



«Альпина Пабlishер»  
Книжное издательство

## «Работа с данными в любой сфере»

Огромные массивы информации помогают удовлетворить человеческие потребности и предсказать будущее. Благодаря своей ценности они становятся предметом кражи, а иногда служат основой для произведений современного искусства. Вопросами анализа, обработки и визуализации больших данных занимается отдельная наука. О том, какие методы, алгоритмы и приемы используются в работе с данными, рассказывает книга Кирилла Еременко [«Работа с данными в любой сфере: Как выйти на новый уровень, используя аналитику»](#), готовящаяся к выходу в издательстве Альпина Пабlishер в переводе Д. Шалаевой. Специально для своих читателей *N + 1* публикует фрагмент, посвященный тому, как большие данные помогают удовлетворить базовые потребности человека.



## Как данные удовлетворяют наши потребности

В науке о данных не так уж много таинственного — она, в конце концов, полностью вписана в современные реалии. И все же преобладает неверное представление, будто данные сложны и даже непостижимы. К сожалению, многие сегодня либо охотно отказываются видеть, как широко применяется наука о данных, либо намеренно отвергают ее как нечто недоступное или неприменимое к их работе. Наука о данных как дисциплина *предполагает* что-то весьма замысловатое. Это похоже на то, чем люди занимаются в маленьких кабинетах без окон, сгорбившись над своими столами.

Такой взгляд совершенно неверен.

В этой главе мы точно узнаем, насколько данные вездесущи, как широко они генерируются и собираются и почему наука о данных никогда не может считаться причудой.

### Проникновение данных

Чтобы проиллюстрировать, насколько важны данные для всех аспектов нашей жизни — что это необходимость, а не роскошь, я буду использовать пирамиду потребностей Маслоу, которая, я уверен, знакома многим бизнес-практикам. В литературе по бизнес-психологии о ней написано очень много. Я считаю, что эта модель на удивление хорошо сочетается с распространенностью и преимуществами данных.\*

\*Это не даст нам исчерпывающих сведений о том, как и где наука о данных используется в нашей жизни, поскольку пирамида Маслоу принижает неосновные человеческие потребности. Такие области, как, например, военная оборона и освоение космического пространства, сюда не будут включены, поскольку они не являются основными потребностями человека. Я использую пирамиду потребностей Маслоу в качестве примера для описания всеобъемлющей силы данных, но, если вы хотите узнать больше о том, как эта иерархия может быть применена в бизнесе, см. Conley (2007) — прим. автора.

Иерархия потребностей была разработана Абрахамом Маслоу в 1943 г. для отображения сложной мотивации, обуславливающей поведение людей. Иерархия представлена в форме пирамиды, которая в последовательности снизу-вверх включает в себя потребности — от наиболее к наименее фундаментальным (рис. 2.1). Короче говоря, иерархия организована таким образом, что потребности, находящиеся на самом нижнем уровне пирамиды, должны быть удовлетворены до того, как у индивидуума, о котором идет речь, появится мотивация для удовлетворения потребностей более высоких уровней.



Рис. 2.1. Пирамида потребностей Маслоу

## Наука о данных и физиология

В основе иерархии Маслоу лежат физиологические факторы — основные потребности людей для простого выживания. Как данные могут поспособствовать лучшему удовлетворению этих основных потребностей? Давайте возьмем в качестве примера воздух, которым мы дышим. Загрязнение воздуха — один из наиболее серьезных поводов для глобального беспокойства со времен промышленной революции конца XVIII и начала XIX в. Мы могли бы считать смог феноменом прошлого — так, в 1950-х гг. выбросы, образовавшиеся при сгорании угля, регулярно окутывали Лондон. Но смесь дыма, тумана и пыли остается большой проблемой во многих городах по всему миру, от Китая до Бразилии. Любые технологии, предназначенные для уменьшения загрязнения воздуха в городах, зависят от данных: чтобы улучшить состояние воздуха, его состав необходимо сначала контролировать.

### Кейс: экологические данные и «Зеленый горизонт»

Программа «Зеленый горизонт» (Green Horizon) была запущена компанией IBM в 2014 г. В связи с необходимостью отреагировать на ужасное качество воздуха в Китае путем «преобразования его национальных энергетических систем и поддержки потребностей в устойчивой урбанизации» (IBM, 2017a). «Зеленый горизонт»\* использует данные 12 глобальных исследовательских лабораторий и применяет когнитивные модели к собранным данным, чтобы предоставить информацию, связанную с главной целью проекта — сокращением загрязнения. Данные необходимы для мониторинга колебаний загрязнения воздуха в отдельных районах, а также для того, чтобы ученые могли проанализировать различные факторы, которые прямо или косвенно влияют на качество, температуру и состояние воздуха, и начать улучшать физическую среду в Китае.

Огромное преимущество этих проектов заключается в том, что экологические данные чаще всего являются общедоступными и в глобальном масштабе. Это означает, что технологические разработки, направленные на борьбу с загрязнением воздуха, могут быстро развиваться. Наличие доступа к важным массивам данных, связанных с удовлетворением наших самых основных потребностей, необходимо для понимания того, как имеющиеся технологии могут работать лучше. Вот почему у нас теперь есть специальные стеклянные панели, которые могут быть установлены в зданиях, чтобы окна могли «дышать», очищая воздух внутри помещения и тем самым защищая находящихся там людей. Вот почему у нас есть фильтры, которые могут быть использованы на фабриках в целях уменьшения вредных выбросов и защиты местных жителей от отравления.

## **Возобновляемые продовольственные ресурсы**

Еда еще один пример того, как данные связаны с самыми основными потребностями человека (физиологические факторы в пирамиде Маслоу). Для некоторых это может показаться научной фантастикой, но уже в течение многих лет еда выращивается в лабораториях, а использование искусственного мяса становится все более актуальным феноменом. Memphis Meats, стартап в Кремниевой долине, который с момента своего создания разработал разные виды искусственного мяса, от говядины до домашней птицы, — всего лишь один из подобных институтов.

Поскольку это все еще некая «серая» область для регулирующих органов, религии и науки, искусственное мясо вызвало и похвалы, и гнев мирового сообщества (Devitt, 2017). Но нравится нам это или нет, искусственное мясо в недалеком будущем может стать заменой того, что мы едим. Резко сократив потребление воды и выбросы углерода, оно станет экологически безопасным решением в условиях, когда сельское хозяйство негативно влияет на мир природы. И данные, которые мы собираем для производства такого мяса, в конечном итоге выйдут за рамки исследования ДНК. Поскольку пищевые технологии становятся все более обыденными, дополнительные потребительские данные будут использоваться для других целей, таких как определение оптимальных способов приготовления искусственного мяса, — это позволит не только сделать мясо вкуснее, но и, что особенно важно для производящих компаний, повысить его продаваемость.

## **Наука о данных и безопасность**

Как только физиологические потребности оказываются удовлетворены, приоритетом, согласно пирамиде Маслоу, становится безопасность (физическая, финансовая, личная). Таким образом, безопасность — это уровень, который в значительной степени включает в себя личное здоровье и благополучие, а медицина — одна из тех областей, для которых наука о данных особенно важна. В медицинской промышленности наука о данных радикально меняет инструменты для диагностики и лечения болезней. Все медицинские эксперименты проводятся с опорой на данные участников, и эти собранные данные могут использоваться для уточнения диагноза, подбора разных практических подходов и создания новых продуктов. Чтобы выявить сложные и редкие заболевания, практикующие медики должны владеть информацией о различных их проявлениях и симптомах — это поможет избежать ошибки при постановке диагноза, найти корень проблемы и эффективно ее решать. Когда недуг усугубляется и требует безотлагательного вмешательства врачей, течение болезни может не контролироваться на протяжении недель и месяцев, которые уходят на то, чтобы пациенты записались на прием к нужному специалисту.

От ученых — аналитиков данных требуется разработать передовые алгоритмы и обучить им машины для получения наиболее точных данных. На основе этих данных могут быть

спрогнозированы необычные ситуации. Более того, собранные данные не зависят от благополучия научного сотрудника, работающего с ними (извините). Как только специалисты-медики выходят на пенсию, вместе с ними уходят их специфические знания. Когда аналитики данных уходят на заслуженный отдых, алгоритмы, которые они оставили, или собранные ими данные могут использоваться как основа для расширения существующих знаний. Наука о данных всегда опирается на то, что осталось, на информацию о нашем прошлом.

Именно эта способность позволяет столь эффективно использовать плоды науки о данных в медицине: пока данные сохраняются, накопленные знания не будут зависеть от отдельных людей.

### **Кейс: диагностика с помощью SkinVision**

На рынке существует множество цифровых приложений, которые собирают данные по различным темам, от звезд в ночном небе до веснушек на вашей коже.

SkinVision — это приложение для мобильных устройств, помогающее тестировать родинки пользователей, чтобы выявить рак кожи. Используя агрегированные пользовательские данные, алгоритм SkinVision может определить вероятность появления у пользователя родинки с злокачественными симптомами. Это действительно очень просто: с помощью приложения вы делаете фото вашей кожи, SkinVision его регистрирует и проанализирует — а потом вы получите рекомендацию относительно следующих шагов, которые вы можете предпринять вместе с врачом.

Не стоит думать, что ставить диагноз с помощью мобильного устройства легкомысленно. По мере того как будет собрано все больше и больше сведений о болезни, базы данных о ее причинах и последствиях увеличатся и станут определять диагноз намного лучше, чем это делает опытный хирург. Чем больше людей используют цифровое приложение подобное SkinVision, чтобы узнать свой диагноз, тем выше вероятность, что технология сможет отличить доброкачественную родинку от злокачественной, потому что у нее будет большой массив данных, с помощью которых можно перекрестно изучить пользовательские данные — представленные изображения. Подумайте, что бы вы предпочли: получить диагноз от человека, которому довелось рассмотреть 1000 отдельных случаев, или от машины, которая накопила информацию о миллионе отдельных случаев?

### **Объем знаний**

Отнюдь не только цифровые приложения прокладывают путь медицине, основанной на данных. Суперкомпьютер IBM Watson, по словам разработчиков, — это «когнитивная технология, которая может мыслить как человек» (IBM, 2017b). Watson прославился, когда стал первым искусственным интеллектом, победившим человека в игре Jeopardy! Но на самом деле это просто пища для СМИ\*. Что же делает Watson столь привлекательным для нас? Эта технология позволяет применять данные в здравоохранении. Watson полезен прежде всего тем, что помогает врачам выявлять болезни пациентов.

\*Что, кстати, является еще одним примером того, как данные меняют наш способ потребления информации. Самые читаемые новостные онлайн-статьи будут выталкиваться на вершину кучи, что делает это войной за самый интригующий заголовок, а не за самый убедительный контент — прим. автора.

Watson применяет тот же принцип, что и приложение SkinVision: собранные данные служат для диагностики — только для этого, естественно, требуются более изощренные алгоритмы. В одном удивительном случае Watson смог диагностировать редкий тип лейкемии у женщины всего за десять минут, в то время как у врачей это заняло бы несколько недель (Otake, 2016).

Все еще сомневаетесь относительно перспективы использования ИИ в медицине?

Разумеется, Watson не является решением всех наших проблем. Искусственный интеллект машин все еще может ошибаться. Но разница между машинами-врачами и людьми-медиками — это данные, и, по мере того как технология обработки растущих объемов информации совершенствуется, меняется и разница между человеком и машиной. В конце концов, люди могут поглощать информацию на конференциях, из медицинских журналов и статей, но все мы имеем ограниченную способность хранить знания. Более того, знания, которыми обладают люди-врачи, в значительной степени зависят от их жизненного опыта. В то же время врач-машина может совершенствоваться, только получая все больше данных. Благодаря мгновенному доступу к данным с других компьютеров через облако общие данные могут способствовать постановке более точных диагнозов и выполнению операций по всему миру. Благодаря экспоненциальному росту эти машины будут хранить информацию о всех видах изменений в человеческом теле, оставляя знания людей далеко позади.

**Подробнее читайте:**

*Еременко, К.* Работа с данными в любой сфере: Как выйти на новый уровень, используя аналитику / Кирилл Еременко ; Пер. с англ. [Д. Шалаевой] — М.: Альпина Паблишер, 2019. — 303 с.