Лекция 5. Прогнозная аналитика

Принято выделять **4 вида аналитики данных**. Они отличаются уровнем сложности работы с информацией и степенью человеческого участия:

1. Описательная (дескриптивная)
2. Диагностическая
3. ***Предиктивная (прогнозная, предсказательная)***
4. Предписывающая (предписательная)

На слайде представлен рисунок видов аналитики данных и вопросы, на которые отвечает каждая из них

- *Описательная (дескриптивная)* отвечает на вопрос «Что случилось?», создавая сводку исторических данных для их дальнейшего анализа, т.е. используется для обобщения и визуализации именно исторических данных. Другими словами, она говорит организациям о том, что уже произошло.

Описательная аналитика — это самый простой вид анализа. Она может быть представлена, например, в виде простой диаграммы с показателями продаж в прошлом году. От достоверности описательной аналитики зависит каждое аналитическое действие. Многие компании по-прежнему полагаются в основном на эту форму аналитики, которая включает в себя информационные панели, визуализированные данные и инструменты отчетности.

Например, непрерывный сбор информации с производственного оборудования с помощью smart-датчиков и других IoT-устройств позволит точно идентифицировать момент сбоя в технологическом процессе.

2-ой вид аналитики: Диагностическая. Анализирует информацию, чтобы ответить на вопрос «Почему это случилось?».

По мере того как аналитика становится более зрелой, организации начинают применять уже более строгие требования к своим историческим данным. Диагностическая аналитика анализирует не просто произошедшие события, а их причину. Для ее выполнения требуется, чтобы аналитики могли отправлять подробные запросы для выявления тенденций и причинно-следственных связей. При использовании диагностической аналитики можно обнаружить новые взаимосвязи между переменными, например: для компании, продающей спортивную одежду, увеличение объемов продаж может быть связано с солнечной погодой. Диагностическая аналитика сопоставляет данные с моделями и позволяет объяснить аномальные или выпадающие данные.

Здесь используются статистические методы анализа данных с целью их кластеризации, классификации, детализации и обнаружения корреляции, чтобы выявить основные факторы влияния на результаты.

В рассмотренном выше примере с промышленным интернетом вещей диагностическая аналитика покажет, что авария случилась по причине выхода из строя модуля приемки сырья.

Первые два вида аналитики рассматривали исторические данные. Следующие 2 вида аналитики: прогнозная и предписывающая сосредоточены вокруг будущих событий.

- ***Предиктивная (прогнозная, предсказательная)*** на основе анализа накопленной информации прогнозирует неизвестные события в будущем, отвечая на вопрос «Что может случиться?». Здесь используется множество различных методов: математическая статистика, моделирование, машинное обучение и другие области **Data Science**, а также интеллектуальный анализ данных (**Data Mining**).

К примеру, предиктивная аналитика текущих и прошлых показателей работы производственного оборудования заблаговременно определит время его профилактического ремонта, чтобы избежать поломки дорогостоящей техники.

- 4-ый вид аналитики: **Предписывающая (предписательная**). Она отвечает, пожалуй, на главный управленческий вопрос «Что делать?». Здесь машинное обучение и другие методы искусственного интеллекта анализируют все накопленные и обработанные данные, чтобы найти наилучшие решения для конкретной ситуации.

В рассматриваемом примере модуль предписывающей аналитики подскажет, какая именно деталь производственного оборудования больше всего износилась и как это исправить наиболее оптимальным с точки зрения экономики образом: заменить на новую или отремонтировать.

На слайде представлена **Аналитическая пирамида**: от описательной к предписывающей аналитике данных.

Что же такое Прогнозная аналитика (ПА) – новое оружие в арсенале ведущих мировых компаний и органов государственного управления. Благодаря развитию информационных технологий открылись новые возможности по использованию больших массивов данных для прогнозирования поведения обычных людей. ПА помогает эффективнее управлять финансами, с высокой точностью прогнозировать объем продаж товаров, предвосхищать желания клиентов и целевую аудиторию новых продуктов, модернизировать технологии, улучшать здравоохранение и образование и даже бороться с преступностью.

Прогнозная аналитика (предсказательная, предиктивная аналитика от англ. predictive analytics) – класс методов анализа данных, который концентрируется на прогнозировании будущего поведения объектов и субъектов с целью принятия оптимальных решений.

История прогнозной аналитики берет свое начало с 40-х годов прошлого столетия, когда команда под руководством Алана Тьюринга пытались взломать шифровальную машину Фашистской Германии “Энигма”. Сложность данной операции заключалась в том, что алгоритм Энигмы менялся каждые 24 часа, и его не успевали взламывать. Алан Тьюринг, британский математик, изобретатель вычислительной “Машины Тьюринга” предположил, что в любом случае есть какая-либо корреляция между символами, осталось лишь ее вычислить. Но для этого требовалось хоть что-то, что есть в каждом зашифрованном сообщении. Немцев подвела их идеология – в каждом их сообщении была ритуальная для них фраза “Да здравствует Гитлер”. Вычислив алгоритм составления этой фразы, а именно, соответствие зашифрованных символов символам из реальной фразы, команде тьюринга удалось разгадать “Код Энигмы”. Кстати, считается, что как раз это и позволило значительно сократить Вторую мировую войну.

После этот метод предиктивной аналитики начал успешно применяться и в других вычислительных задачах.

Сейчас *прогнозная аналитика* – важная, стремительно развивающаяся отрасль науки. Способная предсказывать ваше будущее поведение и выявлять ваши намерения, она представляет собой чрезвычайно мощный инструмент, имеющий значительный потенциал для злоупотреблений. Им следует пользоваться с предельной осторожностью.

ПА дает преимущество перед конкурентами. Сами подумайте: если Вы всегда владеете прогнозной информацией, знаете, что будет с бизнесом завтра, и какое решение более оптимальное в этой ситуации.

100%-ная точность прогнозирования невозможна. Даже погода прогнозируется всего лишь с 50 %-ной точностью, а предсказать поведение людей, будь то пациентов, клиентов или преступников, ничуть не проще.

Прогноз и не должен быть точным на 100 %, чтобы представлять собой большую ценность. Например, одним из самых простых и эффективных применений технологии прогнозирования в коммерческой области является выбор целевой группы для прямой почтовой рассылки рекламных материалов. Если маркетологи могут выявить определенную группу людей, которые, скажем, отреагируют на эти материалы положительно с вероятностью в три раза большей, чем средний потребитель, компания может существенно сэкономить, удалив из списка рассылки людей, которые «не реагируют» на рекламу. А эти люди, в свою очередь, выиграют оттого, что получат по почте меньше спама.

Прогнозирование, даже не отличающееся высокой точностью, всегда лучше создает реальную стоимость, чем чистые догадки. Гораздо лучше иметь хотя бы смутное представление о том, что произойдет в будущем, чем пребывать в полной неизвестности. В этом заключается первый ***эффект*** прогнозирования, который звучит следующим образом: малым достигается многое.

Всего **эффектов**, на которые опирается прогнозная аналитика, пять:

1. *Эффект прогнозирования*: малым достигается многое.
2. *Эффект данных*: данные всегда обладают прогнозным потенциалом.
3. *Эффект индукции*: машинным обучением движет искусство. Стратегии, частично являющиеся продуктом неформальной человеческой творческой мысли, будучи оформлены в виде компьютерных программ, дают успешные результаты при разработке эффективных прогностических моделей, хорошо проявляющих себя в новых случаях.
4. *Эффект ансамбля*: объединенные в ансамбль прогнозные модели компенсируют недостатки друг друга; следовательно, ансамбль моделей в большинстве случаев обладает большей прогнозной точностью, чем составляющие его модели.
5. *Эффект воздействия*: несмотря на свою нематериальную природу, подверженность человека влиянию может быть спрогнозирована при помощи методики моделирования воздействия (uplift modeling), предполагающей построение прогнозной модели на основе двух различных обучающих наборов данных, отражающих результаты применения двух альтернативных подходов.

Построенная на фундаменте компьютерных наук и статистики и активно развиваемая благодаря научно-исследовательским программам, ПА превратилась в самостоятельную дисциплину. Но ПА шагнула далеко за пределы теоретической науки и стала мощным практическим инструментом, оказывающим непосредственное влияние на повседневную жизнь. Ежедневно она влияет на миллионы решений, касающихся того, кому позвонить, отправить почту, назначить диагностику или профилактические мероприятия, кого пригласить на свидание, предостеречь или посадить в тюрьму. ПА дает возможность принимать *персонализированные* решения в отношении каждого человека. Отвечая на массу мелких вопросов, ПА на самом деле может дать нам ответ на ключевой вопрос: *как можно повысить эффективность всех этих многосложных функций в таких сферах, как государственное управление, здравоохранение, бизнес, правоохранительная и некоммерческая деятельность?*

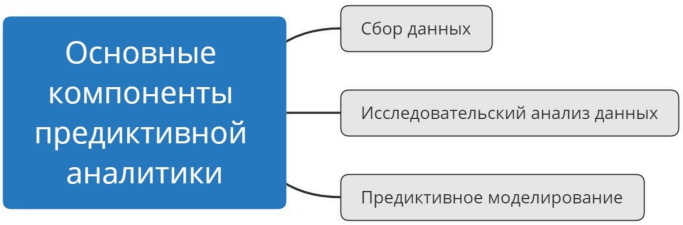
Таким образом, ПА кардинально отличается от стандартного прогнозирования. Последнее производит совокупные прогнозные оценки на макроскопическом уровне. Как будет развиваться экономика? Какой кандидат в президенты наберет больше голосов в Башкирии? Если прогнозная оценка скажет, сколько стаканчиков мороженого будет куплено в Пушкино в следующем месяце, то ПА позволит узнать, какие именно жители Пушкино вероятнее всего соблазнятся на эту покупку.

ПА является ведущим направлением в рамках растущей тенденции по принятию решений, «основанных на данных», опирающихся не на «чутье», а на объективные эмпирические факты. В данной области масса замысловатых названий, таких как *наука о данных, бизнес-аналитика, обработка больших данных* и т. п. Хотя ПА входит в каждое из перечисленных определений, эти красочные термины имеют больше отношения к общей культуре и сферам профессиональной компетенции специалистов, занимающихся инновационными и творческими манипуляциями с данными, чем к конкретным технологиям или методам. Это многозначные термины; иногда они могут означать всего лишь стандартные отчеты в Excel – т. е. вещи важные и требующие значительного мастерства, но не опирающиеся на науку или сложную математику. Другими словами, в каждом конкретном случае их наполнение субъективно. Еще один термин **Data Mining** – «извлечение знаний из данных», или интеллектуальный анализ данных – может

использоваться как синоним прогнозной аналитики, но эта образная метафора может описывать и другие способы добычи знаний из данных, а также часто употребляется в более широком смысле.

На первый взгляд это кажется непосильной задачей: переработать миллионы примеров, чтобы узнать, каким образом использовать различные факты, известные о конкретном человеке, чтобы научиться составлять более-менее обоснованные прогнозы. Но эту задачу можно разбить на несколько частей, что намного ее упростит.

Рассмотрим основные компоненты прогнозной аналитики



2 этап. Исследовательский анализ

На данном этапе начинается анализ данных – обнаружение в организованных данных ранее неизвестных, незаурядных, практически полезных, формализуемых знаний, которые необходимы исследователям в различных сферах человеческой деятельности для принятия необходимых решений.

В англоязычных источниках для обозначения сферы анализа данных используется термины **Data Mining** и **Machine Learning**.

Исследовательский анализ информации решает задачи:

1. *Классификация.* Присвоение одного элемента к группе других по определенным параметрам.
2. *Регрессия.* Выявление зависимости результатов от исходных данных.
3. *Кластеризация.* Объединение объектов в группы по различным параметрам.
4. *Ассоциация.* Определение закономерностей между событиями.
5. *Последовательная ассоциация.* Определение, через какое время после одного события случится другое.
6. *Анализ отклонений.* Определение некоторого количества исключений из правил.

3этап предиктивной аналитики - Предиктивное моделирование

То, ради чего и нужна система предсказательной аналитики – создание высокоточных прогнозов.

После предыдущих этапов у аналитика есть массив данных (различные классы, кластеры, зависимости, ассоциации и отклонения от нормы) и необходимо данные интерпретировать.

На данном этапе необходимо:

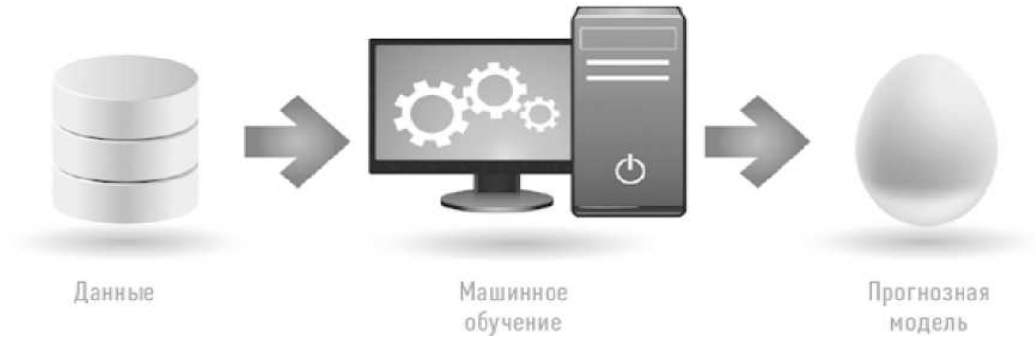
– Поставить задачу перед аналитикой. Прогноз того, что необходимо получить и на какой промежуток времени, или время до определенного события. Это может быть прогноз прибыли на год, спроса на рынке в августе, или сколько еще проработает станок на производстве.

– Выбрать математическую или статистическую модель. Она и сделает этот прогноз. Если проще, то принять во внимание множество факторов, которые влияют на заданный прогноз, распределить их удельный вес в конечном результате и ввести исходные сведения.

Прогнозная модель – механизм, который предсказывает поведение. Например, поведение индивида: предсказывает щелчок мышью, покупку. Прогнозная модель использует в качестве входных данных характеристики конкретного объекта, например, индивида и на выходе выдает прогнозную скоринговую оценку. Чем выше оценка, тем больше вероятность того, что индивид проявит прогнозируемое поведение.

Другими словами, прогнозная модель учитывает все известные характеристики объекта и на их основе вырабатывает прогноз. Существует много способов это сделать. Один из них состоит в том, чтобы оценить влияние каждой характеристики и затем суммировать эти влияния. Если каждый признак увеличивает или уменьшает итоговую скоринговую оценку для конкретного объекта, такая модель называется *линейной;* она считается достаточно простой и ограниченной в своих возможностях, хотя, как правило, это гораздо лучше, чем ничего.

Модель создается посредством машинного обучения:



Модель сама по себе является продуктом машинного обучения. Поэтому машинное обучение также называют *прогнозным моделированием* – обычно в коммерческой сфере употребляется именно этот термин. Если взять ранее упомянутый термин **Data Mining** («извлечение знаний из данных»), то прогнозная модель и есть тот самый добытый бриллиант.

Прогнозное моделирование полностью создает модель с нуля. Все формулы, удельные веса или правила вырабатываются автоматически с помощью компьютера. Для этого и предназначен процесс машинного обучения – механически приобретать новые знания и развивать новые способности, опираясь на анализ данных. Другими словами, присущий ПА «дар предвидения» вырастает из автоматизации.

Прежде чем запустить ПА-систему в действие, организации проверяют ее работоспособность посредством «прогнозирования прошлого» (так называемого *бэктестинга).* Модель должна доказать свою прогнозную точность на исторических данных. Она может тестироваться на данных за прошлую неделю, прошлый месяц или прошлый год. В модель загружаются входные данные, которые были известны в некий исходный момент времени, и она выдает прогноз, который сопоставляется с тем, что фактически произошло в дальнейшем.

Прогнозирование начинается с малого. Строительный элемент ПА - *предикторная переменная,* отдельное значение, измеряемое для каждого объекта. Например, *новизна* -

количество недель, прошедшее с момента последней покупки, совершения последнего преступления или проявления медицинского симптома, - часто отражает вероятность того, что это повторится в ближайшем будущем. Во многих случаях, будь то маркетинговый контакт, уголовное расследование или клиническая оценка, имеет смысл начинать с тех людей, которые проявляли активность в последнее время.

Другим общепринятым и продуктивным параметром является *частота* - сколько раз объект/индивид проявлял данное поведение. Если человек делает что-то достаточно часто, высока вероятность того, что он сделает это снова.

На самом деле именно то, что люди *делали в прошлом,* позволяет спрогнозировать, что они *сделают в будущем.* Поэтому ПА выходит за рамки скучных, но важных демографических данных, таких как место проживания и пол, и обращает пристальное внимание на *поведенческие предикторы,* т. е. прогнозные факторы, такие как новизна, частота, история покупок, финансовая активность и использование продукта. Поведенческие предикторы зачастую являются наиболее ценными из всех, поскольку обычно наша задача - предсказать *поведение*, а поведение предсказывает поведение.

Вот некоторые из наиболее колоритных открытий, каждое из которых соответствует одной предикторной переменной:

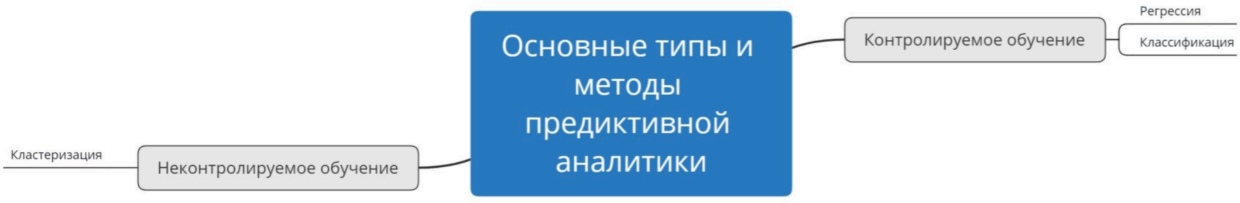
- Среди покупателей подгузников больше потенциальных покупателей пива
- Покупка степлера говорит о найме нового сотрудника
- Пользователи Mac бронируют более дорогие отели
- Ваша предрасположенность к покупкам варьируется в зависимости от времени суток
- Генетика обуславливает неверность жен
- Выход на пенсию подрывает здоровье
- Террористы-смертники не страхуют свою жизнь
- Уровень преступности растет после публичных спортивных мероприятий
- Уровень преступности повышается после выборов
- Голодные судьи склонны выносить более жесткие решения
- Музыкальные вкусы говорят о политических пристрастиях
- Люди с привлекательной внешностью пользуются меньшим вниманием
- Продвижение по службе может привести к уходу сотрудника
- Вегетарианцы реже пропускают авиарейсы
- Сольные рок-музыканты умирают более молодыми, чем участники рок-групп

ПА достигает предсказательной силы путем объединения десятков и даже сотен предикторов. Вы сообщаете машине все, что знаете о каждом человеке, и запускаете программу обработки информации. Ключевая технология обучения, объединяющая эти элементы, - вот где происходит научная магия.

Типы и методы прогнозной аналитики

Прогнозная аналитика использует статистические методы, методы интеллектуального анализа данных, теории игр, анализирует текущие и исторические факты для составления предсказаний о будущих событиях. В бизнесе прогнозные модели используют паттерны, найденные в исторических и выполняемых данных, чтобы идентифицировать риски и возможности. Модели фиксируют связи среди многих факторов, чтобы сделать возможной оценку рисков или потенциала, связанного с

конкретным набором условий, руководя принятием решений о возможных сделках.



На слайде рисунок-схема, на котором обозначены основные типы и методы ПА

Тип 1. Контролируемое обучение

Или обучение с учителем, подразумевает под собой построение (обучение) модели по исходным данным и выходящим результатам. То есть в построении модели известны и параметры события, и результат, на который они влияют.

Например, если мы знаем, что на выручку влияет число покупок и средний чек, а нам необходимо узнать, каким образом влияет тот или иной параметр на её размер, то мы прибегнем к контролируемому обучению. Оно включает два ключевых метода предиктивной аналитики: регрессию и классификацию.

Регрессия

Это самый популярный метод. Применяется для получения количественных ответов или числовой ценности. Например, для расчета выручки по конкретным параметрам. При регрессии используется:

- Числовая переменная ответа. То, что пытаются предсказать.
- Предикторы. Параметры, которые влияют на ответ.

Взаимосвязь между параметрами и результатом и есть предиктивная модель. Кстати, помимо взаимозависимости рассчитывается и вес каждого параметра – то, в какой степени каждый из параметров влияет на конечный результат.

Пример. Даны показатели выручки, среднего чека и количества клиентов за три месяца:

Месяц	Количество клиентов	Средний чек	Выручка
1	10	3 000	30 000
2	11	3 000	33 000
3	10	3 300	33 000

Из этих данных видно, что зависимость выручки от количества клиентов и среднего чека прямая пропорциональная.

Выручка = Количество клиентов * Средний чек.

Зная эту формулу, можно прогнозировать выручку и влиять на нее, сосредотачивая усилия на росте предикторов. Ну или же понять, сколько необходимо привлечь клиентов и при каком среднем чеке, чтобы получить желаемую выручку.

Это выглядит просто, когда известна зависимость. Но даже если в этом уравнении

разложить, из чего складывается количество клиентов, и какой параметр в какой степени влияет на этот показатель, то получится большая и достаточно сложная цепочка.

Классификация

Этот метод связан с причислением объекта к какому-либо классу по определенным параметрам. Его задача определить, к какому именно.

Работает это так: в базу данных загружаются все известные переменные объектов, например, по каждому человеку загружают пол, возраст, профессию и уровень дохода. Далее алгоритм вычисляет зависимость одного от другого и предсказывает неизвестный параметр объекта по известным. Обычно в бизнесе этот метод применяется для различных сегментаций.

Пример. В оптовой торговлей одежды размер скидок зависит от объема закупок товара. Первый способ определить уровень скидки новому клиенту – поработать с ним определенное время.

Если же использовать классификационный метод, то имея информацию о прошлых клиентах, например, о местоположении, об ассортименте, можно рассчитать влияние параметров на объем закупок Вашей продукции. Зная это, можно предугадать, какой объем закупок следует ожидать от нового клиента. Ну, и не стоит забывать, что чем больше данных, тем более точными будут прогнозы.

2 тип прогнозной аналитик: Неконтролируемое обучение

В этом типе предиктивное моделирование происходит только по входящим данным без привязки к ответу. Ответ подбирается автоматически в процессе обучения. Это требуется для поиска и анализа скрытых закономерностей внутри сведений о которых ранее было неизвестно. Основной метод – кластеризация.

2.1. Кластеризация

К этому методу предиктивной аналитики относятся задачи, основными из которых являются:

1. Анализ эффективных схем группировки данных.
2. Проверка гипотез принадлежности одного объекта к проверяемой группе.

Для бизнеса она полезна тем, что на основе кластерного анализа можно более четко представлять взаимосвязи и зависимости. Помимо этого, он помогает выявлять отклонения и новые тенденции.

Пример. Возьмем тот же пример, что и в классификационном методе. Только если там нам и нашей модели уже известна зависимость объема закупок от параметров (местоположение, рекламные вложения и ассортимент), то в этом случае мы их не знаем. Загружаем данные о клиентах и алгоритм определяет, есть ли взаимозависимость между ними, и если есть, то какая.

Выбор инструментов и ПО

Есть много инструментов и программных продуктов. Они отличаются между собой функциональностью и удобством пользования. Некоторые из них нужны для создания предиктивных моделей, некоторые для их интерпретации, а самые продвинутые – для того и другого. При выборе инструмента обратите внимание на:

1. *Поддержку полного цикла аналитики.* От исследования данных до создания моделей и оценки их эффективности.
2. *На интеграцию знаний.* Знания, полученные в процессе аналитики, должны интегрироваться в другие сферы бизнеса.
3. *На поддержку интеграции.* Она необходима с различными источниками получения и обработки данных.
4. *Удобство пользования.* Программа должна быть понятна для разных типов пользователей: от статистиков до менеджеров.
5. *Адаптивность к работе.* Работоспособность с минимальным вмешательством программистов и технических специалистов.

Аналитические системы

Внедрение аналитических **Big Data** систем – это комплексный поэтапный проект. Предписывающая аналитика находится на вершине пирамиды и опирается на предыдущие уровни: предиктивную, диагностическую и описательную. Поэтому для формирования оптимальных управленческих решений на основе данных необходимо, прежде всего, накопить релевантный объем этой информации, достаточный для корректного обучения алгоритмов Machine Learning. Некоторые аналитические задачи решаются с помощью современных прогнозно-аналитических систем, платформ. Рассмотрим некоторые из них.

Название	Цена	Описание	Преимущества
Прогнозно-аналитическая система **Loginom**	Бесплатно	Low-code платформа для реализации всех аналитических процессов: от интеграции и подготовки данных до моделирования, развертывания и визуализации	1. Эксперты в предметных областях могут самостоятельно, без долгого общения с IT-департаментом, решать сложные задачи анализа: комбинировать данных из любых источников, накапливать их, повторно использовать наработки. Это улучшает качество анализа, сокращает время ожидания и увеличивает удовольствие от работы 2. Loginom делает продвинутую аналитику доступной большинству сотрудников. Разработчики избавятся от рутины и направят дефицитные ресурсы для решения важных задач. Пользователи смогут самостоятельно решать большинство задач, публиковать веб-сервисы без

			кодирования, при сохранении безопасности, централизованного контроля и снижении рисков возникновения «теневого IT»
Язык программирования R	Бесплатно	Фаворит рынка, это связано с тем, что в процессе обучения специалистов подобного профиля задействован именно этот язык программирования	1. Открытый исходный код. 2. Расширяемая аналитическая среда. 3. Возможность визуализации представления данных. 4. Большое сообщество пользователей. 5. Разрабатывался статистиками для статистиков.
Язык программирования Python	Бесплатно	Набирает популярность. Основная идея: хороший язык программирования – простой и доступный	1. Простой и интуитивно-понятный. 2. Встроен инструмент для тестирования. 3. Многоцелевой язык.
RapidMiner	Бесплатно	Среда для прогнозной аналитики, которая поддерживает все этапа анализа, проверки, визуализацию и оптимизацию данных	1. Не нужно знать программирования, метод визуального программирования. 2. Расширяемая система, поддержка языка R. 3. Возможность оценки тональности текста. 4. Сообщество пользователей и поддержка новичков.
Knime	Бесплатно	Система для анализа данных, которая даже в базовом функционале имеет мощные инструменты	1. Широкие возможности анализа текста. 2. Возможность веб-анализа, анализа изображений и социальных сетей. 3. Интуитивно-понятный интерфейс без необходимости программирования.
IBM SPSS Modeler	От 80$	Низкая требовательность к новичкам, благодаря	1. Автоматическое моделирование и выбор наиболее эффективное модели.

		автоматическому подбору необходимой статистической модели	2. Геопространственная аналитика. 3. Поддержка технологий с открытым исходным кодом (R, Python). 4. Аналитика текста.
IBM Watson Analytics	От 250$	Один из наиболее мощных инструментов для предиктивной аналитики и анализа больших данных	1. Возможность работы в облаке. 2. Расширенные возможности визуализации. 3. Интуитивно-понятный интерфейс без необходимости программирования. 4. Быстрота обработки данных.
SAS Enterprise Miner	От 160$	Система разработанная для проектирования точных предсказательных и описательных моделей на основе big-data	1. Клиент-серверное решение – позволяет оптимизировать процессы аналитики. 2. Нет необходимости в программировании. 3. Продвинутый скоринг – применение модели к новым данным. 4. Самодокументируемая проектная среда.
SAP BusinessObjects Predictive Analytics	От 200$	SAP в 2015 году был награжден статусом лидера рынка в предсказательной аналитике	1. Большая автоматизированность, легкость в переобучении модели. 2. Расширенные возможности визуализации. 3. Возможность расширения языком R.
Oracle Big Data Preparation	От 150$	Благодаря интуитивному и интерактивному интерфейсу привлекает пользователей без навыков программирования	1. Работа в облаке. 2. Простота использования. 3. Широкие возможности интеграции с другими облачными сервисами.

На практике многие предприятия, вступившие на путь цифровой трансформации, создают собственные системы аналитики больших данных. При этом используются

разнообразные технологии Big Data, например, Apache Hadoop – для хранения информации (в HDFS или HBase), Kafka – для сбора данных из различных источников, а Spark или Storm – для быстрой аналитической обработки потоковой информации. В частности, именно так реализована рекомендательная система стримингового сервиса Spotify. Таким образом, организация предиктивной и, тем более, предписывающей аналитики данных – это одна из ключевых задач цифровизации бизнеса.

Применение аналитики больших данных

Любое применение ПА определяется двумя **факторами:**

1. *Предмет прогнозирования***:** какое поведение, действие или событие должно быть спрогнозировано в отношении конкретного человека, акции или другого субъекта.

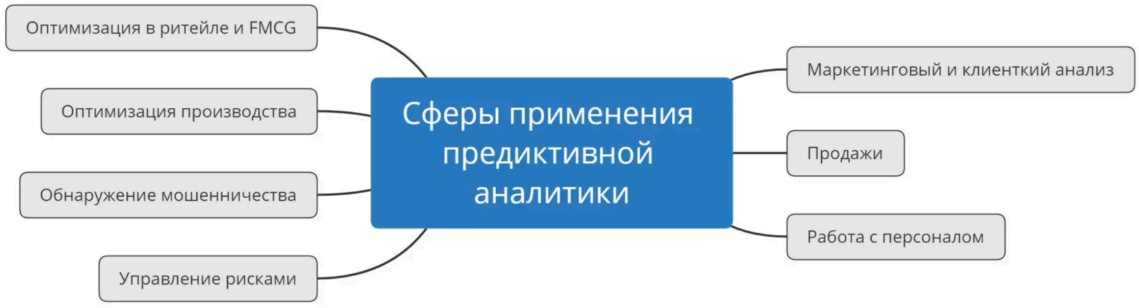
2. *Цель прогнозирования***:** какие решения будут приняты или какие действия предприняты организацией в ответ на каждый прогноз или под его влиянием.

Список потенциальных областей применения ПА неограничен, а перечень уже достигнутых успехов невероятно обширен. Области применения прогнозной аналитики: цены акций, риск, правонарушения, несчастные случаи, продажи, пожертвования, клики, отмены, проблемы со здоровьем, госпитализация, мошенничество, уклонение от уплаты налогов, преступления, неисправности, дебит нефти, отключение подачи электричества, предоставление государственных пособий, мысли, намерения, ответы, мнения, ложь, оценки, отсев учащихся, дружба, романтические отношения, беременность, разводы, рабочие места, увольнения, победы, выборы и многое другое. ПА стремительно проникает во все новые сферы нашей жизни.

В бизнесе свое главное применение ПА находит в области массового маркетинга.

Фактически все, что делает человек, стоит того, чтобы стать предметом прогнозирования, – а именно то, как мы потребляем, думаем, работаем, уходим, голосуем, любим, воспроизводим потомство, разводимся, создаем проблемы, обманываем, воруем, убиваем или умираем.

Рассмотрим некоторые примеры.



1.Нацеливание прямого маркетинга

Предмет прогнозирования: какие клиенты положительно откликнутся на маркетинговый контакт.

Цель прогнозирования: нацеливание маркетинговых усилий на клиентов с наибольшей вероятностью положительного отклика.

Такое использование ПА хорошо иллюстрирует эффект прогнозирования: *малым*

достигается многое.

Как рассчитать величину чистой выгоды, создаваемой благодаря действию эффекта прогнозирования. Допустим, у компании список рассылки, включающий 1 млн потенциальных клиентов. Стоимость прямой почтовой рассылки в расчете на клиента составляет $2, и в прошлом только 1 из 100 человек покупал продукт (т. е. компания получала 10 000 откликов). Предположим, что компания охватывает рассылкой весь список.

Если прибыль компании составляет $220 в расчете на каждый (редкий!) положительный отклик, то она заработает в общей сложности:

Общая прибыль = Выручка - Затраты
= ($220 X 10 000 откликов) - ($2 млн)= $200 000

Прогнозно-аналитическая система говорит, какие клиенты вероятнее всего откликнутся на прямую рассылку, выделяет четверть списка и заявляет: «Эти люди дадут положительный отклик с вероятностью в три раза выше средней». Теперь есть короткий список из 250 000 потенциальных клиентов, из которых 3 %, т. е. 7500 человек, могут стать реальными покупателями.

Если ограничить рассылку только этим коротким списком, можно заработать:

Общая прибыль = Выручка - Затраты
= ($220 X 7500 откликов) - ($2 x 250 000) = $1 150 000

Прибыль увеличилась в 5,75 раза просто за счет того, что разослали рекламные брошюры меньшему числу людей (и при этом также спасли от вырубки несколько деревьев), спрогнозировав, кто вряд ли откликнется на рекламу, и просто оставив этих людей в покое. Таким образом, сократили затраты на три четверти в обмен на снижение продаж всего на одну четверть.

Определить реальную денежную отдачу от прогнозирования несложно. Если составление самих прогнозов требует применения сложных математических методов, то для того, чтобы оценить совокупное влияние на итоговый результат (такой, как прибыль) любого прогноза, точного или не очень, достаточно простейшей арифметики. Прогнозная аналитика - не некая абстрактная наука. Это бизнес.

1. Нацеливание рекламы

Предмет прогнозирования: какое рекламное объявление вероятнее всего спровоцирует клик у каждого конкретного пользователя.

Цель прогнозирования: повышение эффективности отображения рекламных объявлений (на основе вероятности клика, а также вознаграждения, выплачиваемого рекламодателем).

2. Механические торговые системы

Предмет прогнозирования: будет ли цена акции расти или падать.

Цель прогнозирования: использовать прогнозы для покупки акций, которые будут расти в цене, и продажи акций, которые будут падать в цене.

3. Прогнозирование беременности

Предмет прогнозирования: кто из покупательниц в ближайшие месяцы ожидает рождения ребенка.

Цель прогнозирования: делать соответствующие маркетинговые предложения будущим родителям

4. Удержание сотрудников

Предмет прогнозирования: какие сотрудники могут уйти.

Цель прогнозирования: выбор того, какие действия предпринять в отношении своих подчиненных на основе прогнозов, оставляется за руководителями. Это пример применения ПА для поддержки принятия решений, а не для автоматического принятия решений.

5. Прогнозирование преступлений

Предмет прогнозирования: место совершения будущего преступления.

Цель прогнозирования: усиленное патрулирование этого района с целью предотвращения преступлений.

6. Выявление мошенничества

Предмет прогнозирования: какие транзакции или заявки на выдачу кредитов, предоставление льгот, пособий, возмещений и т. п. являются мошенническими.

Цель прогнозирования: повысить эффективность работы инспекторов за счет более точного отбора подозрительных транзакций и заявок.

7. Системы обнаружения вторжений в Сеть

Предмет прогнозирования: какие низкоуровневые интернет-коммуникации исходят от злоумышленников.

Цель прогнозирования: блокирование таких взаимодействий.

8. Фильтрация спама

Предмет прогнозирования: какие сообщения по электронной почте являются спамом.

Цель прогнозирования: направлять подозрительные сообщения в папку со спамом.

9. Настольные игры

Предмет прогнозирования: какая позиция на игровом поле приведет к победе.

Цель прогнозирования: сделать ход, который приведет к такой позиции на игровом поле, которая в свою очередь приведет к победе.

10. Прогнозирование вероятности рецидивизма для правоохранительных органов

Предмет прогнозирования: вероятность повторного совершения преступления.

Цель прогнозирования: учитывать эти прогнозы при вынесении судьями и комиссиями по условно-досрочному освобождению решений о необходимости содержания человека под стражей.

11. Выявление тревожных записей в блогах

Предмет прогнозирования: какие записи в блогах выражают тревогу.

Цель прогнозирования: рассчитать совокупный показатель массового настроения.

12.Удержание клиентов через прогнозирование их ухода

Предмет прогнозирования: какие клиенты могут уйти.

Цель прогнозирования: нацелить маркетинговые усилия по удержанию клиентов на выявленные группы риска.

13.Оценка будущей стоимости ипотечных кредитов

Предмет прогнозирования: кто из держателей ипотечных кредитов может досрочно погасить кредит в течение ближайших 90 дней.

Цель прогнозирования: оценка будущей стоимости ипотечных кредитов для принятия решений об их удержании или продаже другим банкам.

14.Рекомендации фильмов

Предмет прогнозирования: какую оценку клиент поставит фильму.

Цель прогнозирования: рекомендовать клиентам такие фильмы, которые с большой вероятностью им понравятся и получат высокую оценку.

15.Поиск правильного ответа на открытый вопрос

Предмет прогнозирования: оценить правильность ответа посредством прогнозирования правильности пары «вопрос / предполагаемый ответ».

Цель прогнозирования: выявить предполагаемый ответ с наивысшей прогнозной оценкой вероятности и использовать его как окончательный ответ.

16.Образование – повышение эффективности целенаправленного обучения

Предмет прогнозирования: слабые и сильные места в знаниях обучающегося.

Цель прогнозирования: сосредоточить силы на слабых местах, чтобы восполнить пробелы в знаниях.

17.Удержание клиентов при помощи моделирования оттока

Предмет прогнозирования: какие клиенты могут уйти.

Цель прогнозирования: нацелить усилия по удержанию на клиентов из группы риска.

18.Нацеливание маркетинговых усилий через прогнозирование реакции клиентов

Предмет прогнозирования: какие клиенты положительно откликнутся на контакт (например, совершат покупку).

Цель прогнозирования: нацелить усилия на тех клиентов, которые с наибольшей вероятностью положительно откликнутся на контакт.

19.Удержание клиентов при помощи моделирования воздействия

Предмет прогнозирования: каких клиентов можно убедить остаться.

Цель прогнозирования: нацелить усилия по удержанию на этих подверженных воздействию клиентов.

20. Политическая кампания с использованием технологии моделирования воздействия

Предмет прогнозирования: на каких избирателей можно оказать искомое позитивное влияние при помощи контакта, такого как телефонный звонок, посещение, раздача агитационных материалов или телевизионная реклама.

Цель прогнозирования: охватить контактами избирателей, подверженных позитивному влиянию, и избегать контактов с избирателями, для которых существует риск негативного влияния.