгоживущий этнос Китая привычно закрылся Великой стеной (файрвол «Золотой щит») и организовал собственные версии «*что-там-популярное-в-остальном-мире*»; и даже молодой этнос-«солянка» в США почувствовал неладное и возопил **«MAGA!»** (Make America Great Again). Оба примера хорошо показывают силу долгосрочных трендов и (для данной статьи) используемых технологий.

Примечательно, что одни и те же карантинные действия (после вспышки коронавируса) в США и Китае привели к одинаковым последствиям в коротких трендах (спад производства, безработица, депрессия) и совершенно **противоположным результатам в длинных трендах.** Привычный изоляционизм Китая, суженный до района-города-провинция, — для защиты большей части «стаи» — не привел к хаосу переплетение разнообразных трендов этноса. В то время как в США, **столетний тренд «Америка — везде!»** (авианосцы как символ захвата и присутствия вне территории США) не мог безболезненно «сжаться до плитуса» одной страны-штата-города-района. (Белый, теперь ты не можешь захватывать новые земли? — *Тогда покаяйся, а мы захватим тебя*).

Хороший пример тренда (а заодно наглядный пример ограниченности «экспертности» на базе прошлых/устаревших данных) — ЧТО (контекст) именно имеет ввиду русскоязычный этнос под словом «Китай»: в декабре 2019 года (левая часть картинки, до коронавируса) и в феврале 2020 года (правая часть):



Пример длинного тренда, неизменно и постоянно «живущем» в нашем этносе, — слово «Бог». Каждый день, вне зависимости от погоды, кто премьер-министр, уровней самоизоляции, стоимости доллара, праздничных или рабочих дней, в публичном русскоязычном потоке соцмедиа мы упоминаем слово «Бог» **~350 000 раз** ежедневно):



Технократическо-медийная властная структура в США, управляющая каналами «коврового бомбометания» (СМИ, ТВ, классическая интернет-реклама) **жила в своем «информационном пузыре»:** со страниц газет и журналов эксперты говорили о предстоящей победе Клинтон, а билборды в городах и баннеры на сайтах рассказывали уже о планах демократов на будущее. Но «дама пик» (короткие тренды) натолкнулась на «козырного туза» (долгосрочный

Мы уже привыкли, что сознание (интеллект) отдельного человека можно сравнить с конструктором Lego, где каждый элемент отражает работу разных участков головного мозга. Вместе все эти элементы и определяют личность. Если вы понимаете, что за человек перед вами, то вы можете составить точный прогноз, чего от него можно ожидать. На этом базисе «построена» психология. Новые вызовы требуют рассматривать ком-

муниципирующие в пространстве и времени большие группы (стаи) людей не как сумму наборов Lego, а как самостоятельные сущности (экстеллект) и для нового направления нужны новые подходы, методики и решения (потребовалось много раз слов «новое»).

## Прогноз на завтра

Всё, что имеет начало, имеет и конец. Даже статья. Под занавес — «социо»-прогноз, опубликованный в декабре 2019 года (проверить оригинал и «нередактируемость» прогноза можно в Телеграм-канале soMedia, публикация от 24 декабря):

### СОЦИОПРОГНОЗ НА 2020 ГОД: БОРЬБА ПОКОЛЕНИЙ.

Около месяца понадобилось soMedia для формирования в нескольких словах прогнозных тенденций в социальном поле человечества в разных странах и этносах. Китайцы, когда хотят проклясть кого-то, желают тому «жить в эпоху перемен». 2020 год — именно такой: перемены, зародившиеся в начале 10-х, теперь стали мейн-стримом.

- 1. Локальные «революции»:** от Армении, Франции, Боливии, Гонконга, до России, Израиля, Швеции и... США. Все страны в мире ощутили на себе расслоение социума и раздрой в народе: где по этническому, где по политическому (уровню жизни), где по доступу к ресурсам.
- 2. Экономический кризис:** разница в вопросе не «будет/не будет», а «когда?» — 2020 или дотянем до 2021? Не дотянем. Кризис уже среди нас, и кризис не на месяц или год.
- 3. Разделение полов:** женская революция — как реакция социума на остановку прогресса. «Расходники» (мужчины) в законсервированном обществе «мечты по Инстаграму», в принципе, не особо нужны.

**Теория трендов в современной трактовке,** усиленная мощностью анализа высокоскоростных потоков «человеческих» данных (соцмедиа) и невозможных без методов Data Science и Machine Learning, позволяет **жить с «широко раскрытыми глазами»** (от испуга ли, или от восторга?) на завтрашний день.

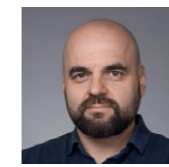
Разнообразие форм публичных конфликтов внутри и снаружи социо-этносов мешает разглядеть тектонические основы перемен — борьба поколений. 2020 год вытаскает наружу и четко обозначит два лагеря: имеющих (возрастные) и жаждущих (молодые).

Для равновесия противоборствующих армий (резкий рост количества возрастных, ввиду роста продолжительности жизни) социум вынужден увеличивать(затягивать) период «детства»: с некогда 12-14 лет до фактически 22-25 нынешних и до 30 лет в ближайшее десятилетие).

**Выводы: Конфликт — не со зла, а потому что очередной перелом (виток) революции социума. Конфликт — это развитие, пусть и через боль. Отсутствие конфликта — стагнация и смерть. Социум не может согласиться на смерть.**

Есть ли решение? Вывать к совести (любимая присказка взрослых) молодых, так же как призывать поживших — дать наконец-то нормально (?) пожить молодым (любимая мечта юных) — бесполезно. В битве никто не слышит другую сторону. Делиться «по-доброму» не получится — понятия о добре у поколений очень разные. Внимательно наблюдаем за продолжением эксперимента после передачи власти молодым в Гонконге и Финляндии. И за США, где за власть борются 80-летние старперы и старперши.

| P.P.P.S. «Что делать завтра?»



Аркадий Сандлер

МТС, Директор Центра искусственного интеллекта МТС

# Предиктивная аналитика решает задачи бизнеса в области телекоммуникаций

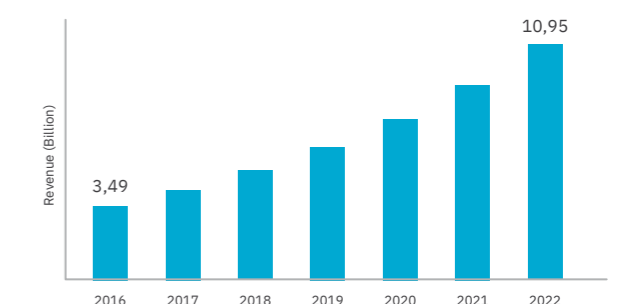
Сфера телекоммуникаций — одна из наиболее масштабных и оперативно меняющихся профессиональных отраслей в мире. Компании, основной задачей которых еще несколько лет назад было предоставление услуг связи, в наши дни становятся организациями с большим объемом накопленных структурированных и неструктурированных данных, обеспечивают качественный доступ к информации и коммуникации, предлагают клиентам широкий набор услуг и контент-сервисов. Анализ данных позволяет сформировать профиль каждого клиента и предлагать ему только те услуги и продукты, в которых он заинтересован. Однако для качественной работы с данными игрокам рынка телекоммуникаций необходимо использовать инновационные технологии, и здесь на выручку приходит искусственный интеллект, в частности предиктивная аналитика и интеллектуальные системы принятия решений.

Эрик Сигель определяет предиктивную аналитику (от англ. predictive analytics) как класс методов анализа данных, концентрирующийся на прогнозировании будущего поведения объектов и субъектов с целью принятия оптимальных решений [193]. Эта технология использует статистические методы, методы интеллектуального анализа данных, теории игр, анализирует текущие

и исторические факты для составления предсказаний о будущих событиях. Неудивительно, что прогнозные модели теперь успешно применяются в бизнесе для идентификации рисков и возможностей. С помощью предиктивной аналитики компании могут получать максимально полную информацию о своих клиентах; кроме того, данную технологию можно использовать для автоматизации и трансформации внешних и внутренних бизнес-процессов.

Согласно исследованию Zion Market Research, объем мирового рынка прогнозной аналитики к 2022 году достигнет примерно 10,9 млрд долларов США, увеличившись в среднем на 21% с 2016 года [194].

Global Predictive Analytics Market Revenue, 2016–2022 (USD Billion)



Отрасль телекоммуникаций — одна из тех, где применение предиктивной аналитики обособленно и служит разным целям:

### 1. АНАЛИЗ СИСТЕМЫ ВЗАИМОТНОШЕНИЙ С КЛИЕНТАМИ, СООТВЕТСТВИЕ ИХ ОЖИДАНИЯМ:

одним из главных принципов работы с клиентами является оценка пользовательского опыта на каждом этапе. Такой подход определяет действия компании еще до того, как человек стал клиентом, и продолжает работать вплоть до конца его истории взаимодействия с компанией. Цель — понять и предугадать потребности клиента и сопровождать его, обеспечивая максимальный комфорт. Такая работа на опережение позволяет бизнесу спрогнозировать следующие запросы клиентов и вовремя предоставить необходимые им услуги, а не реагировать на негативный пользовательский опыт. Благодаря предиктивной аналитике телекоммуникационные компании могут точно определить тенденции в потребностях клиентов. Заслуживает внимания опыт Comcast Labs, где была разработана система, способная оценивать необходимость вызова специалистов на дом. Соответствующее приложение нашло применение в службе поддержки клиентов Comcast, куда жалобы поступают по телефону. В некоторых случаях проблема определяется удаленно (например, если нужно заменить аккумулятор), а иногда сразу помогает оператор (например, перезагрузить роутер). Система верно предсказывает необходимость вызова сервисной службы на дом по запросу клиента в 90% случаев и может использоваться для автоматической сортировки звонков [195]. Модели, предсказывающие потребности клиентов, чрезвычайно полезны в любой отрасли, где существует необходимость клиентской поддержки.

Такие технологии позволяют оптимизировать работу сотрудников, равномерно распределять уровень нагрузки на них, а также поддерживать высокий уровень лояльности клиентов.

### 2. ПРОГНОЗИРОВАНИЕ И ПРЕДОТВРАЩЕНИЕ ОТТОКА АБОНЕНТОВ:

некоторые программы предиктивной аналитики позволяют сократить или предотвратить отток клиентов. Телекоммуникационной компании

Sox Communications это удалось благодаря прогнозу аналитических систем, которые позволили оперативно опросить несколько миллионов абонентов, собрав их наблюдения по использованию услуг компании, и на основе этих данных персонализировать предложения в 28 регионах [193]. Другим примером использования комплексного подхода к прогнозированию оттока клиентов является американская T-Mobile, которая сумела уменьшить отток клиентов на 50%. Для оценки вероятности отключения того или иного абонента компания использовала огромное количество метрик: как часто и кому человек звонил, как долго длились звонки, количество потребленного трафика в месяц, продолжительность нахождения в зоне покрытия сети и так далее [196].

Объем данных, накапливающийся в контакт-центрах и во всех точках соприкосновения клиента и компании, позволяет строить системы, которые обеспечивают высочайший уровень качества поддержки, тем самым способствуя сокращению оттока клиентов.

### 3. ОПТИМИЗАЦИЯ РАБОТЫ КОЛЛ-ЦЕНТРОВ:

с помощью моделей предиктивной аналитики, использующих накопленные данные по обращениям пользователей, можно правильно распределить нагрузку между сотрудниками и исключить ситуации, когда операторы перегружены или сидят без дела. Кроме того, эти системы формируют готовый набор предложений для клиента, уменьшая временные затраты оператора на «ручной» поиск и предложение релевантных услуг. «Дом.ру» и GlowByte Consulting проанализировали несколько миллионов входящих звонков за год, и на основе этого анализа был создан комплекс предиктивных моделей для повышения эффективности работы с клиентами. Система учитывает причину обращения того или иного абонента, историю взаимодействия с ним и в каждом конкретном случае предлагает максимально эффективный способ удержания, а также наиболее подходящее решение [197].

Таким образом, внедрение технологий прогнозной аналитики позволяет удовлетворять индивидуальные потребности клиентов, сокращать время на поиск и выбор рекомендаций и услуг, а значит, и на обслуживание каждого обращения.

### 4. ОПТИМИЗАЦИЯ ПРОЦЕССА ЗАКУПОК И ЮРИДИЧЕСКОГО СОПРОВОЖДЕНИЯ ДЕЯТЕЛЬНОСТИ КОМПАНИИ:

с помощью решений, в основе которых лежат технологии прогнозной аналитики, поддерживается корректная работа с поставщиками и предотвращаются просрочки и нарушения. Например, в функционал системы контрактного менеджмента [197], разработанной в Центре искусственного интеллекта МТС, входит возможность автоматического чтения юридических документов и анализ их содержания с извлечением ключевой информации с помощью технологий CV (Computer Vision) и NLP (Natural Language Processing). На основе полученных данных по каждой сделке выстраивается граф знаний, где осуществляется мониторинг выполнения задач. С учетом истории предыдущих сделок система прогнозирует возможные нарушения на раннем этапе (в случае отклонения от корректного действия в рамках плана сделки) и сформирует соответствующие уведомления для поставщика и заказчика. Решение может быть адаптировано для организаций из любой профессиональной отрасли. Как известно, крупные и средние компании работают с огромным количеством документов, строго регламентируя все бизнес-процессы, а технологические решения, в основе которых лежит предиктивная аналитика, ориентированы на автоматизацию рутинных операций.

Использование таких систем позволяет перенаправлять высококвалифицированных специалистов на работу с более сложными задачами, сокращать затраты компании на урегулирование конфликтов с контрагентами, а также предотвращать нарушения и просрочки.

### 5. ВЫЯВЛЕНИЕ МОШЕННИЧЕСКИХ СХЕМ:

мошенничество — одна из причин потери доходов телекоммуникационных компаний. Системы обнаружения мошенничества построены на алгоритмах интеллектуального анализа данных, позволяющих выявлять мошеннические действия и подозрительное поведение абонентов. Согласно исследованию SAS и ACFE, технологии искусственного интеллекта в борьбе с мошенничеством используют в 13% организаций в разных странах мира, а использование предиктивной аналитики для решения этой проблемы планируется в 52% компаний [199]. Китайский оператор China Mobile создал мобильное приложение «Не-

бесный щит», основанное на анализе Big Data и технологии машинного обучения, которое распознает характерные для мошенников фразы, перехватывает спам-рассылки и звонки. Приложение также умеет идентифицировать группы пользователей, наиболее уязвимые для спама, и предупреждать их об угрозах с помощью СМС-оповещений. Если система заподозрит, что какие-то абоненты стали жертвами мошенников, она направит номера телефонов возможных пострадавших полицейским, чтобы правоохранители оперативно с ними связались.

По данным оператора, на сегодняшний день точность при отправке полиции номеров потенциальных жертв мошенничества составляет 80% [200].

Предиктивная аналитика в отрасли телекоммуникаций оказывает качественное влияние на оптимизацию деятельности компаний, уменьшение трудозатрат, поддержку клиентского сервиса на высоком уровне и предотвращение оттока абонентов, обеспечение безопасности как клиентов, так и самих компаний. В условиях постоянных изменений в индустрии, а также высокой конкуренции организациям необходимо использовать всю мощь накопленных данных, искать новые способы их обработки и применения для улучшения своих сервисов с помощью комплекса технологий искусственного интеллекта. Бизнес в полной мере использует современную предиктивную и событийную аналитику, которая часто раскрывает такие сведения, которые иным способом обнаружить невозможно. Эти передовые статистические модели работают незаметно, в автоматическом режиме, чтобы выявлять тренды, отклонения от норм, возможности и риски.

Безусловно, модели предиктивной аналитики широко используются в разных бизнес-процессах и показывают отличный результат, однако следует отметить, что они представляют собой лишь часть более сложных систем на основе искусственного интеллекта, которые применяются комплексно. Главная роль предиктивной аналитики — в формировании контекста для применения ИИ-систем большего масштаба и выявления новых областей для автоматизации с помощью искусственного интеллекта. Таким образом, хорошая система предиктивного анализа является не просто правильным решением, а насущной необходимостью для современных телекоммуникационных компаний.



# Актуальные кейсы

04

В этом выпуске мы вводим новый раздел — кейсы. Здесь будут публиковаться конкретные кейсы компаний по актуальным темам и ссылки на наиболее интересные доклады на конференциях.





### Сергей Сорокин

Генеральный директор, BOTKIN.AI,  
г. Москва



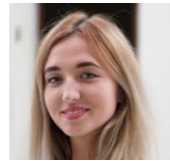
### Валентин Синицын

Д.м.н., профессор, руководитель  
отдела лучевой диагностики  
Медицинского научно-  
образовательного центра МГУ  
им.М.В.Ломоносова, Москва



### Иван Дрокин

Директор по исследованиям  
и разработкам, BOTKIN.AI,  
г. Санкт-Петербург



### Полина Пиллюс

Медицинский консультант ООО  
«Интеллоджик» (Botkin.AI), врач-  
рентгенолог Медицинского  
научно-образовательного центра  
МГУ им.М.В.Ломоносова, Москва



### Роман Матанцев

Старший специалист по анализу  
данных, BOTKIN.AI, г. Санкт-  
Петербург

## Спектр применения систем искусственного интеллекта

### для диагностики поражения легких при пандемии COVID-19 и после нее <sup>[209-212]</sup>

С ноября 2019г и по настоящее время пандемия COVID-19 стала основной угрозой здоровью населения всего мира, ее появление потребовало неотложного решения множества вопросов, связанных с ее диагностикой и лечением.

Традиционно лабораторные методы диагностики имеют приоритет между методами лучевой диагностики (медицинской визуализации) в диагностике вирусных инфекций. Основным критерием диагностики COVID-19 в остром периоде являются высокочувствительные лабораторные тесты на наличие вирусной ДНК в биологических жидкостях организма. Обычно для этой цели применяется полимеразная цепная реакция (PCR). Однако, практика показала, что этот метод имеет ограничения — в первую очередь, оказалось, что у значительной части больных этот тест изначально может быть отрицательным — несмотря на наличие тяжелых проявлений заболевания и вирусовыделения. Исследования показали, что тест PCR может у части пациентов в первые часы и даже дни заболевания быть отрицательным и только после 2–3 кратного повторения он становится положительными. Кроме того, такой тест не может различить пациентов с легкими и тяжелыми (требующими неотложного лечения и госпитализации) формами заболевания. Также известно, что у части пациентов с клинической картиной болезни лучевые методы выявляют характерные изменения легких при изначально негативных результатах лабораторных тестов, которые становятся положительными гораздо позднее. COVID-19 — системное заболевание, поражающее множество органов и тканей, но его основным проявлением и основной причиной заболеваемости и смертности являются воспалительные изменения в легких по типу интерстициальных (вирусных) пневмоний <sup>[201]</sup>.

По этой причине большую роль в выявлении и оценке тяжести COVID-19 приобрели такие лучевые методы, как традиционная рентгенография грудной клетки, рентгеновская компьютерная томография (КТ) и (в гораздо меньшей степени) УЗИ легких. Из этих трех методов КТ легких зарекомендовала себя как лучший метод визуализации для выявления наиболее важных и типичных типов (паттернов) изменений паренхимы легких при пневмонии COVID-19. Из-за применения тонких срезов и объемного характера сбора данных, КТ является лучшим из этих методов. Она выявляет изменения в легких (участки уплотнения ткани легких по типу «матового стекла», консолидации или сетчатого фиброза), которые могли быть не видны при рентгенографии (из-за малых размеров или маскировки их тенями других органов), не говоря о более тяжелых изменениях. КТ позволяет измерить плотность, размеры

и объем очагов, проводить их классификацию по типам, а также обнаруживать сопутствующую патологию легких (опухоли), изменения сердца и сосудов.

Методические руководства и рекомендации национальных и международных радиологических обществ (включая рекомендации российского Минздрава РФ) рекомендуют селективное использование КТ грудной клетки для скрининга, диагностики и наблюдения за пациентами COVID-19 — в зависимости от подозрения на его наличие, тяжести течения заболевания, доступности ресурсов для диагностики <sup>[202-205]</sup>. Они обычно советуют использовать КТ для дифференциальной диагностики COVID-19, выявления его осложнений, а также в тех случаях, когда результаты КТ могут повлиять на ведение пациента и прогноз.

С другой стороны, практика быстро показала, что КТ является одним из самых чувствительных методов выявления поражений легких при COVID-19 и в реальности ее применение оказалось намного шире, чем было рекомендовано изначально.

Были установлены типичные паттерны проявлений пневмонии COVID-19 на изображениях КТ (множественные периферические участки уплотнения по типу «матового стекла» с зонами консолидации или без них и некоторые другие проявления). Было показано, что в условиях пандемии COVID-19 (при высокой распространенности заболевания) КТ обладает высокой чувствительностью для выявления больных COVID-19 — более высокой, чем тест PCR.

По этой причине, в большинстве больниц России и амбулаторных учреждений России (как и во многих других странах) КТ грудной клетки проводится всем пациентам, поступающим с подозрением на пневмонию, вызванную COVID-19. В Москве была создана сеть из 57 амбулаторных центров, которые занимались сортировкой пациентов с подозрением на COVID-19 с помощью КТ: пациенты с легкими формами поражений легких (в сочетании с другими клиническими факторами) отправлялись лечиться домой в условиях самоизоляции, а более тяжелые больные немедленно госпитализировались и выполнение КТ при поступлении в больницу им уже не требовалось.

Но такой подход привел к быстрому росту нагрузки на кабинеты КТ и врачей-рентгенологов, а также возникла проблема стандартизации описаний КТ и оценки тяжести заболевания.

Были созданы международные рекомендации профессиональных и экспертных сообществ по применению КТ и анализу изображений рентгенологами при COVID-19. Широкую известность приобрели рекомендации Северо-Американского Общества Радиологов, Американской Коллегии Радиологов и Общества Торакальной Радиологии (RSNA/ACR/STI), Британского Общества Торакальной Радиологии (BSTI), Европейского Общества Радиологов (ESR), которые легли в основу рекомендаций Российского Общества Рентгенологов и Радиологов и Методических рекомендаций Минздрава РФ.

В этих рекомендациях изменения на КТ предлагалось классифицировать как типичные, неопределенные или атипичные в отношении вероятности их связи с COVID-19 и именно такая классификация применялась при формировании заключений. Однако, менее ясная ситуация сложилась при оценке тяжести заболевания. Было предложено использовать простые субъективные шкалы оценки тяжести изменений в легких, таких как легкая-умеренная-тяжелая или легкая-умеренная-тяжелая-критическая. В частности, наибольшую популярность в нашей стране получила визуальная шкала оценки объема поражений в легких с «шагом» в 25% (степень КТ-1 — поражение менее 25%, КТ-2 — от 25 до 50%, КТ-3 — от 50 до 75%, КТ-4 — более 75%) [205]. Но практический опыт показал, что визуальная оценка объема и площади множественных мозаичных участков «матового стекла» или консолидации в легких очень субъективна и сильно варьирует от одного рентгенолога к другому.

По этой причине практически сразу целым рядом компаний было начато создание или усовершенствование программ автоматизированного анализа изображений и программ с применением искусственного интеллекта (ИИ) для выявления, качественного и количественного анализа изменений в легких при COVID-19. В ряде ранних статей сообщалось, что программы ИИ выявляют изменения в легочной паренхиме при COVID-19 даже лучше, чем рентгенологи [206, 207]. Это результаты обоснованно критиковались рядом экспертов, указывалось на методические неточности

при выполнении данных работ [208]. Дело в том, что врачи-рентгенологи практически не делают ошибок в выявлении участков воспалительной инфильтрации лёгких на КТ изображениях (в отличие от обнаружения мелких, потенциально злокачественных узлов при поиске рака легкого).

Однако, программы ИИ оказались очень полезными и востребованными для количественной и качественной характеристики изменений в легких при COVID-19. В МНОЦ МГУ, на основании выполнения более чем 1200 КТ легких у пациентов с имеющимся или подозреваемым COVID-19, было установлено, что программы, применяющие ИИ для обнаружения, разграничения и анализа зон «матового стекла»/консолидации в легких позволяют получить гораздо более точные и воспроизводимые результаты, чем визуальный анализ врача-рентгенолога. Более того, они оказались крайне востребованными при наблюдении пациентов в динамике (за время нахождения в стационаре КТ повторялось, в среднем 3–4 раза). Сопоставление данных объемного анализа различных типов поражений имело тесную корреляцию с характером течения заболевания и его тяжестью, позволяло оценивать эффективность лечения.

Для эффективного применения такого подхода большую сеть оказывает интеграция программы ИИ с радиологическим информационным сетью (РИС) медучреждения. В этом случае после выполнения КТ, данные исследования в РИС проходят предобработку с помощью интегрированной в нее программы ИИ и рентгенолог получает для описания и для анализа изображения с цветной кодировкой различных типов поражений и анализом их объемов в абсолютных (куб.см или мл) и относительных (%) значениях.

**Благодаря применению таких программ, удалось собрать большой объем данных КТ пациентов с COVID-19, которые можно использовать для дальнейшего развития и обучения систем ИИ.**

Для количественного анализа объема вызванных COVID-19 изменений паренхимы легких в МНОЦ МГУ использовались программы:

- **Botkin.AI** (ООО «Интеллоджик», Москва, <https://botkin.ai/>)

- и программа «Гамма Мультивокс», созданная лабораторией медицинских компьютерных систем НИИЯФ им. Д.В. Скобельцына МГУ им. М.В. Ломоносова и факультетом фундаментальной медицины (ФФМ) МГУ им. М.В. Ломоносова (<https://multivox.ru>).

Опыт их применения показал, что важнейшими компонентами подобных систем являются:

- Модели анализа изображений. В зависимости от поставленных задач, программа может оценивать КТ-исследования и рентгенограммы легких, а также маммограммы.
- Рабочее место радиолога (DICOM вьюер), которое предоставляет основные и привычные для работы врача-диагноста инструменты, показывает результаты работы модели и позволяет их валидировать.
- Механизмы управления потоком исследований, позволяющие подключиться к серверу PACS, получить доступ к исследованиям, исключив из них персональные данные пациента, отправить эти исследования в модель ИИ, а затем — на подтверждение врачу-рентгенологу, предоставив список исследований пациентов, отсортированный по вероятности наличия патологических изменений в легких и их тяжести.

Основным направлением применения подобных программ до пандемии COVID-19 был поиск изменений в легких, подозрительных на злокачественные новообразования (ЗНО) — в первую очередь, выявление малых форм рака легкого в рамках программ скрининга. На решение таких задач был, в частности, направлен проект Botkin.AI.

После начала пандемии COVID-19, вектор применения таких программ ИИ был смещен. Было осуществлено обучение моделей ИИ с целью распознавания признаков вирусных пневмоний, а также автоматизация процесса количественного анализа изменений и отслеживания динамики выявленных изменений.

Стоит отметить, что для задач скрининга COVID-19 и оценки изменений паренхимы легких могут быть использованы как алгоритмы классического компьютерного зрения, основанные на оценке радиологической плотности тканей, так и на базе глубоких сверточных сетей. Послед-

ний подход, хоть и требует обучающей выборки, ведет к более точным результатам анализа КТ исследований.

Отдельно стоит обратить внимание на сложность разметки КТ исследований пациентов с COVID-19: большая площадь поражений и сложности в разграничивании радиологических паттернов требуют очень большого ручного труда радиологов. Чтобы решить эту проблему, в Botkin.AI использовали парадигму Human-AI loop. В рамках этого подхода радиолог не размечает исследование с нуля, а корректирует предсказание модели. Сами же модели обучаются итерационно на все более разнообразном корпусе данных.

Для оценки качества модели использовались как внутренние данные, собранные и размеченные самостоятельно, так и публично доступные датасеты.

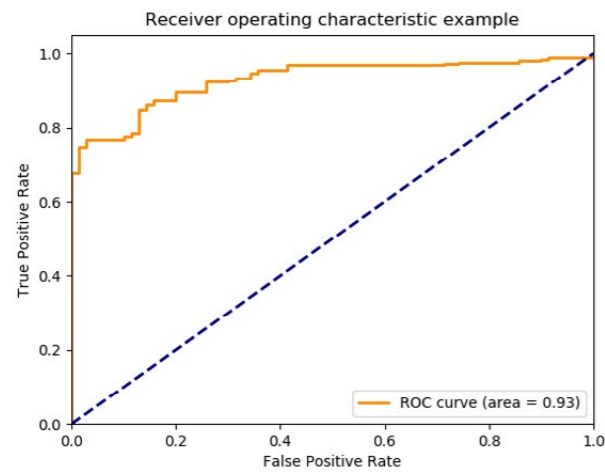
Для оценки качества локализации патологий использовался коэффициент Дайса, тестирование производилось на самостоятельно размеченном тестовом корпусе данных в количестве 50 КТ исследований. Были получены следующие значения: для матового стекла — 0.826; для консолидации — 0.80. Что свидетельствует о достаточно высоком качестве сегментирующих моделей и позволяет использовать их в практике согласно методическим рекомендациям № 43

**«Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика) ГБУЗ «научно-практический клинический центр диагностики и телемедицинских технологий департамента здравоохранения города Москвы»**

Для оценки классифицирующей модели на наличие или отсутствие патологий, в том числе ассоциированных с COVID-19, мы использовали публично доступный датасет MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset [205]. В ходе текущих испытаний была поставлена бинарная задача классификации, для чего все исследования были разбиты на два класса: класс 0 содержит все исследования категории КТ-0, класс 1 — все исследования категорий КТ-1, КТ-2, КТ-3, КТ-4. Платформа перед тестированием не калибровалась на данном наборе данных.

В ходе тестирования, были получены следующие метрики:

Рис 1. ROC-кривая метода определения патологии платформы Botkin.AI



Площадь под ROC-кривой	0.929
Удельный вес ложноотрицательных результатов	10.9%
Удельный вес ложноположительных результатов	3.1%
Чувствительность	0.859
Специфичность	0.857
Точность (Precision)	0.87
Отношение правдоподобия положительного результата	6.02
Отношение правдоподобия отрицательного результата	0.16
Прогностическая ценность положительного результата	0.955
Прогностическая ценность отрицательного результата	0.631
Частота ложных срабатываний	0.1428

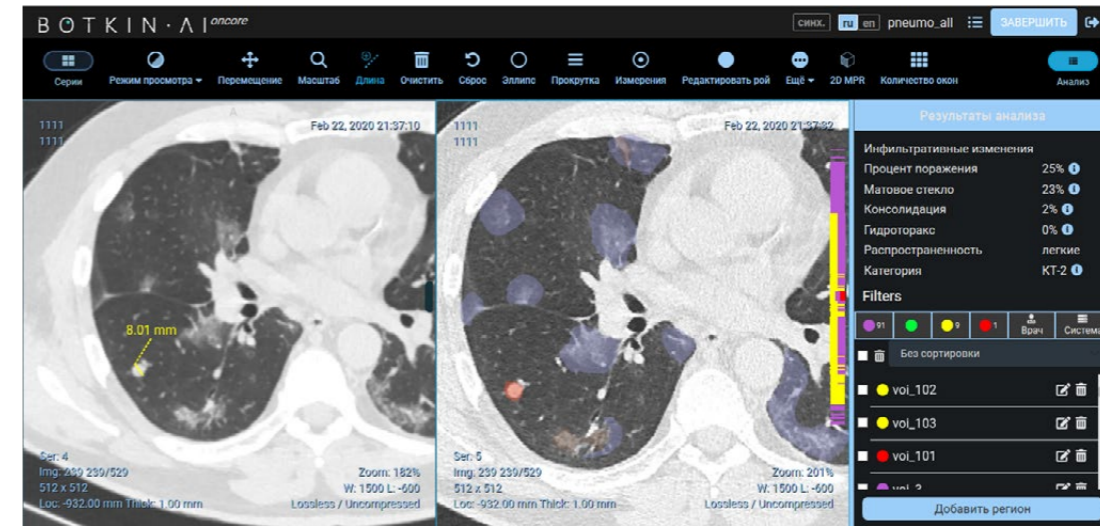


Рис. 3. КТ того же пациента, срезы на другом уровне. Правое изображение — результат обработки программы: помимо разметки КТ-синдромов, характерных для коронавирусной инфекции («матовое стекло» и консолидация), программа обнаружила округлый узел диаметром 8 мм в правом легком, подозрительный на злокачественное новообразование (размечен красным цветом). Слева на изображении без разметки: определяется периваскулярный солидный очаг, размером до 8 мм в диаметре (указан стрелкой), что соответствует категории Lung-RADS-3 и требует дальнейшего наблюдения в динамике

На рисунках 2–4 приведены примеры использования платформы «Botkin AI» при COVID-19.



Рис. 2. КТ легких больного с пневмонией COVID-19. Правая сторона: результат цветной разметки набора данных КТ с помощью ИИ в программе Botkin AI: сиреневым цветом выделены участки «матового стекла», коричневым — области консолидации, неповрежденная легочная ткань — без разметки. Левое изображение: изображение КТ легкого на соответствующем уровне без анализа программы: определяются множественные участки «матового стекла» и зоны консолидации. В таблице справа приводятся результаты количественной оценки всех выявленных изменений с присвоением классификационной категории (КТ1–4), согласно рекомендациям МЗ РФ

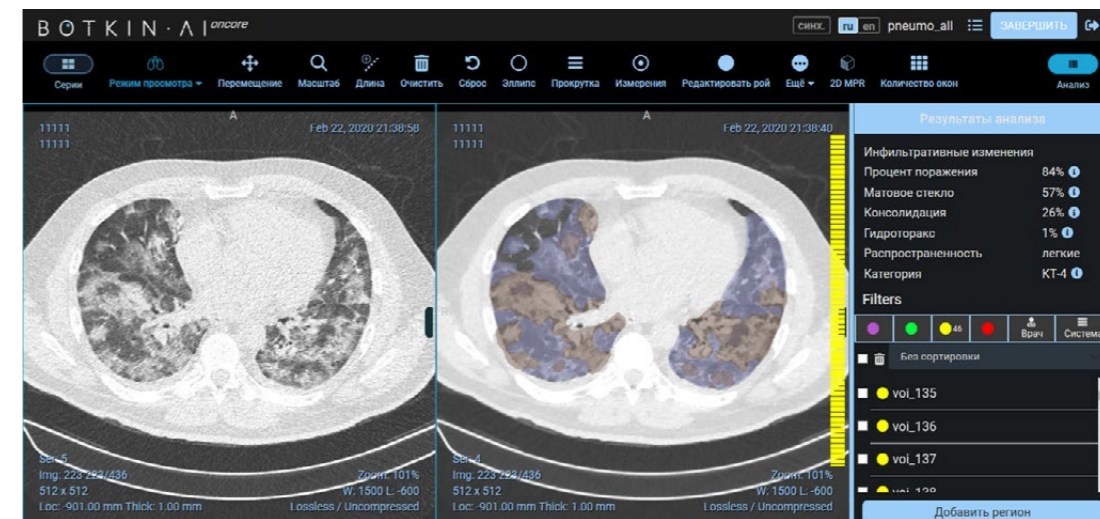


Рис. 4. Оценка в динамике данных КТ у того же пациента с COVID-19 через 10 дней (клинически — ухудшение состояния). Правое изображение размечено с помощью программы Botkin AI, левое — без обработки. В окне анализа можно увидеть результат количественной оценки видимых изменений: общий процент поражения обоих легких вырос до 84% (ранее составлял 25%), отдельно процент «матового стекла» — до 57% (ранее — 23%), консолидации — до 26% (ранее — 2%). Также появилось небольшое количество жидкости в плевральных полостях (гидроторакс)

В ходе клинического применения программы ИИ (применительно к Botkin.AI) были получены следующие результаты:

- Была показана высокая чувствительность системы ИИ при определении основных синдромов, характерных для вирусных пневмоний: по предварительным данным, она составила не менее 87%.
- Продемонстрирована возможность оценки не только паттернов изменений паренхимы легких, характерных для вирусной инфекции, но и выявления и анализа узлов в легких. С учетом того, что частота выполнения КТ легких во время пандемии увеличилась в разы, как в условиях стационара, так и при амбулаторном посещении, появилась дополнительная возможность для массового скрининга рака легкого на основании массивов накопленных данных.
- Доказана важность коррекции результатов разметки системы ИИ (при необходимости) врачом-рентгенологом. В программу встроены удобные инструменты, с помощью которых можно добавить или убрать разметку, а также подкорректировать контуры имеющейся разметки патологического объекта. Эта функция может быть использована для увеличения точности определения площади поражения легких или анализа объема узла.

- Установлено, что коррекция разметки разметка врачами-рентгенологами на этапе обучения системы ИИ позволяет минимизировать ложно-положительные результаты, например, в тех случаях, где синдромы, характерные для COVID-19 относятся с высокой долей вероятности к клинически незначимым изменениям (плевро-апикальные спайки, небольшие наложения по междолевой плевре, мелкие лимфоузлы).
- Продемонстрирована возможность оценивать в динамике изменения общей площади и объема поражения легких, соотношение «матового стекла» и консолидации, присоединение гидроторакса и оценку его объема.
- Была реализована возможность сортировки исследований по объемам поражения легких в общем списке исследований, что позволяет ранжировать исследования в списке и в первую очередь обращать внимание на наиболее тяжелые случаи (рис.4).

Следует отметить, что пациентам, которым многократно выполнялись исследования в динамике, выполнялась низкодозная КТ легких. При этом снижения чувствительности при анализе программой ИИ отмечено не было.

Подводя итог всему вышеизложенному, можно сделать следующие выводы:

- КТ зарекомендовала себя как лучший метод визуализации для выявления и количественной оценки поражений легких при COVID-19.
- КТ легких и грудной клетки обладает высокой чувствительностью для выявления признаков коронавирусного заболевания.
- Технология ИИ может оказывать существенную помощь в оценке степени тяжести COVID-19: экономит время и помогает рентгенологу быть более объективным в оценке тяжести заболевания и его прогрессирования или репрессирования при лечении.
- Подобные программы ИИ станут полезными инструментами для оценки исходов COVID-19, оценки прогноза пациентов и эффективности применения препаратов и других видов лечения в предстоящих клинических исследованиях.

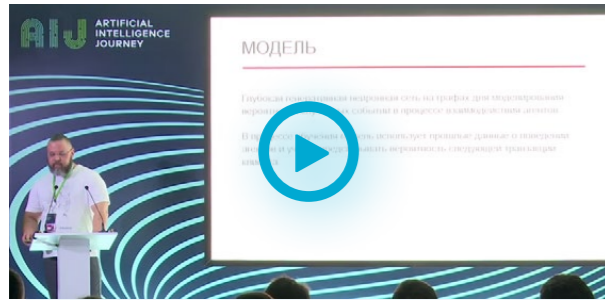
- Накопленный опыт позволит применять такие программы и для анализа состояния легких при других видах пневмоний, которые являются одним из самых частых заболеваний населения в медицинской практике.

За время пандемии COVID-19 было проведено значительное количество исследований КТ легких и грудной клетки в целях выявления больных COVID-19. Одной из возможных сфер применения технологий ИИ после пандемии COVID-19 может быть пересмотр КТ исследований в целях поиска признаков злокачественных новообразований в легких, которые могли быть пропущены в период пандемии в связи с большой нагрузкой на рентгенологов.

Рис. 4. Ранжирование списка анонимизированных исследований с помощью программы Botkin AI соответственно степени поражений легких

ID пациента	Дата загрузки	Дата исследования	Модальность	Статус патологии	Процент поражения	Состояние
1243dd	09-06-2020	27-03-2020	СТ	🔴	81.13	В работе
1204ddd	08-06-2020	26-03-2020	СТ	🟡	80.62	Завершен
1278dd	08-06-2020	27-03-2020	СТ	🔴	80.41	Новый
114538	09-06-2020	08-04-2020	СТ	🔴	80.39	Новый
1236	09-06-2020	21-05-2020	СТ	🟡	80.11	Завершен
Анонимous	11-06-2020	21-05-2020	СТ	🔴	78.28	В работе
1222	09-06-2020	26-03-2020	СТ	🟡	78.09	Завершен
1363	09-06-2020	21-05-2020	СТ	🔴	78.08	Новый

# Видео кейсы применения

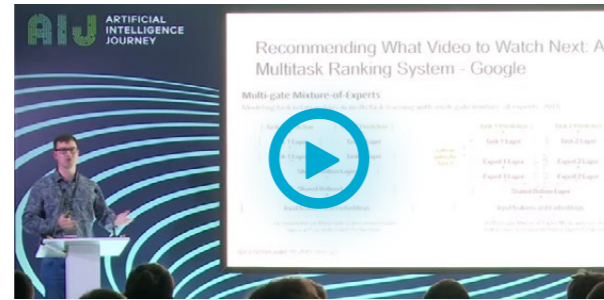


## Практика моделирования стохастических процессов в сложных системах с помощью больших данных и глубокого обучения

Рассказывает про новые методы и рекомендательную систему в приложении Сбебанка, которая предсказывает его следующую покупку и рекомендует магазины партнеров.

Марк Попов, Сбербанк, Руководитель направления исследования данных

[youtu.be/55NvRe35gYg?t=1682](https://youtu.be/55NvRe35gYg?t=1682)

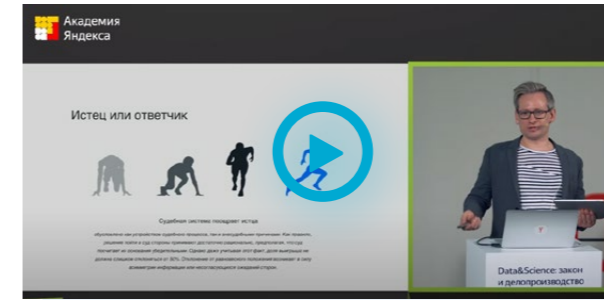


## Обзор трендов рекомендательных систем от Пульса

Пульс – аналог Дзена. Рассказал про тренды, сделал обзор каждого на примерах.

Андрей Мурашев, Mail.Ru Руководитель команды машинного обучения рекомендательных систем

[youtu.be/55NvRe35gYg?t=5013](https://youtu.be/55NvRe35gYg?t=5013)

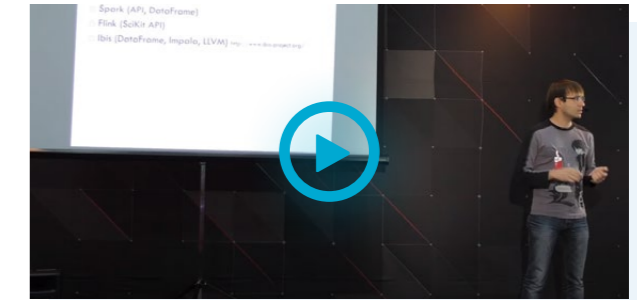


## Как Право ги верно предсказывает исход судебного дела в 82% случаев

Как может работать предиктивная аналитика в legal tech и какова вероятность верного предсказания исхода судебного дела. Показал, какие данные анализирует право.ру, чтобы составить прогноз, и рассказал о перспективах развития этого направления на рынке.

Александр Сарапин Исполнительный директор Право.ру.

[www.youtube.com/watch?v=5qxWqgV3lk0](https://www.youtube.com/watch?v=5qxWqgV3lk0)

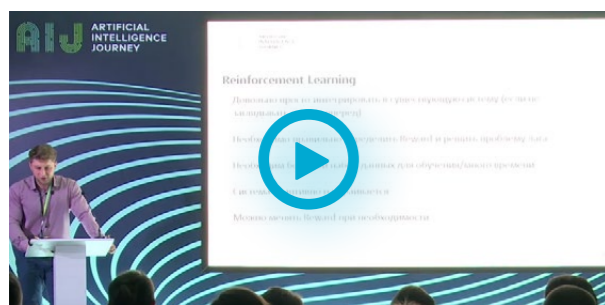


## Минимум инструментов для построения системы рекомендаций

привел несколько примеров простых решений, которые позволят построить сервис по созданию персонализированных рекомендаций с минимальными временными затратами. Технический доклад.

Алексей Дёмин / Java Server Side Developer, InData Labs

[www.youtube.com/watch?v=6VOH-sgiP\\_E](https://www.youtube.com/watch?v=6VOH-sgiP_E)

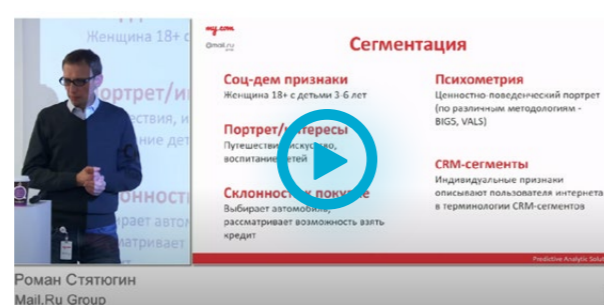


## Рекомендательные системы в ритейле

Рассказал про индивидуальные предложения для клиентов сети x5, про применение аплифта, RL.

Валерий Бабушкин, X5

[youtu.be/55NvRe35gYg?t=3094](https://youtu.be/55NvRe35gYg?t=3094)

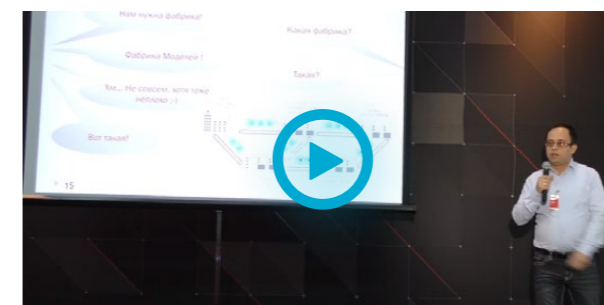


## От Big Data к предиктивной аналитике: инструкция для маркетолога

Рассказал, как они персонализируют клиентов, предсказывают покупку товара определенного пользователя с кейсами.

Роман Стятугин

[www.youtube.com/watch?v=xGYveWhbKP8](https://www.youtube.com/watch?v=xGYveWhbKP8)



## Предиктивная аналитика для управления клиентской базой

Что такое клиентская база, в чем ее ценность, а также как использовать ее в качестве основы клиентской стратегии. Рассказал также о метриках и системах КПЭ для решения задач ценностного управления.

Максим Мозговой, Director of CRM, Wargaming

[www.youtube.com/watch?v=7O1PX6qzntI](https://www.youtube.com/watch?v=7O1PX6qzntI)

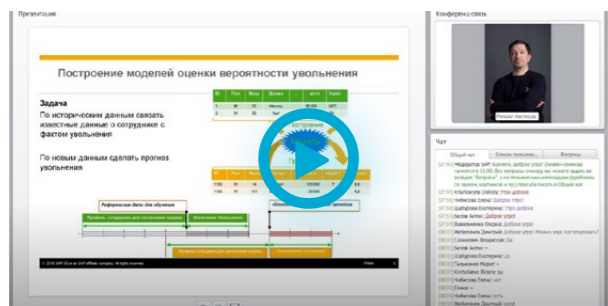


## Использование предиктивной аналитики для мониторинга состояния оборудования и ранней диагностики

Рассказала про решение PRiSM Predictive Asset Analytics. На основе исторических и текущих данных система уведомляет о будущей тревоге.

Мария Гервиц, Business Development Manager, Wonderware Russia.

[www.youtube.com/watch?v=7estlVBohGE](https://www.youtube.com/watch?v=7estlVBohGE)

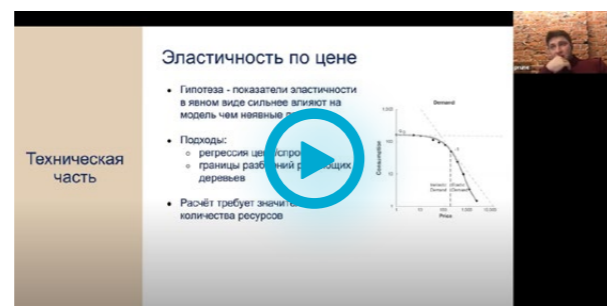


**Predictive HR – управление персоналом через прогнозирование поведения сотрудников**

Про прогнозирование поведение отдельного сотрудника, систему управления персоналом на основе знаний о будущем поведении. Демонстрация возможностей САП Предиктив

Михаил Аветисов, ведущий эксперт по предиктивной аналитике, SAP CHG.

[www.youtube.com/watch?v=5XcvKn1SqMk](http://www.youtube.com/watch?v=5XcvKn1SqMk)



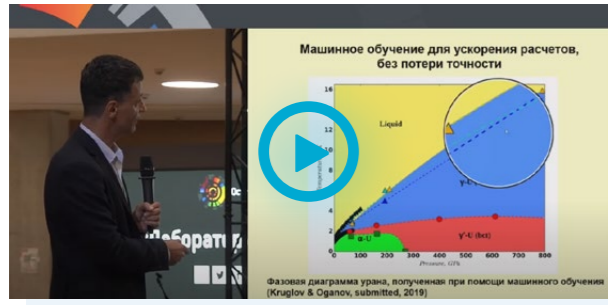
**Прогнозирование спроса в X5 Retail Group**

Зачем нужно прогнозировать спрос в сети магазинов, и как это происходит в X5 Retail Group.

Примеры гипотез, их проверка и влияние. Технический доклад

Валерий Бабушкин

[www.youtube.com/watch?v=6n1Fg8IcHU8](http://www.youtube.com/watch?v=6n1Fg8IcHU8)



**Предсказательное материаловедение: от снов Менделеева до искусственного интеллекта**

Рассказал про историю материаловедения и можно ли предсказать лучший материал из возможных соединений.

Артем Оганов

[www.youtube.com/watch?v=3ba0826HXzc](http://www.youtube.com/watch?v=3ba0826HXzc)



# Ключевые компании мира

05

# Топ компании мира

В этом разделе мы приводим список топ-компаний в мире в области предиктивной аналитики, рекомендательных систем и систем принятия решений.

В этот раз задача определения ключевых компаний оказалась еще сложнее, чем в предыдущих отчетах. Сначала мы подробно разберем основную причину, а потом объясним свой подход к выбору топ компаний. Этот подход в равной мере применим для мировых и российских компаний.

## ВНУТРИ И СНАРУЖИ

В отличие от рассмотренных нами ранее технологий (обработка естественного языка и компьютерное зрение), предиктивная аналитика и рекомендательные системы уже не являются разработкой или технологией. Они являются частью бизнеса. Причем настолько важной и существенной частью бизнеса, что большинство компаний не только не продает свои разработки на сторону, но и предпочитает не распространяться о своих результатах.

И это легко понять. В ходе эволюции те животные, которые приобрели способность хоть немного вперед прогнозировать ситуацию и поведение окружающих, получили колоссальные эволюционные преимущества. Многие исследователи считают, что главное преимущество интеллекта человека как раз и состоит в том, что человек умеет построить у себя в голове модель окружающего мира и проиграть разные варианты течения событий, а потом выбрать наиболее благоприятный. Причем, скорость этого прогнозирования и принятия решения напрямую определяет эволюционную успешность, поэтому у homo sapiens так сильно эволюционировала кора головного мозга (неокортекс), отвечающая, в том числе, за эти задачи.

Очевидно, что такая же способность в бизнесе является колоссальным конкурентным преимуществом. Компания, которая правильно прогнозирует движение рынка или поведение своих клиентов всегда будет на шаг впереди других компаний. И будет оберегать свое преимущество.

Поэтому, проанализировав область предиктивной аналитики мы пришли к нескольким наблюдениям, которые привели нас к любопытной классификации.

## НАБЛЮДЕНИЯ

1. Чем успешнее компания применяет предиктивную аналитику и системы принятия решений — тем больше она вкладывает в это направление и тем меньше о нем рассказывает. Предельным случаем, очевидно, является биржевая торговля, в которой сотни компаний используют ПА и СППР для автоматической торговли, но нет ни одной публикации или доклада на конференции с какими-то сильными результатами.
2. Чем больше компания выросла за последние 10-15 лет и чем сильнее в нее интегрированы онлайн платформы сбора данных и коммуникаций с клиентами — тем выше вероятность, что она успешно использует ПА и СППР.
3. Успешные компании стараются не продавать свои алгоритмы и модели, а создавать на их базе сервисы и продавать эти сервисы. Тем самым они позволяют другим компаниям отчасти разделить успешность правильных предсказаний, но не отдают им ключевые компетенции. Каждая такая компания старается выстроить целую экосистему на своих сервисах.

## ПРОМЕЖУТОЧНЫЙ ВЫВОД

В результате, для области ПА и СППР для наших целей мы выработали новую классификацию компаний. Мы не можем опираться на финансовые показатели, потому что тогда это будет рейтинг крупнейших компаний по стоимости или выручке. Мы не можем опираться на открытую информацию о разработках компаний, потому что в данной области ее мало, компании ведут свои разработки в условиях строгой секретности и мало публикуются и регистрируют патенты.

# Классификация

## ЛИДЕРЫ ЭКОСИСТЕМЫ

Это компании, которые построили у себя внутри системы ПА и СППР и успешно их используют, развивая свой бизнес. При этом они не продают наружу напрямую эти технологии или компетенции, но продают их в форме сервисов, сформулированных как правило в виде «облака». Это компании у которых есть доступ к большим массивам данных и хорошие возможности по их обработке. Они лидируют в вопросах разработки новых решений. Сюда мы относим, например, Google и Facebook.

## ПРЕДСКАЗАТЕЛИ ДЛЯ СЕБЯ

Компании, которые построили у себя внутри системы ПА и СППР и успешно их используют, развивая свой бизнес. При этом они совсем не продают наружу технологии или решения, используя собственные разработки только для себя, чтобы стать еще успешнее. Сюда мы относим например Netflix и Alibaba или например крупные банки. Как правило это уже большие компании, которые сильно выросли за последние 10–15 лет благодаря активному использованию данных и работе с ними.

## РАЗРАБОТЧИКИ

Есть компании, которые исторически занимаются разработкой софта для бизнеса и нарабатывали у себя компетенции по ПА и СППР. В первую очередь это SAS, SAP и IBM. Теперь к ним добавились еще и множество маленьких компаний, делающих нишевые решения для разных бизнесов. По нашему мнению, эта категория компаний обречена на нишевый рынок с не очень доходными клиентами, потому что все крупные компании осознали ценность своих данных и собственных прогнозов и теперь стремятся реализовать компетенции ПА и СППР внутри, в первую очередь скупая маленькие компании. Тем не менее, бизнес разработчиков в этой области будет стабильно расти и развиваться, ведь все больше компаний захотят использовать ПА и СППР для своих задач, а возможности формировать внутренние команды есть у всех.

## ИНТЕГРАТОРЫ

Это компании, чьими клиентами являются крупные и средние компании. У них есть непосредственный контакт с клиентами и высокий уровень доверия. У них не всегда есть собственные компетенции в ПА и СППР, но они могут найти разработчика под заказ. Такие интеграторы играют свою роль в формировании рынка ПА и СППР. Сюда мы относим, например, аудиторские компании большой четверки — например PwC или Deloitte, а также системных интеграторов — например Крок или Ланит.

## СПИСОК ТОП-10

В рамках построенной нами классификации можно выделить достаточно много сильных компаний. Мы выбрали наиболее известные компании из каждого сегмента. Про каждую из них мы далее приводим краткую справку.

# Топ компании мира

На основании экспертных оценок

Топ-10 компаний мира	Отрасль
1. SAS	Бизнес-аналитика
2. SAP	Бизнес-аналитика
3. IBM	Бизнес-аналитика
4. Google	Интернет
5. Facebook	Интернет
6. Amazon	Ритейл
7. Netflix	Медиа
8. Alibaba	Ритейл
9. Baidu	Интернет
10. Huawei	Телекоммуникации

## SAS

**Описание:** Компания SAS является крупнейшей в мире частной IT-компанией, специализирующейся на разработке и продаже решений и услуг в области бизнес-аналитики. В России и странах СНГ компания SAS начала работу в 1996 году. Заказчикам компания SAS предлагает полный спектр решений и услуг в области бизнес-аналитики: консалтинг, внедрение, обучение и техническую поддержку. Клиентами SAS в России и СНГ являются все 10 крупнейших российских банков (Сбербанк России, ВТБ, Газпромбанк и др.), РЖД, «Аэрофлот», лидеры рынка розничной торговли (X5 Retail Group, Азбука Вкуса и др.), крупнейшие компании из телекоммуникационного и топливно-энергетического секторов, государственные организации.

**Руководитель компании:** Джеймс Гуднайт — CEO.

### Руководители направления PA:

Saratendu Sethi Head of Artificial Intelligence & Machine Learning at SAS.

Tamara Fischer Operationalizing Predictive Analytics | Senior Analytics Expert at SAS,

Alexander Efimov Advanced Analytics & Industry Solutions Lead, Russia & CIS at SAS.

### Основные направления исследований в PA:

Интеллектуальный анализ данных, Статистический анализ, Прогнозирование и оптимизация, Текстовая аналитика, Optimization & Simulation.

### Некоторые PA-проекты:

- **SAS® Enterprise Miner™** — Создание предсказательных и описательных аналитических моделей, в т.ч. на основе больших данных.

- **SAS® Econometrics** — Анализируйте, описывайте и прогнозируйте проблемы, связанные с экономическими и финансовыми системами, применяя комплексные математические и статистические методы.
- **SAS® Simulation Studio** — Создание моделей, имитирующих сложные реальные системы, посредством инструментов дискретно-событийного моделирования позволит лучше понять и оптимизировать работу этих систем.
- **SAS Real-Time Decision manager** — система поддержки принятия решений, разработанная SAS для различных бизнес-применений. Используется в финансах и страховании, промышленности и энергетике, государственном секторе и здравоохранении для автоматизации и масштабирования процессов принятия решений, риск-аналитики, автоматизации отчетности и др.

- **AIoT** — объединение искусственного интеллекта и интернета вещей, ИИ-решения для получения данных от миллиардов IoT устройств в реальном времени.
- **Kids' Cancer Project** — некоммерческий проект, в основе которого лежат облачные технологии на базе предсказательной аналитики, применяемые для таргетирования спонсорской аудитории и увеличения пожертвований, направленных на борьбу с онкологическими заболеваниями среди детей. За 1 год данный проект позволил увеличить пожертвования на 34%.



**\$3,1 млрд**  
Выручка <sup>[216]</sup>

**\$374.93 МЛН**  
Капитализация <sup>[217]</sup>

**\$1 млрд**  
Затраты на RnD на 3 года <sup>[216]</sup>

**172**  
Патентов <sup>[218]</sup>

**133**  
Публикаций <sup>[218]</sup>

## SAP

**Описание:** Ведущий мировой провайдер бизнес-приложений для предприятий, IT-сервисов, аналитических, облачных и мобильных решений, а также технологий для баз данных.

**Руководитель компании** — Дженнифер Морган (со-CEO), Кристиан Кляйн (со-CEO).

### Руководители направления PA:

Anirban Banerjee Product Owner at SAP Predictive Analytics.

Jesper Schleimann Chief Innovation Officer at SAP.

### Основные направления исследований в PA:

Глубокое обучение, обучение с минимальным подкреплением, сентимент-анализ, интерпретируемое машинное обучение, пристрелочная классификация, поиск и извлечение информации.

### Некоторые PA-проекты:

- **SAP Analytics Cloud** — облачное программное обеспечение для прогнозной аналитики работает вместе с инструментами BI и планирования в SAP Analytics Cloud — можно находить, визуализировать, планировать и прогнозировать в контексте. Используйте технологию оперативной памяти и машинное обучение, чтобы в реальном времени находить релевантные прогнозные данные.

- **SAP Data Intelligence** — Предоставляет пользователям интеллектуальные, релевантные контекстуальные сведения с помощью обработки данных в вашей IT-среде. Обнаружение и подключение нескольких типов данных независимо от того, где они находятся. Уточнение и повторное использование потоков аудио, изображений и видео, а также данных с устройств на основе Интернета вещей. Оптимизируйте управление и минимизируйте риск соответствия с надежными правилами управления метаданными Управление и выполнение модульных конвейеров данных в распределенных инфраструктурах.
- **SAP Cash Application** — система на основе алгоритмов машинного обучения для автоматической обработки платежей, автоматизации процессов сверки счетов и сопоставления критериев исторических периодов.
- **SAP Intelligent Robotic Process Automation** — интеллектуальные боты на основе алгоритмов машинного обучения для автоматизации рутинных бизнес-процессов (операционной деятельности, планирования и распределения ресурсов и др.).



**\$31.135 млрд**  
Выручка <sup>[213]</sup>

**\$163,2 млрд**  
Капитализация <sup>[214]</sup>

**\$5 млрд**  
Затраты на R&D <sup>[213]</sup>

**н/д**  
Патентов

**43**  
Публикаций <sup>[215]</sup>



## IBM

**Описание:** Один из крупнейших в мире производителей и поставщиков компьютеров и программного обеспечения, а также IT-сервисов и консалтинговых услуг. Компания применяет различные NLP-технологии, а также разрабатывает модели для поиска и извлечения информации из неструктурированных данных на базе своей экспертной системы Watson.

**РА — подразделения:** IBM Research China, IBM Research Israel, IBM Research Ireland, IBM Research Almanden, IBM Research Austin, IBM Research Cambridge.

**Основные направления исследований:** вопросно-ответные системы, поиск и извлечение информации из текста, машинный перевод, обработка естественного языка.

### Лидеры в РА:

Rob High (Роб Хай) — вице-президент и СТО IBM Watson.

Dario Gil (Дарио Хиль) — директор IBM Research.

Jeffrey Welser — директор IBM Research Almanden.

Oded Cohn — директор IBM Research Israel.

Kevin Nowka — директор IBM Research Austin.

### Некоторые РА проекты:

- **IBM Decision Optimization** — Это семейство программных продуктов для анализа и оптимизации бизнес-процессов
- **IBM Watson** — Экспертная система на основе алгоритмов машинного обучения, предоставляющая организациям из разных отраслей широкий набор возможностей по работе с большими данными — их агрегации, подготовке и анализу. На базе Watson в нескольких штатах США созданы системы принятия врачебных решений, обеспечивающие более высокие качество и безопасность лечения пациентов.
- **IBM Cloud Pak для данных** — Облачная платформа для предсказательной аналитики, которая помогает собирать, систематизировать и анализировать данные.
- **IBM SPSS** — программное обеспечение для статистической обработки данных. Используется в финансах, рекламе, ритейле и многих других отраслях.



**\$112B**  
Капитализация <sup>[222]</sup>

**\$77B**  
Выручка <sup>[223]</sup>

**\$45B**  
R&D <sup>[222]</sup>

**5071**  
Публикаций <sup>[224]</sup>

**5563**  
Патентов <sup>[224]</sup>

## Netflix

**Описание:** Крупнейший поставщик медиаконтента (фильмов, сериалов) на основе стриминговых видеосервисов, уникален своей бизнес-моделью на основе рекомендательных ИИ-алгоритмов.

**Руководитель компании** — Рид Хастингс (председатель и CEO),

**Руководители направления РА:** Tony Jebara (Тони Джебара) — директор по машинному обучению Netflix Research, руководитель лаборатории машинного обучения Columbia University.

**Основные исследования в РА:** Рекомендательные системы для пользователей сервиса онлайн платформы, машинное обучение, анализ поведения пользователей, машинное обучение для сегментирования пользователей и когортного анализа.

### Некоторые РА-проекты:

- **Рекомендательный системы** — алгоритмы машинного обучения для персонализации пользователей и создания автоматических рекомендаций фильмов и сериалов на основе персональных данных и ранее понравившегося контента.
- **Платформа для анализа поведения пользователей и автоматического выявления поведенческих аномалий.**
- **Определение местоположения для кинопроизводства (Pre-Production)** — Анализ данных, для определения, где и когда лучше всего арендовать съемочную площадку — с учетом ограничений планирования (времени актеров / съемочной группы), бюджета (место проведения, расходы на перелет / гостиницу) и производственных требований (дневная или ночная съемка, вероятность неблагоприятных погодных явлений).

## NETFLIX

**\$21.403 млрд**  
Выручка <sup>[219]</sup>

**\$196.05 млрд**  
Капитализация <sup>[220]</sup>

**\$1.8 млн**  
Затраты на RnD <sup>[219]</sup>

**Н/д**  
Патентов

**21**  
Публикаций <sup>[221]</sup>

## Baidu

**Описание:** Крупнейшая поисковая система в Китае и 4-й самый посещаемый сайт в мире. Baidu предлагает множество сервисов, аналогичных Google и Яндекс, а также поддерживает Baidu Baike — китайский аналог Wikipedia. Основные доходы компания получает от таргетированной рекламы. В группе компаний Baidu исследованиями занимается Baidu Research, в которой в свою очередь есть несколько исследовательских подразделений, которые занимаются CV — Institute of Deep Learning, Robotics and Autonomous driving Lab, Silicon valley AI Lab.

**Руководитель компании:** Робин Ли — CEO Baidu.

**Руководители направления PA:** Wang Haifeng — CTO компании, Ping Li — глава лаборатории когнитивных вычислений, Dejing Dou — глава лаборатории больших данных.

**Основные направления исследований в PA:**

- **Алгоритмы глубокого обучения для работы с большими данными,** математические методы обработки больших данных, поиск и извлечение неструктурированных данных. Некоторые PA-проекты: Baidu создали аналитическую систему для моделирования распространения ВИЧ, сифилиса, гонореи

среди населения и предсказания новых случаев заболеваний. Во время пандемии SARS-Covid-19 эта система была использована для борьбы с распространением вируса.

- **Виртуальная клавиатура от Baidu,** обучающаяся на конкретном пользователе, способна предсказывать слова и фразы с вероятностью 91.6% — один из самых высоких показателей в мире.
- **Алгоритмы Baidu используются для предсказания туристических потоков** внутри Китая во время долгих праздников и выходных.
- **Ultra-high-bandwidth DCI network** — Baidu совместно с Huawei создали сверхмощную сеть ЦОД, приуроченную к ежегодному празднику CCTV Spring Festival Gala 2019. В этот день участники праздника обменивались друг с другом и с людьми по всему миру виртуальными «красными пакетами», символизирующими традиционный в этот день в Китае обмен подарками в красной упаковке между близкими людьми. Суммарное число запросов в этот день достигло 20,8 миллиардов, при этом максимальная частота запросов составляла 1 миллиард/минуту. Для распределения нагрузки на ЦОДы и онлайн-мониторинга их состояния использовались алгоритмы ПА.



**\$15.4B**  
Выручка <sup>[225]</sup>

**\$40.7B**  
Капитализация <sup>[226]</sup>

**\$2.6B**  
Затраты на RnD <sup>[225]</sup>

**36**  
Патентов <sup>[227]</sup>

**199**  
Публикаций <sup>[227]</sup>

## Alibaba Group

**Описание:** Китайский транснациональный холдинг, содержащий множество компаний, специализирующийся на e-commerce, ритейле, ИИ и технологиях. Помимо торговли выстраивает целую экосистему платежей, логистики и сервисов.

Основные статьи дохода — B2B торговые операции и розничная онлайн-торговля. У компании есть иностранные акционеры — японский Softbank (примерно 29% акций) и американская Altaba (дочерняя компания Yahoo!, примерно 15% акций). Компания инвестирует не только в Китае, но также в США, Индии, Израиле, Тайване, Индонезии, Пакистане, Сингапуре, ОАЭ и Швейцарии. Область инвестиций включает биотех, безопасность, пищевые технологии, медиа, облачные вычисления, финтех и технологии распознавания голоса и лиц.

**Руководитель компании:** Jack Ma (Джек Ма) — президент Alibaba Group,

**Руководители направления PA:** Jeff Zhang (Джефф Жанг) — CTO компании, Dr. Hui Xue директор отдела алгоритмов и анализа данных в Alibaba Group, Yangqing Jia — вице-президент Alibaba Group, Yangqing Jia (Янцин Цзя) — директор по ИИ в Alibaba, ранее работал в Facebook над ИИ-проектами (Caffee2), Jingren ZhouHe — глава лаборатории аналитики данных.

**Основные направления исследований в PA:** рекомендательные системы, машинное обучение в ПА, новые алгоритмы data mining, распределенные гетерогенные системы хранения и обработки больших данных, вычисления на графах.

**Некоторые PA-проекты:**

- **Alibaba Data Analytics Solutions** — набор инструментов для хранения, предобработки, визуализации больших данных и построения предсказательных моделей на их основе.
- **Алгоритмы для борьбы с кибермошенничеством** — Alibaba постоянно подвергается огромному количеству кибератак, поэтому кибербезопасность — одно из приоритетных направлений исследований компании. Alibaba ежедневно идентифицирует десятки миллионов подозрительных устройств по их трафику.
- **Alibaba Cloud** — облачная платформа для хранения и работы с большими данными для решения задач бизнес-аналитики.
- **Tmall** — торговая онлайн-площадка, на которой алгоритмы предсказательной аналитики используются для продвижения товаров и брендов, анализа спроса, сегментирования потребителей и создания персональных рекомендаций.



**\$56.1 B**  
Выручка <sup>[228]</sup>

**\$638.3B**  
Капитализация <sup>[229]</sup>

**\$890M**  
Затраты на RnD <sup>[228]</sup>

**171**  
Патентов <sup>[230]</sup>

**189**  
Публикаций <sup>[230]</sup>

## Google

Миссия Google состоит в том, чтобы организовать мировую информацию и сделать ее универсально доступной и полезной. Благодаря таким продуктам и платформам, как Поиск, Карты, Gmail, Android, Google Play, Chrome и YouTube, Google играет значительную роль в повседневной жизни миллиардов людей и стала одной из самых широко известных компаний в мире. Google является дочерней компанией Alphabet Inc.

### Исследовательские подразделения:

Google AI, Google Brain. Также исследовательским подразделением является DeepMind (дочерняя компания Google), которая ведет исследования в области AGI (сильного ИИ). У DeepMind очень много публикаций в области AI/ML, в ней работают несколько ведущих исследователей по ML в мире.

**Руководитель компании:** Сундар Пичаи – CEO Google.

**Руководитель PA направления:** Ashwin Ram Technical Director of AI.

**Основные направления исследований:** рекомендательные системы, предсказательная аналитика в рекламе.

### Некоторые PA-проекты:

- **Google AI Platform** — набор инструментов, позволяющих ML разработчикам, специалистам по данным и инженерам данных легко и быстро выполнять свои проекты. На платформе доступен большой выбор методов хранения, обработки и анализа больших данных.
- **BigQuery ML** — платформа, которая позволяет исследователям и аналитикам данных за считанные минуты создавать и внедрять модели машинного обучения на массивных структурированных или полуструктурированных наборах данных.
- **Рекомендательные системы на базе Google Ads** — на основе анализа пользовательского трафика позволяют маркетологам запускать эффективные рекламные кампании.



**\$160.7B**

Выручка <sup>[240]</sup>

**\$1000B**

Капитализация <sup>[241]</sup>

**\$26B**

Затраты на RnD <sup>[240]</sup>

**428**

Патентов <sup>[242]</sup>

**1888**

Публикаций <sup>[242]</sup>

## Facebook

**Описание:** самая большая группа социальных сетей в мире, в портфеле компании следующие платформы: Facebook, Instagram, Whatsapp. Основной доход компания получает от рекламы. Алгоритмы машинного обучения применяются во всех продуктах Facebook. Исследовательское подразделение в ИИ — Facebook AI Research (FAIR).

**Руководитель компании** — Марк Цукерберг — CEO Facebook.

**Руководитель PA направления** — Mike Schroepfer Chief Technology Officer at Facebook.

**Основные исследования в PA:** кибербезопасность, рекомендательные системы, алгоритмы машинного обучения, направленные на решение задач предсказательной аналитики.

### Некоторые PA-проекты:

- **Facebook Analytics** — инструмент для анализа поведения людей на различных платформах. Позволяет понять, как люди взаимодействуют с компанией на сайте, в приложении, на Странице Facebook и в других ресурсах и помогает оптимизировать рекламные кампании.
- **I Voted** — инструмент для анализа политической активности пользователей и предсказания результатов выборов на основе данных профилей в Facebook.
- **Topic Data** — технология Facebook, которая показывает маркетологам ответы аудитории касательно различных брендов, событий и мероприятий таким образом, чтобы личная информация пользователей оставалась конфиденциальной. Маркетологи используют данную информацию для оптимизации рекламных кампаний.



**\$70.7B**

Выручка <sup>[234]</sup>

**\$629.04B**

Капитализация <sup>[235]</sup>

**\$13,6B**

Затраты на RnD <sup>[234]</sup>

**43**

Патентов <sup>[236]</sup>

**1082**

Публикаций <sup>[236]</sup>

## Amazon

**Описание:** Транснациональная компания, профилирующаяся в электронной коммерции (маркетплейс, аукционы), облачных вычислениях (AWS) и искусственном интеллекте. Компания ведет разработки по всем дисциплинам, относящимся к термину «искусственный интеллект».

**Руководитель компании:** Джефф Безос — CEO компании.

**Руководитель PA направления:** Вернер Фогельс — СТО компании.

### Некоторые PA-проекты:

- **Collaborative filtering engine (CFE)** — технология, используемая в рекомендательных системах компании. Позволяет эффективно увеличивать продажи за счет персональных рекомендаций на основе анализа предыдущих покупок, вишлиста и трафика пользователей.

- **Kindle Highlighting** — рекомендательная система и социальная сеть на основе Amazon Kindle. Позволяет пользователям подчеркивать фразы, понравившиеся в прочитанных книгах. На основе этих фраз строит рекомендации новых книг.
- **Anticipatory Shipping Model** — запатентованная Amazon модель, позволяющая с использованием алгоритмов ПА прогнозировать, какие продукты, где и когда пользователи будут заказывать. Товары, которые предсказала модель, автоматически отправляются в ближайший распределительный центр или на склад, чтобы в нужное время быть готовыми к отправке.
- **Amazon Web Services** — облачные инструменты для хранения, анализа и визуализации больших данных.



\$280.5B

Выручка <sup>[237]</sup>

\$1000B

Капитализация <sup>[238]</sup>

\$37.3B

Затраты на RnD <sup>[237]</sup>

252

Патентов <sup>[239]</sup>

200

Публикаций <sup>[239]</sup>

## Huawei

**Описание:** Одна из крупнейших мировых компаний в сфере телекоммуникаций, производитель сетевого оборудования, смартфонов, планшетных компьютеров, терминалов, мобильных приложений и ПО.

**Руководитель компании** — Жэнь Чжэн-фэй (CEO).

**Руководители направления PA:** Александр Филиппов — руководитель лаборатории искусственного интеллекта московского исследовательского центра Huawei, William McColl — директор по исследованиям Huawei.

**Основные направления исследований в PA:** Высокопроизводительные вычисления, глубокое обучение, анализ больших объемов фото и видео, облачные технологии для высокопроизводительных вычислений.

### Некоторые PA-проекты:

- **Huawei разработала и** представила первую в мире библиотеку обработки потоковых данных с открытым исходным кодом на базе Spark.
- **Yunlu + Huawei Cloud EI.** Система, разработанная компаниями Yunlu Technology и Huawei, предназначена для умных городов. Она обеспечивает мониторинг состояния строительных конструкций, выявляет повреждения, обнаруживает потенциальные угрозы безопасности и обеспечивает техническую поддержку отделов строительства и технического обслуживания.
- **Huawei предоставляет интегрированное решение на** основе сервера Atlas AI и программного обеспечения ISV для дорожной полиции Шэньчжэня. Данное решение позволяет автоматизировать составление дорожного расписания, автоматически агрегировать и обрабатывать видеопотоки с дорожных камер.



\$123B

Выручка <sup>[231]</sup>

\$350M

Капитализация <sup>[232]</sup>

\$19B

R&D <sup>[231]</sup>

767

Патентов <sup>[233]</sup>

1400

Публикаций <sup>[233]</sup>

A collection of light blue, thin-lined geometric shapes, including triangles and polygons, scattered in the upper right corner of the page.

# Ключевые компании России

06

# Топ компании России

В этом разделе мы приводим список топ-компаний в России в области предиктивной аналитики, рекомендательных систем и систем принятия решений.

В этот раз задача определения топ-компаний оказалась еще сложнее, чем в предыдущих отчетах. Сначала мы подробно разберем основную причину, а потом объясним свой подход к выбору топ компаний.

## ВНУТРИ И СНАРУЖИ

Отличие от рассмотренных нами ранее технологий (обработка естественного языка и компьютерное зрение), предиктивная аналитика и рекомендательные системы уже не являются разработкой или технологией. Они являются частью бизнеса. Причем настолько важной и существенной частью бизнеса, что большинство компаний не только не продает свои разработки на сторону, но и предпочитает не распространяться о своих результатах.

И это легко понять. В ходе эволюции те животные, которые приобрели способность хоть немного вперед прогнозировать ситуацию и поведение окружающих, получили колоссальные эволюционные преимущества. Многие исследователи считают, что главное преимущество интеллекта человека как раз и состоит в том, что человек умеет построить у себя в голове модель окружающего мира и проиграть разные варианты течения событий, а потом выбрать наиболее благоприятный. Причем, скорость этого прогнозирования и принятия решения напрямую определяет эволюционную успешность, поэтому у homo sapiens так сильно эволюционировала кора головного мозга (неокортекс), отвечающая, в том числе, за эти задачи.

Очевидно, что такая же способность в бизнесе является колоссальным конкурентным преимуществом. Компания, которая правильно прогнозирует движение рынка или поведение своих клиентов всегда будет на шаг впереди других компаний. И будет оберегать свое преимущество.

Поэтому, проанализировав область предиктивной аналитики мы пришли к нескольким наблюдениям, которые привели нас к любопытной классификации.

## НАБЛЮДЕНИЯ

1. Чем успешнее компания применяет предиктивную аналитику и системы принятия решений — тем больше она вкладывает в это направление и тем меньше о нем рассказывает. Предельным случаем, очевидно, является биржевая торговля, в которой сотни компаний используют ПА и СППР для автоматической торговли, но нет ни одной публикации или доклада на конференции с какими-то сильными результатами.
2. Чем больше компания выросла за последние 10–15 лет и чем сильнее в нее интегрированы онлайн платформы сбора данных и коммуникаций с клиентами — тем выше вероятность, что она успешно использует ПА и СППР.
3. Успешные компании стараются не продавать свои алгоритмы и модели, а создавать на их базе сервисы и продавать эти сервисы. Тем самым они позволяют другим компаниям отчасти разделить успешность правильных предсказаний, но не отдают им ключевые компетенции.

## ПРОМЕЖУТОЧНЫЙ ВЫВОД

В результате, для области ПА и СППР мы предлагаем новую классификацию компаний. Мы не можем опираться на финансовые показатели, потому что тогда это будет рейтинг крупнейших компаний по стоимости или выручке. Мы не можем опираться на открытую информацию о разработках компаний, потому что в данной области ее мало, компании ведут свои разработки в условиях строгой секретности и мало публикуются и регистрируют патенты.

# Классификация

## ЛИДЕРЫ ЭКОСИСТЕМЫ

Это компании, которые построили у себя внутри системы ПА и СППР и успешно их используют, развивая свой бизнес. При этом они ничего не продают наружу, за исключением сервисов, сформулированных как правило в форме «облака». Сюда мы относим, например, Яндекс. Это компании у которых есть доступ к большим массивам данных и хорошие возможности по их обработке.

## ПРЕДСКАЗАТЕЛИ ДЛЯ СЕБЯ

Компании, которые построили у себя внутри системы ПА и СППР и успешно их используют, развивая свой бизнес. При этом они ничего не продают наружу, используя собственные разработки чтобы стать еще успешнее. Сюда мы относим например Озон и Сбербанк. Как правило это уже крупные компании, которые сильно выросли за последние 10–15 лет благодаря активному использованию данных и работе с ними.

## РАЗРАБОТЧИКИ

Это как правило начинающие маленькие компании, которые имеют у себя компетенции по ПА и СППР и пытаются их продавать. По нашему мнению, эта категория компаний обречена на нишевый рынок с не очень доходными клиентами, потому что все крупные компании осознали ценность своих данных и собственных прогнозов и стремятся реализовать компетенции ПА и СППР внутри, в том числе скупая такие маленькие компании или (что чаще) переманивая к себе специалистов. Сюда можно отнести, например, CleverData или Fidesys.

## ИНТЕГРАТОРЫ

Это компании, чьими клиентами являются крупные и средние компании. У них есть непосредственный контакт с клиентами и высокий уровень доверия. У них не всегда есть собственные компетенции в ПА и СППР, но они могут найти разработчика под заказ. Такие интеграторы играют свою роль в формировании рынка ПА и СППР. Сюда мы относим, например, аудиторские компании большой четверки — например PwC или Deloitte, а также системных интеграторов — например Крок или Ланит.

## СПИСОК ТОП-10

В рамках построенной нами классификации можно выделить достаточно много сильных компаний. Чтобы провести точку отсечения, мы решили выбрать по одной компании из каждой отрасли и на ее примере показать, как можно использовать ПА и СППР в этом бизнесе.

Все российские компании активные в области искусственного интеллекта есть на карте Искусственного интеллекта России <http://aiRussia.online>.

В результирующий список вошло 10 компаний, которые на наш взгляд являются лидерами в области ПА в своей отрасли в России. Про каждую из них мы далее приводим краткую справку.

# Топ компании России

На основании экспертных оценок.

Топ-10 компаний России	ИИ-профиль
1. Яндекс	Лидеры экосистемы
2. Mail.Ru Group	Лидеры экосистемы
3. Лаборатория Касперского	Пользователи
4. Сбербанк	Пользователи
5. МТС	Пользователи
6. X5 Retail Group	Пользователи
7. Газпромнефть	Пользователи
8. Цифра	Разработчики
9. Крок	Интеграторы
10. Ланит	Интеграторы

## Яндекс

Ведущая поисковая система в интернете в России по количеству обрабатываемых поисковых запросов. Также, как и в прошлом Альманахе, Яндекс оказался на лидирующих позициях в наших рейтингах, так как с самого начала он был построен на технологиях ИИ и является лидирующей российской компанией в этой области.

Основной фокус разработок в ИИ направлен на улучшение качества поиска, а для этого — на все более тщательное изучение пользователей и данных о них. Поэтому Яндекс развивает целую серию проектов, предоставляющих услуги пользователям — и собирающих данные о них и их предпочтениях, например Яндекс.Музыка. Яндекс.Еда, Яндекс.Дзен, Яндекс.Такси и другие. Например, популярный сервис персональных рекомендаций Яндекс.Дзен уже имеет аудиторию более 10 млн. чел, его основной задачей является рекомендация пользователям статей, которые их вероятно заинтересуют.

**РА подразделения:** предсказательная аналитика применяется в большинстве подразделений Яндекса, в частности: Яндекс.Музыка. Яндекс.Еда, Яндекс.Дзен и другие.

**Лидер в ИИ:** Михаил Биленко — руководитель управления машинного интеллекта и исследований.

**Ведущие исследователи в РА:** Евгений Соколов, Виктор Ламбурт.

**Основные направления исследований:** рекомендательные системы, предиктивная аналитика.

**Некоторые РА проекты:**

- **Яндекс.Дзен:** это сервис персональных рекомендаций и издательская платформа. Дзен постоянно изучает интересы читателей и подбирает для каждого индивидуальную ленту рекомендаций.
- **Яндекс.Такси,** динамическое ценообразование: прогнозируют высокий спрос, привлекают дополнительных водителей и корректируют стоимость заказа.
- **Clickhouse** — высокопроизводительная аналитическая СУБД с открытым исходным кодом. Система быстро отвечает на аналитические запросы по обновляемым в реальном времени данным. Clickhouse способна масштабироваться до десятков триллионов записей и петабайтов данных, сохранять весь поток данных без предварительной агрегации и быстро получать отчеты в любых разрезах.
- **Рекомендательные системы** — внедряются во все сервисы компании Яндекс (поисковик Яндекс, Яндекс.Музыка. Яндекс.Еда, Яндекс.Дзен и другие), на основе запросов пользователей и данных о посещаемых ими web-страницах строят кастомизированные персонализированные рекомендации товаров и услуг.

## Яндекс

\$2,224B

Выручка <sup>[249]</sup>

\$21,5B

Капитализация <sup>[250]</sup>

\$370M

Затраты на RnD <sup>[249]</sup>

9

Патентов <sup>[251]</sup>

55

Публикаций <sup>[251]</sup>

## Mail.ru Group

Mail.ru Group — одна из крупнейших интернет- и IT-компаний в России. В экосистему Mail.ru Group входят ведущие русскоязычные социальные сети — ВКонтакте и Одноклассники, лидирующий почтовый сервис, один из самых посещаемых порталов в рунете (Mail.ru), игровые проекты под брендом MY.GAMES и четыре мессенджера.

В 2019 году было запущено подразделение Mail.ru Group Tech Lab. Направление отвечает за технологические проекты в области искусственного интеллекта, распознавания голоса и изображений, а также разработку новых экспериментальных коммуникационных продуктов.

Сферы применения: в Mail.ru Group технологии ИИ, машинного обучения и нейронных сетей применяются в ряде продуктов и сервисов. В их числе голосовой помощник Маруся, рекомендации контента ВКонтакте и в Одноклассниках, рекламный таргетинг в myTarget, компьютерное зрение Vision, прогнозирование спроса в Delivery Club, сервис умных ответов в Почте Mail.ru, технология поиска Mail.ru, машинное обучение в Mail.ru Cloud Solutions и другие.

В начале 2016 года в группе было создано бизнес-направление PREDICT (Predictive Analytic Solutions). Команда направления занимается разработкой сервисов клиентской аналитики и систем поддержки принятия решений на основе анализа большого объема данных и технологий ма-

шинного обучения. Проекты направлены на повышение эффективности маркетинга и продаж, исследований рынка, управления рисками, планирования, управления персоналом и других рабочих процессов компаний из самых разных индустрий.

**Руководитель направления:** Роман Стягуин, директор по развитию бизнеса PREDICT Mail.ru Group.

**Некоторые РА проекты:**

- **Пульс Mail.ru** — персональная лента материалов, собранная по интересам пользователя. Алгоритм новой ленты работает на базе технологий машинного обучения и анализирует большой объем информации, включая поведение пользователей в социальных сетях и на других сайтах, историю их кликов на сервисе, популярность источников и другие факторы.
- **Смотри Mail.ru** — в основе продукта лежит концепция персонального телеканала. Он включает музыкальные клипы, новости, телешоу, видеоблоги, фильмы и сериалы, которые подбираются персонализировано на основе предпочтений пользователя.
- **Диктор Mail.ru** — платформа для создания новостных и репортажных видео студийного качества. При создании виртуальных ведущих компания использовала собственные разработки в области искусственного интеллекта.

@ mail.ru  
group

₽87,1  
млрд

Выручка <sup>[243]</sup>

\$5,86  
млрд

Капитализация <sup>[243]</sup>

9

Патентов <sup>[251]</sup>

55

Публикаций <sup>[251]</sup>

## Лаборатория Касперского

**Описание:** международная компания, специализирующаяся на разработке систем защиты от компьютерных вирусов, спама, хакерских атак и прочих киберугроз. Компания ведёт свою деятельность более чем в 200 странах и территориях мира. Технологии «Лаборатории Касперского» защищают более 400 миллионов пользователей и 250 тысяч корпоративных клиентов во всём мире. Ведут разработку KasperskyOS — операционную систему для подключенных к интернету встраиваемых систем с особыми требованиями к кибербезопасности.

**Стратегия по ИИ:** Многоуровневая защита нового поколения «Лаборатории Касперского» активно применяет методы машинного обучения на всех стадиях процесса обнаружения угроз: от масштабируемых методов кластеризации, используемых для предварительной обработки

потока входящих файлов в инфраструктуре, до надежных и компактных моделей для поведенческого анализа, которые создаются на основе глубоких нейронных сетей и работают непосредственно на пользовательских устройствах.

**Руководитель:** Евгений Касперский.

**ИИ подразделение:** н/д.

**Лидер в ИИ:** н/д.

**Основные РА проекты:**

- **Технология Astraea** — собирает всю статистику и метаданные о подозрительных действиях и угрозах со всего мира в реальном времени, и на основе анализа этих данных принимает решения об обнаружении вредоносных объектов. Эти вердикты незамедлительно становятся доступны всем пользователям Kaspersky Security Network.

**KASPERSKY** Lab

**\$684,6M**

Выручка <sup>[245]</sup>

н/д

Капитализация

н/д

Затраты на RnD

0

Патентов <sup>[251]</sup>

0

Публикаций <sup>[251]</sup>

## Сбербанк

**Описание:** крупнейший банк в России и один из ведущих глобальных финансовых институтов. На долю Сбербанка приходится около трети активов всего российского банковского сектора. Сбербанк является ключевым кредитором для национальной экономики и занимает крупнейшую долю на рынке вкладов. Банк активно занимается построением экосистемы, в т.ч. включающей компании лидеры в разработке ИИ решений: ЦРТ (технологии Natural Language Processing и Speech Analytics), Vision Labs (технологии Computer Vision) и др.

**ИИ подразделения:** С 2017 года в Сбербанке функционирует Управление развития компетенций по исследованию данных, Лаборатория по искусственному интеллекту и внутреннее DS\AI сообщество. В лаборатории была разработана библиотека AutoML, автоматизирующая разработку типовых моделей принятия решений.

**Руководитель:** Герман Греф — Президент, Председатель Правления.

**Лидеры в ИИ:**

Александр Ведяхин — вице-президент, координирует деятельность Лаборатории по искусственному интеллекту, Управления развития компетенций по исследованию данных, Управления контроля и координации деятельности, Контрольно-экспертной группы.

Максим Еременко — старший управляющий директор Управления развития компетенций по исследованию данных, CDS Банка. Курирует разработку и реализацию инициатив в области искусственного интеллекта и анализа данных, создает среду для коммуникации и обмена опытом в области искусственного интеллекта и анализа данных.

**Основные направления исследований:** чат-боты, поиск и извлечение информации из текста, распознавание речи, рекомендательные системы.

**Некоторые РА проекты:**

- **AI трансформация. Преобразование Сбербанка в AI-native компанию,** когда ИИ — неотъемлемая часть бизнес-процессов. Масштабное создание бизнес-решений с использованием технологий ИИ и анализа данных. Технологии ИИ уже находят применение в множестве бизнес-процессов: интеллектуальная система управления, риск-менеджмент, комплаенс, подбор персонала, финансы, кибербезопасность, взаимодействие с клиентами, персонализация продуктов и сервисов, учёт и анализ обратной связи, управление ИТ-системами, фрод-аналитика, робототехника и др.
- **Создание промышленных единых ИИ-платформ** по основным технологиям ИИ (Computer Vision, Natural Language Processing, Speech Analytics, Optical Character Recognition и др.) для ускорения и стандартизации разработки ИИ решений. На базе платформ выстраиваются ИИ решения для банка, экосистемы и внешних клиентов.
- **Сбербанк проводит автоматизированную оценку соискателей** при найме консультантов в свои отделения. С помощью искусственного интеллекта банк присваивает скоринг-балл кандидату и вычисляет, скоро ли тот решит уволиться.
- **Big Five** — технология позволяет составить психологический портрет личности и оценить его благонадёжность для банка на основе пяти черт характера: добросовестность, открытость, общительность, законопослушность и эмоциональная неустойчивость
- **Рекомендательная система для платежей и переводов** в приложении Сбербанка — предсказывает покупки клиента и рекомендует ему магазины партнеров. Кроме того, рекомендует повторяющиеся платежи. Используется >1000 параметров для персонализации мобильного приложения

**СБЕРБАНК**

**\$11,2B**

Прибыль <sup>[243]</sup>

**\$62,7B**

Капитализация <sup>[243]</sup>

н/д

Затраты на RnD

0

Патентов <sup>[251]</sup>

3

Публикации за 2019 г <sup>[251]</sup>



## МТС

**Описание:** Публичное акционерное общество «Мобильные ТелеСистемы» (ПАО «МТС») — крупнейший российский телекоммуникационный оператор и провайдер цифровых услуг. Компания предоставляет финансовые услуги, сервисы электронной коммерции, облачные услуги, телемедицинский сервис, продукты для киберспорта и ИТ-решения в других областях. В 2017 году МТС создала отдельное подразделение для внедрения решений на базе ИИ — Центр AI.

Компания недавно начала работу с технологиями ИИ, но уже сформировала мощное подразделение, отмечаемое многими участниками рынка, несмотря на то что пока большинство технологий используется для внутреннего применения. Фокус исследований в сфере предиктивной аналитики и рекомендательных систем направлен на персональные продуктовые рекомендации для каждого клиента.

**Президент МТС:** Алексей Корня.

**Руководитель ИИ направления:** Аркадий Сандлер, руководитель Центра AI МТС.

**Исследователи по PA:** н/д.

**Основные направления исследований:** рекомендательные системы.

**Некоторые PA проекты:**

- В 2019 году МТС запустила сервис по подбору тарифного плана для каждого клиента.
- МТС Снежинка** — внутренний сервис, аккумулирующий данные о коммуникациях сотрудников. Сервис позволяет оценить активность сотрудника и получить рекомендации о развитии компетенций сотрудника и вероятности его увольнения.
- В марте 2020 года МТС и «Первый канал» договорились о стратегическом партнёрстве и создании совместного предприятия для формирования медийного предложения нового уровня. Партнёры планируют разработать рекомендательную систему контента, усовершенствовать инструменты мониторинга и аналитики, а также адаптировать контентную и рекламную политику с учётом современных способов потребления видео.



\$6,3B

Выручка <sup>[246]</sup>

\$8,3B

Капитализация <sup>[247]</sup>

н/д

Затраты на RnD

0

Патентов <sup>[251]</sup>

0

Публикаций <sup>[251]</sup>

## X5 Retail Group

**Описание:** X5 Retail Group — ведущая продуктовая розничная компания России.

Под управлением компании находится 17 025 магазинов под брендом «Пятёрочка», «Перекрёсток» и «Карусель», с лидирующими позициями в Москве и Санкт-Петербурге и значительным присутствием в европейской части страны.

Штат подразделения Big Data составляет 340 человек.

В апреле-июне рост чистой розничной выручки ключевых форматов «Пятёрочки» и «Перекрёстка» составил 15,6% при росте LFL продаж на 4,7%.

**Онлайн-продажи:** выручка от онлайн-продаж Perekrestok.ru и экспресс-доставки из магазинов «Пятёрочка» и «Перекрёсток» составила 5,1 млрд руб. во 2 кв. 2020 г., достигнув пика в 2,1 млрд руб. в мае, когда обе онлайн платформы вышли на положительный уровень EBITDA. X5 стала ведущим игроком по обороту в сегменте онлайн-торговли продуктами питания в России во 2 кв. 2020 г.

Для формирования и развития культуры управления на основе данных реализована политика «демократизации» данных: создан единый каталог данных, содержащий одновариантный корпус данных с указанием источника; внедряются инструменты аналитики для самостоятельного использования. Задача корректного использования больших данных, обеспечивающих поддержку бизнес-стратегии, требует постоянных экспериментов, прототипирования и извлечения новых знаний из данных. Ряд инструментов работы с большими данными уже прошли этот этап разработки и в настоящее время используются в деятельности компании на системной основе.

**Главный исполнительный директор:**

Игорь Шехтерман.

**Директор по большим данным X5 Retail Group:** Антон Мироненков.

**Руководитель ИИ направления:** Андрей Молчанский, директор департамента разработки и сопровождения продуктов больших данных.



\$26,8B

Выручка <sup>[255]</sup>

\$9,67B

Капитализация <sup>[256]</sup>

н/д

Затраты на RnD

0

Патентов <sup>[251]</sup>

0

Публикаций <sup>[251]</sup>

**Исследователи по PA:** Валерий Бабушкин, директор по моделированию и анализу данных в X5 Retail Group.

**Основные направления исследований:** компьютерное зрение, рекомендательные системы.

**Некоторые PA проекты:**

- В X5 Retail Group применяют аплифт-моделирование** — прогнозный метод оценки эффекта воздействия промо-акций на потребителя, снижения оттока клиентов и определения ценовой эластичности.
- Ассортимент.** Система автоматизированного управления (инструменты общего и покатегорийного анализа) помогает формировать ассортимент по 28 категориям товаров, на долю которых приходится более 60% чистой розничной выручки группы.
- Ценообразование.** Внедренная в X5 автоматизированная система управления ценообразованием позволяет нам более точно и своевременно вырабатывать решения в области ценовой политики. Сочетание автоматизированных алгоритмов анализа с постоянным мониторингом цен конкурентов лежит в основе автоматизированного процесса принятия решений в сфере управления ценообразованием.
- Промоакции.** X5 повышает эффективность промоакций за счет использования продвинутой аналитики на основе больших данных и анализа результатов прошлых акций, что способствует принятию более точных и эффективных решений. В настоящее время в ТС «Перекрёсток» проходит тестирование системы управления потребительской ценностью (CVM), используемой для прогнозирования торгового оборота при проведении промоакций. В дальнейшем система будет развиваться для решения задач прогнозирования регулярного спроса на основе предиктивных моделей, что позволит нам повысить эффективность логистики за счет более точного планирования сроков и объемов поставок.

## Газпром Нефть

**Описание:** вертикально-интегрированная нефтяная компания, основные виды деятельности которой — разведка и разработка месторождений нефти и газа, нефтепереработка, а также производство и сбыт нефтепродуктов. Входит в ТОП-3 российских нефтяных компаний по объемам добычи и переработки нефти. Компания занимается разработкой собственных программных продуктов в рамках одного из направлений Технологической стратегии — это направление ЭРА (это система цифровых продуктов, которая оптимизирует всю производственную цепочку добычи нефти).

**Председатель правления, генеральный директор:** Александр Дюков.

**Лидер в ИИ:** Роман Худорожков, руководитель программ продвинутой аналитики и машинного обучения, Газпромнефть.

**Исследователи по РА:** н/д.

**Основные направления исследований:** предсказательное моделирование, системы оптимизации производственных процессов.

**Некоторые РА проекты:**

- **Интеллектуальная система «ЭРА: ОптимА»** — комплексно оптимизирует разработку месторождения, позволяет подобрать эффективную геометрию расположения новых и режимы работы уже существующих скважин на основе трехмерного моделирования. Проведенные расчеты показали, что варианты, найденные системой, на 20–30% превосходят по экономической эффективности предложения, сформированные на основе экспертных методов.

- **Компания применяет машинное обучение** для поиска дополнительных скрытых запасов нефти. Компании удалось добиться дополнительного притока нефти на уровне 70 баррелей в сутки на Ямале.
- **«Газпром нефть» внедрила интеллектуальную систему мониторинга оборудования на АЗС.** Автоматизация мониторинга и предиктивная аналитика МЦИ позволяют отслеживать в онлайн-режиме состояние оборудования и контролировать процессы эксплуатации АЗС, связанные с сохранностью и качеством топлива.
- **Газпром Нефть реализует программу «Цифровая нефть».** Программа нацелена на повышение эффективности зрелых месторождений: анализирует геолого-геофизические исследования по нескольким десяткам тысяч скважин и вырабатывает алгоритм, который позволяет предсказать новые перспективные зоны для поиска залежей углеводородов. Программа обрабатывает сотни гигабайт данных с тысяч скважин, расположенных на месторождении, анализируя около 60 тыс. результатов геофизических исследований одновременно. В результате удалось добиться дополнительной добычи на уровне 70 баррелей нефти в сутки без затрат на дополнительное бурение и создание инфраструктуры.



\$23B

Капитализация <sup>[252]</sup>

₽2,5 трлн

Выручка <sup>[253]</sup>

н/д

Затраты на RnD

0

Патентов <sup>[251]</sup>

0

Публикаций <sup>[251]</sup>

## Цифра

**Описание:** Группа компаний «Цифра» разрабатывает и внедряет решения на основе промышленного Искусственного интеллекта и интернета вещей, а также развивает индустрию роботизированного промышленного транспорта. Идея создания проекта принадлежит Виктору Вексельбергу и Пекке Вильякайнену.

Компании, входящие в состав «Цифры», фокусируются на работе с крупными клиентами. В их числе: «Ростех» и его предприятия, «Росатом», «Роскосмос», «Алмаз-Антей», из нефтяных — «Роснефть», «Газпром нефть», «Сибур», подрядчики «Газпрома» и др. В сентябре 2019 года «Цифра» за 9 миллионов долларов приобрела компанию «Геонавигационные технологии», разработчика технологий «умного бурения».

**Главный директор:** Игорь Богачев.

**Руководитель ИИ направления:** Александр Смоленский, директор по искусственному интеллекту.

**Исследователи по РА:** Сергей Свириденко.

**Основные направления исследований:** сбор данных с промышленных датчиков и предсказательная аналитика на них.

**Некоторые РА проекты:**

- **Zyfra Predictive Maintenance Tool Life** прогнозирует срок службы режущего инструмента станков с ЧПУ с использованием машинного обучения. Решение позволяет правильно рассчитать время для замены инструмента, прежде чем он износится и это негативно повлияет на качество детали. Также система в реальном времени анализирует поступающие данные со станка и сигнализирует о возможных аномалиях, которые приводят к браку, поломке инструмента и общему износу оборудования.
- В 2018 году Цифра внедрила систему предиктивной аналитики для трубопрокатного стана Челябинского трубопрокатного завода. Система снимает показания с нескольких сотен датчиков и подбирает наиболее оптимальный технологический режим, увеличивая скорость обработки трубы. В результате применения удалось снизить себестоимость трубы на 5%.



н/д

Капитализация

\$31.7M

Выручка <sup>[248]</sup>

н/д

Затраты на RnD

0

Патентов <sup>[251]</sup>

0

Публикаций <sup>[251]</sup>

## КРОК

**Описание:** КРОК — один из лидеров российского рынка информационных технологий. ИТ-компания создает новую цифровую экосистему общества, реализуя проекты в 40 странах мира. КРОК предоставляет полный портфель ИТ-услуг в области системной интеграции, консалтинга, тиражируемых продуктов, управляемых B2B-сервисов, перспективных сквозных технологий — Big Data, блокчейн, искусственный интеллект, виртуальная и дополненная реальность, интернет вещей, роботизация, машинное обучение.

**Главный директор:** Борис Бобровников.

Руководитель ИИ направления: Егор Осипов, Руководитель направления Big Data.

**Исследователи по РА:** Алексей Сидорин, архитектор бизнес-решений, руководитель направления HR-технологий и управления знаниями.

**Основные направления исследований:** алгоритмы классификации и регрессии, нейросетевые архитектуры, включая глубокие нейросети, предикативные модели, алгоритмы снижения размерности и кластеризации, методы оценки моделей.

**Некоторые РА проекты:**

- **Отслеживание уникальных объектов.** Поиск и распознавание уникальных QR-кодов, наклеенных на тележки и контейнеры позволил автоматизировать отслеживание их передвижения. На основе анализа траекторий создавалась виртуальная карта и велся подсчет тележек.
- **Поиск пропавших детей.** С помощью видеоаналитики (распознавание лиц) был решен кейс на Востоке. Искали автобус с потерявшимся школьником. Оказывается, там дети часто садятся не в свой школьный автобус и потом катаются в нем по городу, в то время как родители пытаются их отыскать.

- **Люди на «удаленке».** Распознавание лиц для контроля доступа сотрудников к корпоративным системам с мобильных устройств (телефоны, планшеты, ноутбуки). Распознает даже людей в масках.
- **Онлайн-заседание в суде. Пилотный проект, с применением в ходе дистанционного заседания суда по видеоконференцсвязи (ВКС) российской биометрической системы распознавания лиц на основе нейросетей.** В условиях изоляции и ограниченного доступа граждан в суды разработка помогла провести заседание без физического присутствия сторон.
- **Прогнозирование пассажиропотока.** Построение помесечного прогноза по типу пассажиров, билетов, льготным группам. Профилирование по таким параметрам как: температура, осадки, наличие турникетов, контролёров, индексация тарифов. Динамическая корректировка стоимости билетов в зависимости от населенности и числа дней до отправления.
- **Нейросеть на базе ядра диалоговой системы AutoFAQ** внедрена в первую линию службы поддержки в офисах «МегаФон», где будет классифицировать типовые заявки и автоматически отправлять их на исполнение. Это поможет обрабатывать в три раза больше запросов без увеличения штата, более половины всех обращений может быть обработано без участия живых операторов.
- **Прогнозирование грузоперевозок.** Создание долгосрочного прогноза по маршрутам и объемам перевозок на базе исторических данных за длительный период.

## КРОК

Н/д  
Капитализация

\$409M  
Выручка <sup>[243]</sup>

Н/д  
Затраты на RnD

0  
Публикаций <sup>[251]</sup>

0  
Патентов <sup>[251]</sup>

## ЛАНИТ

**Описание:** ЛАНИТ — одна из крупнейших групп компаний на российском рынке информационных технологий. Предоставляет широкий комплекс ИТ-услуг, обеспечивающих этапы выполнения проекта от разработки и внедрения до обучения персонала и сервисной поддержки. Основана в 1989 году, штат насчитывает 8 000 человек. Партнер более 250 мировых производителей оборудования и программных решений: Avaya, Citrix, Ericsson, HP, Intel, Microsoft, Oracle, Siemens PLM Software и др.

За последние годы в составе группы появилось несколько компаний, специализирующихся на технологиях ИИ — например, компания Datana, которая разрабатывает предиктивную аналитику для промышленных предприятий и компания CleverData, разрабатывающая технологии предиктивной аналитики для рекламы и маркетинга.

**Руководитель:** Филипп Генс, президент группы компаний.

**Руководитель ИИ направления:** Денис Афанасьев, Директор центра компетенций больших данных и искусственного интеллекта ЛАНИТ.

**Основные направления исследований:** Большие данные, предиктивная аналитика, оптимизация бизнес-процессов.

**Некоторые РА проекты:**

- **ЛАНИТ Омни** специализируется на услугах для розничных торговых сетей: разработке и внедрении программ лояльности и сервиса персональных рекомендаций покупателям, основанных на использовании методов предиктивной аналитики. Сервис персональных рекомендаций покупателям ClearFuture — инструмент для построения прогноза покупок для каждого покупателя на основании анализа потребительского поведения других похожих на него клиентов, учитывает индивидуальную историю покупок конкретного покупателя.
- В 2020 году Datana внедрила решение на Ашинском металлургическом заводе по предоставлению операторам плавки металла рекомендаций по ведению технологического процесса в реальном времени. В результате было достигнуто снижение на 8% расхода ферросплавов при плавке стали.



₽173,8  
млрд  
Консолидированный оборот <sup>[244]</sup>

Н/д  
Капитализация

Н/д  
Затраты на RnD

0  
Патентов <sup>[251]</sup>

0  
Публикаций <sup>[251]</sup>



# Тренды и аналитика

[07]

**Игорь Пивоваров**

Центр Науки и Технологий  
Искусственного Интеллекта МФТИ

**Наталья Гутенева**

Центр Науки и Технологий  
Искусственного Интеллекта МФТИ

## Тренды и прогнозы

В этой статье мы постарались сформулировать основные технологические тренды в области предсказательной аналитики и систем поддержки принятия решений, а также прогнозы новых применений и развития рынка по состоянию на 2020 год.

### Технологические тренды

#### ПОТОКОВАЯ ОБРАБОТКА ДАННЫХ (REAL TIME DATA STREAMS)

Real time data streams processing / event-driven data processing — методы обработки непрерывного объема данных с минимальной задержкой, практически в реальном времени. Ранее модели обработки потоков данных применялись в узкоспециализированных сегментах, однако с появлением IoT, IIoT и глобальной цифровизацией данные технологии оказались востребованы во многих отраслях, включая промышленность, ритейл, рекламу и др. В потоковую обработку данных входит: фильтрация, агрегация и преобразование потоковых данных, а также зачастую автоматическое построение моделей и принятие решений на их основе. Она используется для:

- Борьбы с финансовыми мошенничествами (fraud detection) в реальном времени;
- Динамического ценообразования;
- Мониторинга промышленных систем и оповещения о неполадках в реальном времени;
- Мониторинга состояния пациентов в реальном времени;
- Создания систем поддержки принятия решений по торгам на фондовой бирже.

Наиболее популярные сервисы потоковой обработки данных: Google Cloud DataFlow, Amazon Kinesis, Apache Kafka/Fink/Aurflow, Azure Stream Analytics.

#### AUTOML

Automating machine learning или AutoML — сравнительно недавно появившийся концепт, заключающийся в частичной или полной (“end-to-end”) автоматизации решений на основе машинного обучения. В предсказательной аналитике AutoML системы применяются для предварительной подготовки данных, выбора архитектуры и настройки гиперпараметров нейросетей. Использование подобных систем позволяет значительно ускорить и упростить разработку и развертывание предсказательных моделей, а также повысить точность предсказаний. Согласно отчету Google AI, именно благодаря AutoML компании удалось добиться повышения скорости предсказаний на 30% и точности на 8%<sup>[254]</sup>. Уже сейчас AutoML активно применяется при выделении клиентских сегментов и построении архитектур рекомендательных систем. Среди бесплатных AutoML-фреймворков можно выделить: Auto Tune Models, TransmogriAI, AutoSklearn, hyperopt-sklearn, Auto Weka, а среди коммерческих — H2O driverless, Datarobot AutoML, Google Cloud AutoML.

#### РАЗМЕТКА ДАННЫХ — АВТОМАТИЗИРОВАННАЯ ИЛИ КРАУДСОРСИНГОВАЯ

Для обучения ML-моделей, которые смогут делать релевантные предсказания, необходимы качественные данные. Ручная или полуавтоматическая подготовка, фильтрация и разметка большого объема данных — сложный и дорогостоящий процесс. Именно он требует основных вложений. Поэтому набирают все большую популярность алгоритмы автоматизированной разметки данных, такие как Amazon Mechanical Turk.

Наряду с этим, появляется множество сервисов, привлекающих краудсорсинг (обычных людей) для разметки, например сервис Яндекс Толока. Применяя машинное обучение для подсказки пользователю вариантов ответа, ему отдается финальное решение на разметку, но весь процесс сильно упрощается и ускоряется.

#### BLACK-BOX OPTIMIZATION

Зачастую нам неизвестна аналитическая форма функции, которую необходимо оптимизировать, известны лишь ее значения в некоторых точках. Соответственно, мы не можем вычислить ее градиенты. В таких случаях применяется black-box optimization — методы оптимизации неизвестных функций. К ним относятся байесовская оптимизация и различные методы оптимизации без использования производных (например, метод роя частиц, генетические и эволюционные алгоритмы и др.). С развитием алгоритмов машинного обучения и увеличением вычислительных мощностей black-box optimization набирает все большую популярность, в особенности, для решения задач предсказательной аналитики, где архитектуры зачастую представляют собой «черные ящики», однако выбор наиболее релевантного для конкретной архитектуры алгоритма оптимизации все еще является достаточно сложной задачей. В открытом доступе находятся следующие фреймворки для black-box optimization: BlaBoO, Kurobako, COCO (COmparing Continuous Optimisers).

#### DATAOPS

DataOps — новая ориентированная на данные методология, аналог DevOps, заключающийся в организации непрерывного и беспрепятственного процесса доступа к данным и извлечения из них аналитической информации. По словам Ашиша Тусу, возглавляющего команду разработчиков инфраструктуры данных Facebook, «DataOps — это новый способ управления данными, стимулирующий интеграцию и взаимодействие между ранее изолированными данными, командами разработчиков и системами. Посредством изменений в определенных процессах, перестройки организации и применения нужных технологий DataOps устанавливает взаимоотношения между всеми, кто имеет дело с данными: разработчиками, специалистами по архитектуре и обработке данных, аналитиками и корпоративными пользователями. DataOps формирует тесную взаимосвязь между теми, кто занимается сбором и подготовкой данных, теми, кто анализирует данные, и теми, кто применяет полученные в результате этого анализа выводы на практике для развития бизнеса».

## Рыночные тренды

### ПРЕДСКАЗАНИЕ НА БИРЖЕ — УЖЕ РЕАЛЬНОСТЬ

Можно уверенно утверждать, что предсказательная аналитика и системы принятия решений уже работают в финансовой отрасли, как минимум в сфере инвестиций и биржевой торговли. Информации по этим применениям крайне мало, но это понятно: компании, у которых предсказательная аналитика работает и приносит доход предпочитают о ней не распространяться. Но целая индустрия торговых биржевых роботов показывает, что в этой области ПА успешно справляется со своими задачами.

Однако известно, что биржа быстро адаптируется к любому алгоритму торговли. Поэтому фактически мы имеем не столько конкуренцию алгоритмов, сколько постоянную конкуренцию команд на финансовых рынках, сконцентрированных, судя по всему, вокруг ведущих инвестиционных банков.

### ПА В МЕДИЦИНЕ

Пожалуй, одно из самых востребованных применений предсказательной аналитики и систем поддержки принятия решений — здравоохранение и медицина. В США и Европе медицинские информационные системы (МИС) и системы принятия врачебных решений активно внедряются в госпитали и поликлиники. Данные системы позволяют снизить нагрузку на врачей, уменьшить процент врачебных ошибок, а также снизить среднюю стоимость приема врача. В России подобные системы только начинают применяться, и то лишь в режиме пробных запусков.

Основная проблема внедрения таких систем — ответственность за результат. Если алгоритм принял (предложил) неправильное решение — кто отвечает за последствия? Этот вопрос сейчас остро дискутируется в профессиональном сообществе, есть несколько возможных подходов, эту тему мы раскроем в одном из следующих выпусков.

### ПОЛНАЯ АВТОМАТИЗАЦИЯ

Все более популярными становятся решения на основе предсказательной аналитики «под ключ» — готовые системы, представляющие собой черные ящики, на вход получающие данные и автоматические выполняющие их фильтрацию, обработку, анализ, а также строящие на их основе модели и выдающие на выходе готовые паттерны и предсказания. Подобные системы разрабатываются крупными ИИ-компаниями не только для частных применений, таких как промышленность, здравоохранения, реклама и т.д., но и зачастую носят универсальный характер. На рынке доступны автоматизированные или частично-автоматизированные решения на основе алгоритмов машинного обучения от SAS, SAP, Amazon, IBM, RapidMiner и др.

### ДЕЦЕНТРАЛИЗАЦИЯ

Постепенно набирают популярность технологии, связанные с распределенными системами обработки данных, блокчейн-архитектурами и edge computing («граничными вычислениями»). В мире насчитывается гигантское количество «умных» сенсоров и других IoT-устройств, которые генерируют зеттабайты данных. Даже самые мощные суперкомпьютеры не могут справиться с такими объемами информации. В связи с этим теряет целесообразность сбор и обработка данных в едином централизованном хранилище. Вместо этого крупные it-компании начинают развертывать свои собственные инфраструктуры для распределенной обработки данных. Пока еще не решены часть вопросов, связанных с обеспечением скорости и безопасности передачи данных, масштабированием архитектур и отказоустойчивостью звеньев распределенной сети. Однако данная парадигма несомненно является перспективной в свете все увеличивающегося объема данных и постоянно возрастающей потребности в быстродействии их обработки.

### РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ ВИЗУАЛЬНОГО ПОИСКА

В последнее время начали появляться системы персональных рекомендаций, подбирающие товары и услуги не только по параметрам, но и по внешнему виду. Они совмещают в себе новейшие достижения в сфере машинного зрения и предсказательной аналитики и автоматически распознают размеры, формы, материалы, цветовые гаммы товаров, а также с помощью системы тегирования и некоторых инструментов распознавания образов определяют их стилистическую принадлежность.

Одним из примеров рекомендательных систем визуального поиска является сервис цветового поиска MulticolorEngine. Посетители интернет-магазинов с помощью этого сервиса могут осуществлять подбор товаров по цвету. В 2013 году система поиска похожих товаров Similar4.ru была разработана компанией Kuzpetch. Она работает по схожему принципу с западными рекомендательными системами, по модели SaaS. При внедрении она подключается через API к каталогу интернет-магазина, на сайте магазина появляется рекомендательный виджет с выбором похожих товаров по форме, цвету, текстуре материала и еще 150 параметрам.



Если Вы хотите быть в курсе всех трендов и познакомиться лично с ведущими людьми и компаниями в ИИ, то для этого есть главное место:

## Открытая конференция по искусственному интеллекту

**03-05 февраля 2021**

Конференц-зал гостиницы "Космос»

**OpenTalks.AI** – ведущая независимая открытая конференция по искусственному интеллекту в России.

Лучшие российские докладчики со всех топовых международных конференций по ML/DL на одной площадке.

### Сбор заявок на доклады

Мы объявляем открытый **call for papers and proposals** (сбор заявок на доклады) для всех желающих. Лучшие доклады и панельные секции отбираются Программным комитетом и из них, а также из приглашенных спикеров, формируется программа.

Программа, спикеры и регистрация на сайте  
[www.OpenTalks.AI](http://www.OpenTalks.AI)

## Приложения

08

# Основные ссылки и методика формирования отчета

1. Юваль Ной Харари. Sapiens. Краткая история человечества. 2017.
2. Краткое изложение книги Эрика Сигела «Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die» <https://lifeclub.org/books/predictive-analytics-eric-siegel-review-summary>.
3. Кукла, используемая в колдовстве вуду. Последователи учения вуду считают, что в результате специального обряда кукла получает особую связь с определённым человеком. Согласно этому представлению, после установления связи благодаря принципам симпатической магии у обладателя куклы появляется возможность воздействовать через неё на человека, которого она символизирует.
4. <https://www.youtube.com/watch?v=WQMuxNiYoz4>.
5. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
6. Оригинал доступен по ссылке: <https://vc.ru/services/65236-prediktivnaya-analitika-na-palcah>.
7. Kumar, Vaibhav & L., M.. (2018). Predictive Analytics: A Review of Trends and Techniques. International Journal of Computer Applications. 182. 31-37. 10.5120/ijca2018917434.
8. Charles Elkan, 2013, "Predictive analytics and data mining", University of California, San Diego.
9. Eric Siegel, 2016, "Predictive Analytics", John Willey and Sons Ltd.
10. Charles Nyce, 2007, "Predictive Analytics White Paper", American Institute of CPCU/IIA.
11. W Eckerson, 2007, "Extending the Value of Your Data Warehousing Investment", The Data Warehouse Institute.
12. Sue Korn, 2011, "The Opportunity of Predictive Analytics in Finance", HPC Wire.
13. M Nigrini, 2011, "Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations", John Willey and Sons Ltd.
14. M Schiff, 2012, "BI Experts: Why Predictive Analytics Will Continue to Grow", The Data Warehouse Institute.
15. F Reichheld, P Schefter, Retrieved 2018, "The Economics of E-Loyalty", Harvard Business School Working Knowledge.
16. V Dhar, 2001, "Predictions in Financial Markets: The Case of Small Disjuncts", ACM Transaction on Intelligent Systems and Technology, Vol-2, Issue-3.
17. J Osheroff, J Teich, B Middleton, E Steen, A Wright, D Detmer, 2007, "A Roadmap for National Action on Clinical Decision Support", JAMIA: A Scholarly Journal of Informatics in Health and Biomedicine, Vol-14, Issue2, Pages-141-145.
18. B Kaminski, M Jakubczyk, P Szufel, 2018, "A framework for sensitivity analysis of decision trees", Central European Journal of Operations Research, Vol26, Issue-1, Pages-135-159.
19. J S Armstrong, 2012, "Illusions in regression analysis", International Journal of Forecasting, Vol-28, Issue-3, Pages-689-694.
20. W S McCulloch, Walter Pitts, 1943, "A logical calculus of the ideas immanent in nervous activities", The bulletin of mathematical biophysics, Vol-5, Issue-4, Pages-115- 133.
21. Peter M Lee, 2012, "Bayesian Statistics: An Introduction, 4th Edition", John Willey and Sons Ltd.
22. R Polikar, 2006, "Ensemble based Systems in decision making", IEEE Circuits and Systems Magazine, Vol-6, Issue-3, Pages-21-45.
23. J H Friedman, 1999, "Greedy Function Approximation: A Gradient Boosting Machine", Lecture notes.
24. C Cortes, 1995, "Support-vector networks", Machine Learning, Vol-20, Issue-3, Pages- 273-297.
25. Ben Hur et al, 2001, "Support Vector Clustering", Journal of Machine Learning Research, Vol-2, Pages125-137.
26. J Lin, E Keogh, S Lonardi, C Chiu, 2003, "A symbolic representation of time series, with implications for streaming algorithms", Proceedings of the 8th ACM SIGMOD workshop on research issues in data mining and knowledge discovery, Pages-2-11.
27. N S Altman, 1992, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression", The American Statistician, Vol-46, Issue-3, Pages- 175-185.
28. H Abdi, L J Williams, 2010, "Principal component analysis", WIREs: Computational Statistics, Vol-2, Issue-4, Pages-433-459.
29. K Das, GS Vidyashankar, 2006, "Competitive Advantage in Retail Through Analytics: Developing Insights, Creating Values", Information Management.
30. N Conz, 2008, "Insurers Shift to Customer-Focused Predictive Analytics Technologies", Insurance & Technology.
31. J Feblowitz, 2013, "Analytics in Oil and Gas: The Big Deal About Big Data", Proceeding of SPE Digital Energy Conference, Texas, USA.
32. G H Kim, S Trimi, J-H Chung, 2014, "Big-data applications in the government sector", Communications of the ACM, Vol-57, Issue-3, Pages-78-85.
33. Recent Advances in Design and Decision Support Systems in Architecture and Urban Planning Editors: van Leeuwen, Jos P., Timmermans, Harry J.P. (Eds.) 2004.
34. Merkert, Mueller, Hubl, A Survey of the Application of Machine Learning in Decision Support Systems, University of Hoffenheim 2015.
35. Sanchez i Marre, Gibert, Evolution of Decision Support Systems, University of Catalunya, 2012.
36. Ltifi, Trabelsi, Ayed, Alimi, Dynamic Decision Support System Based on Bayesian Networks, University of Sfax, National School of Engineers (ENIS), 2012.
37. How a Pioneer of Machine Learning Became One of Its Sharpest Critics, The Atlantic, 2018.
38. Варламов О.О. Перспективы искусственного интеллекта // Знание – Сила, №3, 2020. с. 36 – 46.
39. Варламов О. О. Миварный подход как основа качественного перехода на новый уровень в области искусственного интеллекта // Радиопромышленность. 2017. № 4. С. 13–25.
40. Уоссермен Ф. Нейрокомпьютерная техника. -М.: Мир, 1992.
41. Варламов О. О. Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство. - М.: Радио и связь, 2002. - 288 с.
42. Varlamov O. MIVAR: Transition from Productions to Bipartite Graphs MIVAR Nets and Practical Realization of Automated Constructor of Algorithms Handling More than Three Million Production Rules // arxiv.org: arXiv preprint arXiv:1111.1321, <https://arxiv.org/abs/1111.1321>, 23 p., 2011.
43. Варламов О.О. Роль и место миваров в компьютерных науках, системах искусственного интеллекта и информатике // Радиопромышленность. 2015. №3. С. 10-27.
44. Shadrin, S. S., Varlamov O. O., Ivanov A. M. Experimental Autonomous Road Vehicle with Logical Artificial Intelligence // Journal of Advanced Transportation, vol. 2017, Article ID 2492765, 10 p., 2017.



45. Варламов О. О. Хадиев А. М., Чибирова М. О., Сергушин Г. С., Антонов П. Д. Автоматизированное построение маршрута логического вывода в миварной базе знаний // Патент на изобретение RU 2607995 11.02.2015., опубликовано 11.01.2017, бюл. №2. – 43 с.
46. Varlamov O. O. Wi!Mi Expert System Shell as the Novel Tool for Building Knowledge-Based Systems with Linear Computational Complexity // International Review of Automatic Control, 2018, 11(6), 314-325.
47. Базы знаний интеллектуальных систем/Т.А. Гаврилова, В.Ф. Хорошевский. -СПб.: Питер, -384 с.
48. Попов Э.В. Искусственный интеллект: экспертные системы. -М.: Наука, 1990.
49. Поспелов Д.А. Введение в теорию вычислительных систем. -М.: Сов. радио, 1983.-272с.
50. Поспелов Д.А. Логико-лингвистические модели в системах управления. -М.: Энергоиздат, 1981. -232 с.
51. Поспелов Г.С., Поспелов Д.А. Искусственный интеллект -прикладные системы. -М.: Знание, 1985. -48 с.
52. Таунсенд К., Фохт Д. Проектирование и программная реализация экспертных систем на персональных ЭВМ/Пер. с англ. -М.: Финансы и статистика, 1990. -320 с.
53. Уотермен Д. Руководство по экспертным системам/Пер. с англ. -М.: Мир, 1989. -388 с.
54. Экспертные системы. Базы знаний и данных. Материалы семинара//Под рук. Э.В. Попова. -М.: ЦРДЖ, 1992. -171 с.
55. Кирсанов Б.С., Попов Э.В. Отечественные оболочки экспертных систем // Справочник по искусственному интеллекту. Т. 1. -М.: Радио и связь, 1990. С. 369-388.
56. Попов Э.В., Фоминых И.Б., Кисель Е.Б., Шапот М.Д. Статистические и динамические экспертные системы. -М.: Финансы и статистика, 1996.
57. Яшин А.М. Разработка экспертных систем. -Л.: ЛПИ, 1990.
58. Справочник по искусственному интеллекту в 3-х т.//Под ред. Э.В. Попова, Д.А. Поспелова. -М.: Радио и связь, 1990.
59. Антонов П.Д., Чибирова М.О., Жданович Е.А., Сергушин Г.С., Елисеев Д.В. Практический пример использования миварного подхода для создания экспертной системы в предметной области «Геометрия»//Радиопромышленность.2015.№3. С. 131.
60. Чибирова М.О. Структурное развитие миварного подхода: классы и отношения//Радиопромышленность.2015.3.С.44.
61. Чибирова М.О. Необходимость добавления ограничений и прецедентов для развития миварного информационного пространства // Радиопромышленность. 2015. № 3. С. 67-78.
62. Чибирова М.О. Сравнительный анализ миварного подхода с подходами, основывающимися на онтологиях и когнитивных картах // Радиопромышленность. 2015. № 3. С. 55-66.
63. Хадиев А.М. Разработка и практическая реализация миварной машины логического вывода // Радиопромышленность. 2015. № 3. С. 79-89.
64. Сергушин Г.С. Компьютерно-реализованная система для автоматизированного построения маршрута логического вывода в миварной базе знаний // Радиопромышленность. 2015. № 3. С. 90-99.
65. Чувииков Д. А. Об экспертной системе «Анализ ДТП», основанной на концепции миварного подхода // Проблемы искусственного интеллекта. 2017. № 2(5). С. 78-88.
66. Жданович Е. А., Антонов П. Д., Хадиев А. М., Сергушин Г. С., Чибирова М. О. Постановка диагноза по симптомам на основе миварного подхода / // Радиопромышленность. 2015. № 3. С. 122-130.
67. Чувииков Д. А. Применение экспертного моделирования в получении новых знаний человеком // Радиопромышленность. – 2017. № 2. – С. 72-80.
68. Адамова Л. Е., Белоусова А. И., Протопопова Д. А., Елисеев Д. В., Петерсон А. О. Об одном подходе к созданию интеллектуальной вопросно-ответной системы «Миварный виртуальный консультант» // Радиопромышленность.2015. № 3.с.160.
69. Максимова А.Ю., Варламов О.О. Миварная экспертная система для распознавания образов на основе нечеткой классификации и моделирования различных предметных областей с автоматизированным расширением контекста // Известия ЮФУ. Технические науки. 2011. № 12 (125). С. 77-87.
70. Майборода Ю.И., Синцов М.Ю., Озерин А.Ю. и др. Система автоматического тегирования изображений на основе миварных технологий // Программные системы: теория и приложения. 2014. Т. 5. № 4 (22). С. 159-170.
71. Сергушин Г.С. Разработка миварных АСУ ТП для различных применений в автомобильно-дорожной сфере // Радиопромышленность. 2015. № 3. С. 100-111.
72. Чувииков Д.А., Теплов Е.В., Сараев Д.В., Варламов О.О., Пунам Д. Методика автоматизации системы диспетчерского контроля на основе экспертной системы городского пассажирского транспорта // Радиопромышленность. 2016; 26(4): 85-95.
73. Жданович Е. А., Панферов А. А., Юфимычев К. А., Хадиев А. М., Елисеев Д. В. Применение миварной экспертной системы для планирования движения мобильного сервисного робота // Радиопромышленность. 2015. № 3. С. 243-254.
74. Жданович Е. А., Чернышев П. К., Юфимычев К. А., Елисеев Д. В., Чувииков Д. А. Вычисление произвольных алгоритмов функционирования сервисных роботов на основе миварного подхода // Радиопромышленность. 2015.3.С.226-242.
75. Варламов О.О., Лазарев В.М., Чувииков Д.А., Пунам Д. О перспективах создания автономных интеллектуальных роботов на основе миварных технологий // Радиопромышленность. 2016; 26(4):96-105.
76. Варламов О.О. О метрике автономности и интеллектуальности робототехнических комплексов и киберфизических систем. // Радиопромышленность. 2018; 28(1):74-86.
77. Варламов О.О., Аладин Д.В. О создании миварных систем контроля за соблюдением правил дорожного движения на основе «РАЗУМАТОРОВ» и экспертных систем. // Радиопромышленность. 2018;28(2):25-35.
78. Варламов О. О., Аладин Д. В. О применении миварных сетей для интеллектуального планирования поведения роботов в пространстве состояний // Известия Кабардино-Балкарского научного центра РАН. 2018. № 6-2 (86). С. 75-82.
79. Варламов О.О., Аладин Д. В. Успешное применение миварных экспертных систем для MIPRA - решения задач планирования действий робототехнических комплексов в реальном времени // Радиопромышленность. 2019; №3(29).С.15-25.
80. Фейгенбаум Э., Фельдман Дж. Вычислительные машины и мышление. -М.: Мир, 1967.
81. Финк Д. Вычислительные машины и человеческий разум. -М.: Мир, 1967.
82. Фогель Л., Оуэнс А., Уолш М. Искусственный интеллект и эволюционное моделирование. -М.: Мир, 1969.
83. Хант Э. Искусственный интеллект. -М.: Мир, 1978. -558 с.
84. Уинстон П. Искусственный интеллект. -М.: Мир, 1980.
85. Шалютин С.М. Искусственный интеллект. Гносеологический аспект. -М.: Мысль, 1985.
86. Эндрю А. Искусственный интеллект. -М.: Мир, 1985.
87. Джордж Ф. Основы кибернетики/Пер. с англ. -М.: Радио и связь, 1984. -272 с.

88. Городецкий В.И. Прикладная алгебра и дискретная математика. Часть 1. Алгебраические системы. -МО СССР, 1984. -174 с.
89. Городецкий В.И. Прикладная алгебра и дискретная математика. Часть 2. Формальные системы нелогического типа. -МО СССР, 1986. -200 с.
90. Городецкий В.И. Прикладная алгебра и дискретная математика. Часть 3. Формальные системы логического типа. -МО СССР, 1987. -177 с.
91. Ростовцев Ю.Г. Информация и представление знаний//Известия ВУЗ. Приборостроение, 1981, № 4, с. 92 -96.
92. Ростовцев Ю.Г. Математические методы и модели оценивания военно-политической обстановки. -МО СССР, 1986. -312 с.
93. Ростовцев Ю.Г., Рахматуллин М.Ю. Формализация целей в системе взаимодействующих роботов// Известия ВУЗ. Приборостроение, 1984, № 11, с. 3 -9.
94. Захарченко А.Н. Обработка информации в комплексных системах мониторинга. -М., 2001. -147 с.
95. Фомин Я.А., Савич А.В. Оптимизация распознающих систем.-М.:Машиностроение,1993.
96. Каляев А.В. Многопроцессорные системы с программируемой архитектурой. -М.: Радио и связь, 1984.
97. Нейман Дж. фон. Теория самовоспроизводящихся автоматов/Пер. с англ. -М.: Мир, 1971.
98. Саати Т.Л. Математические модели конфликтных ситуаций. -М.: Сов. радио, 1977.
99. Ладенко И.С. Интеллектуальные системы в целевом управлении. -Новосибирск: Наука, 1987. -199 с.
100. Яглом И.М. Математические структуры и математическое моделирование. -М.: Сов. радио, 1980.
101. Advanced Satisfiability. Real-World Reasoning Tackling inherent computational complexity. (Based on slides of Carla Gomes, Henry Kautz, Cristopher Moore, Ashish Sabharwal, Bart Selman, Toby Walsh) // Электронный ресурс, [время доступа 21.07.2020], режим доступа: <https://courses.cs.washington.edu/courses/cse573/12sp/lectures/08-advsat.pdf>.
102. Цикритзис Д., Лоховски Ф. Модели данных. -М.: Финансы и статистика, 1985.
103. Цаленко М.Ш. Моделирование семантики в базах данных. -М.: Наука. Гл. ред. физ.-мат. лит., 1989. -288 с.
104. Четвериков В.Н., Ревунков Г.И., Самохвалов Э.Н. Базы и банки данных. -М.: Высшая школа, 1987.
105. Кошелев С.А. Программирование без программирования. Использование утилиты Designer пакета Clarion для разработки программ. -М.: Радио и связь, 1992. -192 с.
106. Forgy Ch. A network match routine for production systems // — Working Paper, 1974.
107. Forgy Ch. On the efficient implementation of production systems // Ph. D. Thesis. — Carnegie-Mellon University, 1979.
108. Forgy Ch. Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem // Artificial Intelligence 19, pp. 17—37, 1982.
109. Питерсон Дж. Теория сетей Петри и моделирование систем/Пер. с англ. -М., 1984.-264с.
110. Котов В. Е. Сети Петри. — М: Наука, 1984. — 160 с.
111. Райбекас А.Я. Вещь, свойство, отношение как философские категории. -Томск: Изд-во Томского университета, 1977.
112. Уемов А.И. Вещи, свойства и отношения. -М.: Изд-во АН СССР, 1963.
113. Орфеев Ю.В., Тютин В.С. Мышление человека и искусственный интеллект. -М.: Мысль, 1978.
114. Collaborative Filing: <http://files.grouplens.org/papers/FnT%20CF%20Recsys%20Survey.pdf>.
115. Matrix decomposition methods:[http://pzs.dstu.dp.ua/DataMining/recom/bibl/1symeonidis\\_panagiotis\\_zioupos\\_andreas\\_matrix\\_and\\_tensor\\_fact.pdf](http://pzs.dstu.dp.ua/DataMining/recom/bibl/1symeonidis_panagiotis_zioupos_andreas_matrix_and_tensor_fact.pdf).
116. Alternating Least Squares: <http://stanford.edu/~rezab/classes/cme323/S15/notes/lec14.pdf>.  
Factorization Machines: <https://cseweb.ucsd.edu/classes/fa17/cse291-b/reading/Rendle2010FM.pdf>.  
Google's Wide'n'Deep: <https://arxiv.org/abs/1606.07792>.
117. DeepFM: <https://arxiv.org/abs/1703.04247>.
118. xDeepFM: <https://arxiv.org/abs/1803.05170>.
119. Neural Collaborative Filtering <https://arxiv.org/pdf/1708.05031.pdf>.
120. Association Rules: <https://web.stanford.edu/class/cs345d-01/rl/ar-mining.pdf>.
121. DSSM: [https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/cikm2013\\_DSSM\\_fullversion.pdf](https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/cikm2013_DSSM_fullversion.pdf).
122. Theory of ML: <https://www.amazon.com/Pattern-Classification-Pt-1-Richard-Duda/dp/0471056693>.
123. Theory of ML: <https://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738>.
124. Sample datasets: <https://grouplens.org/datasets/movielens/100k/>
125. <http://1.https://arxiv.org/abs/1907.06902%20>.
126. [https://www.researchgate.net/publication/335766050\\_Deep\\_language-based\\_critiquing\\_for\\_recommender\\_systems%20](https://www.researchgate.net/publication/335766050_Deep_language-based_critiquing_for_recommender_systems%20).
127. [https://club.cnews.ru/blogs/entry/import\\_zachem\\_facebook\\_kupila\\_whatsapp\\_za\\_19\\_mlrd\\_d3b6%20](https://club.cnews.ru/blogs/entry/import_zachem_facebook_kupila_whatsapp_za_19_mlrd_d3b6%20).
128. <https://ai.googleblog.com/2020/05/federated-analytics-collaborative-data.html?m=1%20>.
129. <https://arxiv.org/abs/1912.04977%20>.
130. <https://arxiv.org/abs/1902.04885>.
131. History of smart cities: Timeline // Verdict. 2020. <https://www.verdict.co.uk/smart-cities-timeline/>.
132. Mark Vallianatos. Uncovering the Early History of “Big Data” and the “Smart City” in Los Angeles // Boom California. 2015. <https://boomcalifornia.com/2015/06/16/uncovering-the-early-history-of-big-data-and-the-smart-city-in-la/>.
133. Показатели «умных» устойчивых городов, разработанные ЕЭК ООН–МСЭ. Записка секретариата / Европейская экономическая комиссия, ООН. [http://www.unece.org/fileadmin/DAM/hlm/projects/SMART\\_CITIES/ECE\\_HBP\\_2015\\_4.pdf](http://www.unece.org/fileadmin/DAM/hlm/projects/SMART_CITIES/ECE_HBP_2015_4.pdf).
134. Smart sustainable city: Shaping smarter and more sustainable cities. Striving for sustainable development goals//ITU-T's Technical Reports and Specifications [http://wftp3.itu.int/pub/epub\\_shared/TSB/ITUT-Tech-Report-Specs/2016/en/flipviewerxpress.html](http://wftp3.itu.int/pub/epub_shared/TSB/ITUT-Tech-Report-Specs/2016/en/flipviewerxpress.html).
135. Рейтинг «Умных» городов International Institute for Management Development, IMD <https://www.imd.org/smart-city-observatory/smart-city-index/>.
136. Рейтинг развития городов / IESE Business School <https://citiesinmotion.iese.edu/>.
137. Рейтинг инновационных городов / 2thinknow <https://www.innovation-cities.com/city-rankings/>.

138. Рейтинг устойчивых городов / Arcadis <https://www.arcadis.com/en/global/our-perspectives/sustainable-cities-index-2018/citizen-centric-cities/>.
139. City Science Application Framework / U4SSC. 2019. <https://www.itu.int/en/publications/Documents/tsb/2019-U4SSC-City-Science-Application-Framework/index.html>.
140. Air quality management in Southern California // Case study of U4SSC. 2019. <https://www.itu.int/en/publications/Documents/tsb/2019-U4SSC-Air-quality-management-in-Southern-California-USA/index.html>, <https://envirosuite.com/>.
141. <https://www.smartdubai.ae/initiatives/happiness-agenda>.
142. Smart Dubai Happiness Meter. Dubai, UAE // Case study of U4SSC. 2019. <https://www.itu.int/en/publications/Documents/tsb/2019-U4SSC-Smart-Dubai-Happiness-Meter-in-Dubai-United-Arab-Emirates/index.html>.
143. Crime prediction for more agile policing in cities – Rio de Janeiro, Brazil // Case study of U4SSC. 2019. <https://www.itu.int/en/publications/Documents/tsb/2019-U4SSC-Crime-prediction-for-more-agile-policing-in-cities-Rio-de-Janeiro-Brazil/index.html#p=1>.
144. <https://rio.crimeradar.org/>.
145. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms / FAT/ML community. <https://www.fatml.org/resources/principles-for-accountable-algorithms>.
146. Постановление Правительства Москвы от 21 ноября 2019 г. № 1543-ПП «О проведении эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы». <http://vestnik.mos.ru/pdf/2019/11nov/66.pdf>.
147. XVIII Ассамблея «Здоровая Москва» // Видеозапись Круглого стола «Искусственный интеллект в здравоохранении» <https://moscowhealth.ru/program>.
148. Постановление Правительства Москвы от 28 апреля 2020 № 464-ПП «О внесении изменений в постановление Правительства Москвы от 21 ноября 2019 № 1543-ПП». <http://vestnik.mos.ru/files/pdf/2020/05may/25.pdf>.
149. Официальный портал эксперимента по использованию инновационных технологий в области компьютерного зрения для анализа медицинских изображений и дальнейшего применения в системе здравоохранения города Москвы. <https://mosmed.ai/>.
150. Методические рекомендации №43 «Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика)» / ГБУЗ «Научно-практический клинический центр диагностики и телемедицинских технологий ДЗМ» [http://medradiology.moscow/f/klinicheskie\\_ispytaniya\\_po\\_12022020.pdf](http://medradiology.moscow/f/klinicheskie_ispytaniya_po_12022020.pdf).
151. Набор данных MosMedData: COVID19\_1110: Результаты исследований компьютерной томографии органов грудной клетки с признаками COVID-19. <https://mosmed.ai/datasets/>, [https://mosmed.ai/documents/50/README\\_RU\\_1.pdf](https://mosmed.ai/documents/50/README_RU_1.pdf).
152. Morozov, S.P., Andreychenko, A.E., Pavlov, N.A., Vladzmyrskyy, A.V., Ledikhova, N.V., Gombolevskiy, V.A., Blokhin, I.A., Gelezhe, P.B., Gonchar, A.V. and Chernina, V.Y., 2020. MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset. arXiv preprint arXiv:2005.06465.
153. <https://www.mos.ru/city/projects/contact-center/>.
154. <https://zakupki.mos.ru/>.
155. Раушенбах Б.В. (2001). Предсказывать – глупейшее занятие.
156. <https://www.un.org/ru/digital-cooperation-panel/>.
157. Цели в области устойчивого развития / Организация Объединённых наций. <https://www.un.org/sustainabledevelopment/ru/sustainable-development-goals/>.
158. Artificial Intelligence (AI) in Manufacturing Market Worth \$27 Billion by 2027- Exclusive Report by Meticulous Research. <https://www.globenewswire.com/news-release/2020/03/30/2008236/0/en/Artificial-Intelligence-AI-in-Manufacturing-Market-Worth-27-Billion-by-2027-Exclusive-Report-by-Meticulous-Research.html>.
159. AI in Oil and Gas Market - Growth, Trends, and Forecast (2020 - 2025). [https://www.reportlinker.com/p05881461/AI-in-Oil-and-Gas-Market-Growth-Trends-and-Forecast.html?utm\\_source=GNW](https://www.reportlinker.com/p05881461/AI-in-Oil-and-Gas-Market-Growth-Trends-and-Forecast.html?utm_source=GNW).
160. Tom Davenport “Shining Up a Rusty Industry with Artificial Intelligence”. <https://www.forbes.com/sites/tomdavenport/2018/02/08/shining-up-a-rusty-industry-with-artificial-intelligence/#745ff0d961c4>.
161. Demand Prediction Using a Soft-Computing Approach: A Case Study of Automotive Industry. <https://www.mdpi.com/2076-3417/10/3/829>.
162. Volkswagen AG Data: Lab Munich. <https://www.volkswagenag.com/en/news/stories/2018/11/learning-to-learn.html>.
163. Scaling AI in Automotive Industry. <https://www.capgemini.com/wp-content/uploads/2019/03/Ai-in-automotive-research-report.pdf>.
164. Remoto by Bright Box. <https://remoto.com/>
165. How Artificial Intelligence is Transforming Automotive Industry Monetization Models. <https://remoto.com/white-paper-artificial-intelligence/Artificial-Intelligence-transforms-the-automotive-industry.pdf>.
166. Predictive layer. <https://www.predictivelayr.com/>
167. Predictive layer data correlation forecasts use cases. <https://www.predictivelayr.com/data-correlation-forecasts-usecases>.
168. Monthly energy consumption forecast: A deep learning approach. <https://ieeexplore.ieee.org/document/7966398>.
169. Conventional models and artificial intelligence-based models for energy consumption forecasting: A review. <https://www.sciencedirect.com/science/article/abs/pii/S092041051930600X>.
170. Applications of machine learning methods for engineering risk assessment – A review. <https://www.sciencedirect.com/science/article/pii/S0925753519308835>.
171. Mathieu Ducros and Félix Gonçalves; Kognitus. Artificial Intelligence and Petroleum System Risk Assessment. <https://www.geoexpro.com/articles/2020/03/artificial-intelligence-and-petroleum-system-risk-assessment>.
172. Oil and Gas industry: Will Machine Learning algorithms be helpful? <https://bitrefine.group/industries/big-data-manufacturing/107-articles/ml-articles/manufacturing-ml-article/269-oil-and-gas-industry-will-machine-learning-algorithms-be-helpful>.
173. Artificial Intelligence Optimizes Oil and Gas Production. <https://pubs.spe.org/en/ogf/ogf-article-detail/?art=6989>.
174. Smart Oil Platform Strikes Multiple Benefits with AI. <https://medium.com/bcggamma/smart-oil-platform-strikes-multiple-benefits-with-ai-da8824eeabc1>.
175. Artificial intelligence methods for oil price forecasting: a review and evaluation. <https://link.springer.com/article/10.1007/s12667-015-0151-y>.
176. Lee Jellum. Brilliant Manufacturing Suite. [https://www.ge.com/jp/sites/www.ge.com.jp/files/C-2\\_Lee-Jellum\\_Eng.pdf](https://www.ge.com/jp/sites/www.ge.com.jp/files/C-2_Lee-Jellum_Eng.pdf).
177. ARTIFICIAL INTELLIGENCE IN LOGISTICS. A collaborative report by DHL and IBM on implications and use cases for the logistics industry - 2018. [http://www.globalhha.com/doclib/data/upload/doc\\_con/5e50c53c5bf67.pdf](http://www.globalhha.com/doclib/data/upload/doc_con/5e50c53c5bf67.pdf).

178. Artificial intelligence in manufacturing: Optimization of additives consumption. <https://bitrefine.group/industries/big-data-manufacturing/93-industries/manufacturing/manufacturing-solutions/255-artificial-intelligence-in-manufacturing-consumption>.
179. <https://www.tadviser.ru/>: “Как машинное обучение помогает сталеварам. Кейс НЛМК и «Инфосистемы Джет»”
180. Accenture – Artificial Intelligence and Blockchain. Insights and Actions for the Chemical Industry. [https://www.accenture.com/\\_acnmedia/PDF-102/Accenture-AI-Blockchain-Chemicals.pdf](https://www.accenture.com/_acnmedia/PDF-102/Accenture-AI-Blockchain-Chemicals.pdf).
181. Bitrefine Group - <https://bitrefine.group/>
182. Smartening up with Artificial Intelligence (AI) - What’s in it for Germany and its Industrial Sector? McKinsey Report, 2017. [https://deepsense.ai/wp-content/uploads/2018/04/170419\\_mckinsey\\_ki\\_final\\_m.pdf](https://deepsense.ai/wp-content/uploads/2018/04/170419_mckinsey_ki_final_m.pdf).
183. Machine Vision in Automotive Industry. <https://devisionx.com/industrial-applications/automotive/>
184. Artificial Intelligence Supports BMW Quality Assurance. <https://metrology.news/artificial-intelligence-supports-bmw-quality-assurance/>
185. Capgemini Research Institute Report. “Scaling AI in Manufacturing Operations: A Practitioners’ Perspective.” <https://www.capgemini.com/gb-en/wp-content/uploads/sites/3/2019/12/Report-%E2%80%93-AI-in-Mfg-Ops.pdf>.
186. “Anomaly Detection — Another Challenge for Artificial Intelligence.” <https://medium.com/sciforce/anomaly-detection-another-challenge-for-artificial-intelligence-c69d414b14db>.
187. Sciforce – Anomaly detection solution. <https://ai.sciforce.solutions/anomaly-detection.html>.
188. DWHorison “DATA-DRIVEN DECISIONS IN FINANCIAL SERVICES”. <http://www.dwhorizon.com/lorem-ipsum-dolor-sit-amet-2/>
189. <https://retailrocket.ru/blog/cases/prediktivnyj-marketing-v-internet-magazine-hoff-rost-vhodyashego-oborota-na-13/>
190. <https://retailrocket.ru/blog/sekretyi-retention-rate/>
191. <https://retailrocket.ru/blog/data-driven-marketing-kak-ispolzovat-polzovatelskie-dannye/>
192. <https://habr.com/ru/company/retailrocket/blog/261593/>
193. Эрик Сигель. Просчитать будущее: Кто кликнет, купит, сойдёт или умрёт = Predictive Analytics. — М.: Альпина Паблицер, 2014. — 374 с.
194. Trends in Predictive Analytics Market, Zion Marker Research: <https://www.globenewswire.com/news-release/2018/03/02/1414176/0/en/Trends-in-Predictive-Analytics-Market-Size-Share-will-Reach-10-95-Billion-by-2022.html>.
195. Artificial Intelligence at Disney, Viacom, and Other Entertainment Giants, Emerj: <https://emerj.com/ai-sector-overviews/ai-at-disney-viacom-and-other-entertainment-giants/>.
196. T-mobile reduces churn by up to 50% by identifying and retaining highly-influential ‘tribe leader’ customers with advanced predictive modelling: [https://www.bestpractice.ai/studies/t\\_mobile\\_reduces\\_churn\\_by\\_up\\_to\\_50\\_by\\_identifying\\_and\\_retaining\\_highly\\_influential\\_tribe\\_leader\\_customers\\_with\\_advanced\\_predictive\\_modelling#](https://www.bestpractice.ai/studies/t_mobile_reduces_churn_by_up_to_50_by_identifying_and_retaining_highly_influential_tribe_leader_customers_with_advanced_predictive_modelling#)
197. Предиктивная аналитика и Big Data помогают «Дом.ру» увеличивать выручку, Comnews: <https://www.comnews.ru/content/113554/2018-06-18/prediktivnaya-analitika-i-big-data-pomogayut-domru-velichivat-vyruchku>.
198. Norma Contract Management, MTC: [www.norma.mts.ru](http://www.norma.mts.ru).
199. Исследование SAS и ACFE, 2019: 72% организаций планируют использовать ИИ против мошенников: <https://www.it-world.ru/it-news/analytics/147052.html>.
200. China Mobile uses big data and AI to curb telecom fraud, ChinaDaily: <http://www.chinadaily.com.cn/a/201809/13/WS5b99cd1da31033b4f4655cc3.html>.
201. Ai T., Yang Z., Hou H., Hou H., Zhan Ch., Chen Ch., Lv W. et al. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*. 2020. 200642. DOI: 10.1148/radiol.2020200642.
202. Simpson S, Kay FY, Abbara S, Bhalla S, Chung JH, Chung M, Henry TS, Kanne JP, Kligerman S, Ko JP, Litt D. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA. *Radiology: Cardiothoracic Imaging*, 2020. doi.org/10.1148/ryct.2020200152.
203. Rubin G.D., Ryerson C.J., Haramati L.B., Sverzellati N., Kanne J. P., Raof S. et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society [published online ahead of print, 2020 Apr 7]. *Radiology*. 2020; 201365. DOI: 10.1148/radiol.2020201365.
204. Revel M.P., Parkar A.P., Prosch H., Silva M., Sverzellati N., Gleeson F. et al. COVID-19 patients and the Radiology department – advice from the European Society of Radiology (ESR) and the European Society of Thoracic Imaging (ESTI). *Eur. Radiol*. 2020. DOI: 10.1007/s00330-020-06865-y.
205. Синицын В.Е., Тюрин М.Е., Митьков В.В. Временные согласительные методические рекомендации Российского общества рентгенологов и радиологов (РОПР) и Российской ассоциации специалистов ультразвуковой диагностики в медицине (РАСУДМ) «Методы лучевой диагностики пневмонии при новой коронавирусной инфекции COVID-19» (версия 2). *Вестник рентгенологии и радиологии*, 2020, doi: 10.20862/0042-4676-2020-101-2-72-88.
206. Bullock J, Pham KH, Lam CS, Luengo-Oroz M. Mapping the Landscape of Artificial Intelligence Applications against COVID-19. *arXiv preprint arXiv:2003.11336*. 2020 Mar 25.
207. Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, Cao K et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. 2020 Mar 19:200905. doi: 10.1148/radiol.2020200905.
208. Kundu Sh., Elhalawani, Gichoya J.W., Kahn Ch.E. How Might AI and Chest Imaging Help Unravel COVID-19’s Mysteries? *Radiology: Artificial Intelligence*. 2020. doi.org/10.1148/radiol.2020200905.
209. Христенко Е.А., фон Стакельберг О., Кауцор Х.-У., Лайер Г., Ридэн Т.В. КТ-паттерны при COVID-19 ассоциированных пневмониях – стандартизация описаний исследований на основе глоссария общества Флейшнера.
210. *REJR* 2020; 10(1):16-26. DOI:10.21569/2222-7415-2020-10-1-16-26.
211. Morozov et al. MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset. Preprint <https://arxiv.org/abs/2005.06465>.
212. 2019 Опыт разработки и внедрения системы поиска онкологических образований с помощью искусственного интеллекта на примере рентгеновской компьютерной томографии легких. Дрокин И.С., Еричева Е.В., Бухвалов О.Л., Пилус П.С., Малыгина Т.С., Синицын В.Е. в журнале *Врач и информационные технологии. Искусственный интеллект в здрав*, № 3, с. 48-57.
213. Данные по выручке и затратам на R&D за 2019 год приведены из ежегодного отчета SAP Annual Report 2019: <https://www.sap.com/docs/download/investors/2019/sap-2019-annual-report-form-20f.pdf>.
214. Рыночная капитализация взята на 26.06.2020 по данным <https://www.macrotrends.net/stocks/charts/SAP/sap-se/market-cap>.

215. Информация по общему числу публикаций по предсказательной аналитике и смежным тематикам взята с сайта <https://www.sap.com/products/intelligent-technologies/artificial-intelligence/research.html> на 28.07.2020.
216. Данные по выручке и затратам на R&D приведены из ежегодного отчета SAS Annual Report: <https://www.sas.com/content/dam/SAS/documents/corporate-collateral/annual-report/company-overview-annual-report.pdf>.
217. Рыночная капитализация взята на 26.06.2020 по данным [https://ycharts.com/companies/SASDY/market\\_cap](https://ycharts.com/companies/SASDY/market_cap).
218. Всего патентов и публикаций по предсказательной аналитике и смежным тематикам, по данным Microsoft Academic (<http://academic.microsoft.com/>) на 15.07.2020.
219. Выручка и затраты на R&D за 2019 год приведены из ежегодного отчета Netflix Annual Report 2019: [https://s22.q4cdn.com/959853165/files/doc\\_financials/2019/ar/2019-10-K.pdf](https://s22.q4cdn.com/959853165/files/doc_financials/2019/ar/2019-10-K.pdf).
220. Рыночная капитализация взята на 30.06.2020 из: <https://www.macrotrends.net/stocks/charts/NFLX/netflix/market-cap>.
221. По данным: <https://research.netflix.com/articles?q=Machine%20Learning>.
222. Данные по выручке и затратам на R&D за 2019 год приведены из ежегодного отчета IBM Annual Report: [https://www.ibm.com/annualreport/assets/downloads/IBM\\_Annual\\_Report\\_2019.pdf](https://www.ibm.com/annualreport/assets/downloads/IBM_Annual_Report_2019.pdf).
223. Рыночная капитализация взята на 28.07.2020 из: <https://www.macrotrends.net/stocks/charts/IBM/ibm/market-cap>.
224. Информация по общему числу публикаций и патентов по предсказательной аналитике и смежным тематикам по данным Microsoft Academic (<http://academic.microsoft.com/>) на 28.07.2020.
225. Данные по выручке и затратам на R&D за 2019 год приведены из ежегодного отчета Baidu Annual Report 2019: <http://ir.baidu.com/static-files/ee02be35-ab39-496b-9119-3f30a5e99e6f>.
226. рыночная капитализация взята на 29.07.2020 из: <https://www.macrotrends.net/stocks/charts/BIDU/baidu/market-cap>.
227. Информация по общему числу публикаций и патентов по предсказательной аналитике и смежным тематикам по данным Microsoft Academic на 28.07.2020.
228. Данные по выручке и затратам на R&D за 2019 год приведены из ежегодного отчета Alibaba Annual Report 2019: [https://www.alibabagroup.com/en/news/press\\_pdf/p190515.pdf](https://www.alibabagroup.com/en/news/press_pdf/p190515.pdf).
229. Рыночная капитализация взята на 29.07.2020 из: <https://www.macrotrends.net/stocks/charts/BABA/alibaba/market-cap>.
230. Информация по общему числу публикаций и патентов по предсказательной аналитике и смежным тематикам по данным Microsoft Academic на 28.07.2020.
231. данные по выручке и затратам на R&D за 2019 год приведены из ежегодного отчета Huawei Annual Report 2019: <https://www.huawei.com/en/annual-report/2019>.
232. Рыночная капитализация взята на 28.07.2020 из: <https://finance.yahoo.com/quote/002502.sz/>
233. Информация по общему числу патентов и публикаций в целом по ИИ взята с Microsoft Academic на 28.07.2020.
234. Данные по выручке и затратам на R&D за 2019 год приведены из ежегодного отчета Facebook Annual Report 2019: <https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Fourth-Quarter-and-Full-Year-2019-Results/default.aspx>.
235. Рыночная капитализация взята на 26.06.2020 из: <https://www.macrotrends.net/stocks/charts/SAP/sap-se/market-cap>.
236. Информация по общему числу публикаций и патентов по предсказательной аналитике и смежным тематикам по данным Microsoft Academic на 28.07.2020.
237. Данные по выручке и затратам на R&D за 2019 год приведены из ежегодного отчета Amazon Annual Report 2019: [https://www.annualreports.com/HostedData/AnnualReports/PDF/NASDAQ\\_AMZN\\_2019.pdf](https://www.annualreports.com/HostedData/AnnualReports/PDF/NASDAQ_AMZN_2019.pdf).
238. Рыночная капитализация взята на 28.07.2020 из: <https://www.macrotrends.net/stocks/charts/AMZN/amazon/market-cap>.
239. Информация по общему числу публикаций и патентов по предсказательной аналитике и смежным тематикам по данным Microsoft Academic на 28.07.2020.
240. Данные по выручке и затратам на R&D за 2019 год приведены из ежегодного отчета Google Annual Report 2019: [https://abc.xyz/investor/static/pdf/20200429\\_alphabet\\_10Q.pdf?cache=1a4ecd7](https://abc.xyz/investor/static/pdf/20200429_alphabet_10Q.pdf?cache=1a4ecd7).
241. Рыночная капитализация взята на 28.07.2020 из: <https://www.macrotrends.net/stocks/charts/GOOGL/alphabet/market-cap>.
242. Информация по общему числу публикаций и патентов по предсказательной аналитике и смежным тематикам по данным Microsoft Academic на 28.07.2020.
243. Информация предоставлена компанией.
244. Консолидированный оборот Группы Компаний Ланит за 2019 год по данным: <https://www.lanit.ru/about/>.
245. Выручка за 2019 год по данным компании: [https://www.kaspersky.com/about/press-releases/2020\\_kaspersky-reports-2019-financial-results](https://www.kaspersky.com/about/press-releases/2020_kaspersky-reports-2019-financial-results).
246. Выручка взята за 2019 год из ежегодного отчета MTS: [https://moskva.mts.ru/upload/contents/10677/mts\\_ifrs\\_cons\\_fs\\_18-19-rus.pdf](https://moskva.mts.ru/upload/contents/10677/mts_ifrs_cons_fs_18-19-rus.pdf).
247. Капитализация взята на 20 августа 2020 года по данным: <https://finance.yahoo.com/quote/MBT/>.
248. По данным: <https://dsmedia.pro/company/zyfra/analytics/2019-results>.
249. Выручка и затраты на R&D за 2019 год по данным годового отчета Yandex N.V. зв 2019 год [https://company-docs.s3.yandex.net/prospectus/annual\\_2019.pdf](https://company-docs.s3.yandex.net/prospectus/annual_2019.pdf).
250. Данные взяты на 20 августа 2020: [https://ycharts.com/companies/YNDX/market\\_cap](https://ycharts.com/companies/YNDX/market_cap).
251. Всего патентов и публикаций по предсказательной аналитике и смежным тематикам, по данным Microsoft Academic (<http://academic.microsoft.com/>) на 20.08.2020.
252. Капитализация взята на 20 августа 2020 года по данным: <https://finance.yahoo.com/quote/gzpfy?ltr=1>
253. Выручка взята за 2019 год из ежегодного отчета компании Газпром: <https://www.gazprom.ru/investors/disclosure/reports/2019/>
254. <https://ai.googleblog.com/2018/10/introducing-adanet-fast-and-flexible.html>
255. Капитализация взята на 20 августа 2020 года по данным: <https://www.bloomberg.com/quote/FIVE:LI>
256. Выручка взята за 2019 год из ежегодного отчета X5 Retail Group: <https://www.x5.ru/ru/Pages/Investors/Reports.aspx>

## ДИСКЛЕЙМЕР

Настоящий аналитический отчет подготовлен на основе анализа открытых источников, в том числе научных публикаций, патентов, СМИ, сайтов компаний, сайтов университетов и других. Другим важным источником информации был опрос экспертов индустрии, проводившийся путем анкетирования при помощи Google Forms. Также мы использовали закрытые системы поиска и машинного анализа больших текстов, которые обрабатывают открытые источники, такие как Microsoft Academic Graph и Dimensions.ai.

В приложении приводятся ссылки на источники и процедуры получения тех или иных цифр. При такой методике сбора и обработки информации, возможно, не все компании и люди попали в наше рассмотрение. Мы не включали в рассмотрение военные применения технологий ИИ, так как по ним нет открытой достоверной информации. Данный аналитический отчет выражает мнение редакции и может не совпадать с официальной позицией Центра Национальной технологической

инициативы на базе МФТИ по направлению «Искусственный интеллект» и/или его членов и/или партнеров настоящего издания.

## БЛАГОДАРНОСТИ

Мы благодарим наших партнеров:

- Компанию МТС и лично Аркадия Сандлера — за помощь в издании печатной версии Альманаха.
- Сколтех и персонально Марию Пукальчик за предоставление данных по авторам, на базе публикаций в открытых источниках;
- АйПи Лабораторию за предоставление информации, помощь в верстке и подготовке инфографики, размещение на сайте и продвижение.

Все права принадлежат Центру компетенций Национальной технологической инициативы на базе МФТИ по направлению «Искусственный интеллект». Распространение Альманаха разрешено только в виде публикации ссылки на сайт Альманаха <http://www.aiReport.ru>, копирование и перепост содержимого Альманаха или его части или выкладывание файла с Альманахом или его части без письменного согласования правообладателя запрещается.

© Центр компетенций Национальной технологической инициативы на базе МФТИ по направлению «Искусственный интеллект», сентябрь 2020

## РЕДАКЦИЯ

### Руководитель проекта:

Игорь Пивоваров

### Главный редактор:

Сергей Шумский

### Аналитики:

Наталья Гутенева, Илья Северов

### Программист:

Леонид Картушин

### Координатор верстки и графики:

Елена Чинарина

### Верстка:

Юрий Брусницин

### Инфографика:

Александр Кей

## АВТОРЫ

### Авторы статей:

авторы указаны в начале каждой статьи

### Авторы текстов разделов:

Игорь Пивоваров, Наталья Гутенева

### Авторы концепции:

Игорь Пивоваров, Сергей Шумский

Утверждение концепции — 01.06.2020

Отдано в верстку — 01.08.2020

Подписано в печать — 31.08.2020

Мы приглашаем все заинтересованные организации к партнерству и участию в следующих номерах Альманаха! Пишите нам на сайте Альманаха <http://www.aiReport.ru> или на почту [org@opentalks.ai](mailto:org@opentalks.ai)

Печатный тираж заказан и распространяется ООО «АйПи Лаборатория» по соглашению с Центром Национальной технологической инициативы на базе МФТИ по направлению «Искусственный интеллект»

## ООО «АЙЛАБС»

119234, город Москва, улица Ленинские Горы, дом 1 стр. 75-А, комната 5

Тел.: +7 499 99 IP LAB

E-Mail: [is@ipaccelerator.ru](mailto:is@ipaccelerator.ru)

Подписано в печать: 31.08.2020. Формат 60×90/8

Бумага мелованная. Печать офсетная. Усл.печ.л. 7,5.

Тираж 1000 экз.

Отпечатано в соответствии с предоставленными материалами в ООО «АЙЛАБС». 119234, г. Москва, ул. Ленинские Горы, дом № 1, строение 75-А, комната 5

[www.cherryPie.ru](http://www.cherryPie.ru)

ISBN 978-5-6045197-2-1



9 785604 519721

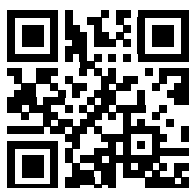
Федеральный закон №436-ФЗ:  
Издание не подлежит маркировке  
в соответствии с п. 1 ч. 2 ст. 1

## 2019

Июнь	ИИ в мире и в России. Стратегии
Сентябрь	Обработка естественного языка, распознавание и синтез речи
Декабрь	Computer Vision

## 2020

Март	Итоги 2019 года
<b>Сентябрь</b>	<b>Предсказательная аналитика и системы поддержки принятия решений</b>



[www.AIreport.ru](http://www.AIreport.ru)  
здесь Вы можете  
получить электронную  
копию Альманаха  
с кликабельными  
ссылками



[www.ozon.ru](http://www.ozon.ru)  
здесь Вы можете  
получить Печатную  
версию