МИРОВОЙ БЕСТСЕЛЛЕР

Эрик Сигель

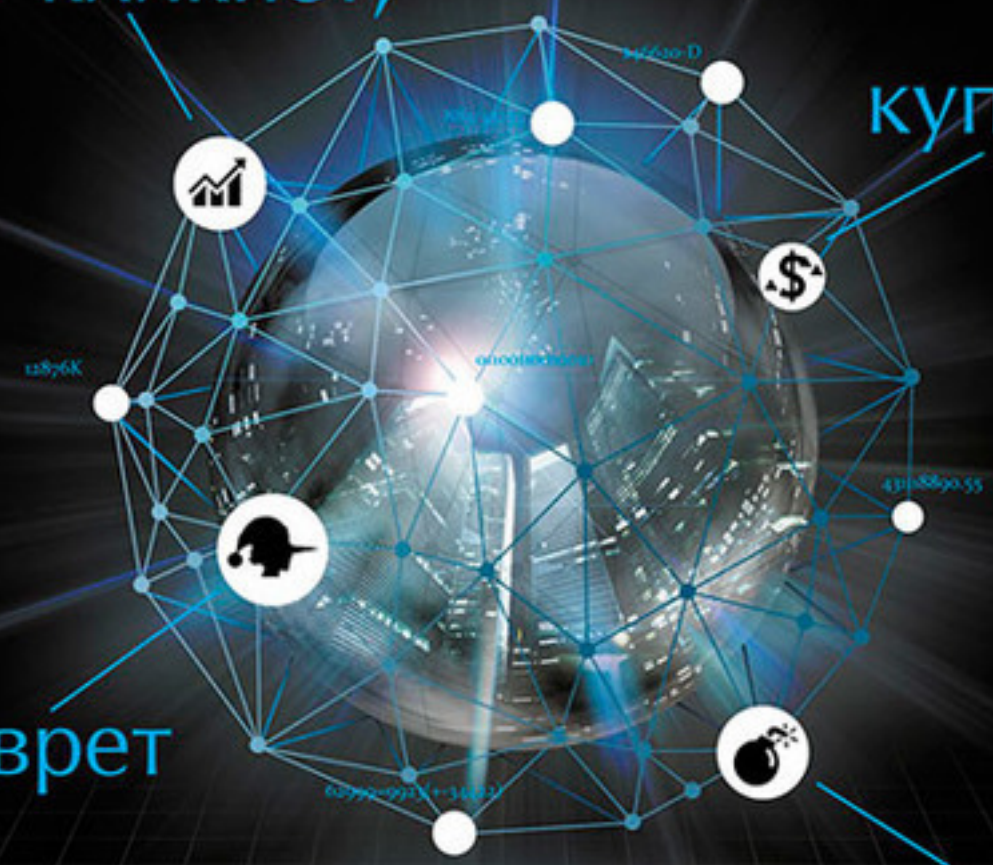
ПРОСЧИТАТЬ БУДУЩЕЕ

Кто кликнет,

купит,

соврет

или умрет



альпина паблишер

Эрик Сигель

Просчитать будущее. Кто кликнет, купит, соврёт или умрёт

«Альпина Диджитал»

2013

Сигель Э.

Просчитать будущее. Кто кликнет, купит, соврёт или умрёт / Э. Сигель — «Альпина Диджитал», 2013

Прогнозная аналитика — 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. Это 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. Эрик Сигель приводит массу примеров эффективного использования этой методики и показывает, как на практике начать использовать возможности больших данных для получения конкурентных преимуществ. Книга, ставшая мировым бестселлером и получившая множество восторженных отзывов от ведущих аналитиков и прогнозистов, будет интересна не только руководителям и менеджерам компаний, но и всем, кто хочет научиться с пользой для себя анализировать большие массивы данных.

© Сигель Э., 2013
© Альпина Диджитал, 2013

Содержание

Предисловие 7
Вступление 9
Введение 12
Прогнозирование в большом бизнесе – судьба активов 14
Позвольте вам представить… ясновидящий компьютер 15
«Накормите меня!» – пища для размышлений для компьютеров 16
Я знал, что вы это сделаете 18
Потребление 18
Любовь, работа, дети и разводы 18
Мышление и принятие решений 19
Уход сотрудников и клиентов 19
Проблемы 20
Болезни и смерть 20
Ложь, мошенничество, кражи и убийства 21
Пределы и возможности прогнозирования 23
Поле чудес 25
Организационное обучение 27
Новая модная профессия: аналитик данных 28
Искусство обучения 29
Глава 1 31
Конец ознакомительного фрагмента. 33

Эрик Сигель
Просчитать будущее. Кто кликнет, купит, соврёт или умрёт

Переводчик *И. Евстигнеева*
Научный редактор *М. Оверченко*
Руководитель проекта *О. Равданис*
Корректоры *Е. Чудинова, Е. Аксёнова*
Компьютерная верстка *К. Свищёв*
Дизайн обложки *М. Борисов*
Арт-директор *С. Тимонов*
В оформлении обложки использовано изображение из фотобанка shutterstock.com

© Eric Siegel, 2013. Все права защищены
Настоящее издание опубликовано по лицензии
© Издание на русском языке, перевод, оформление. ООО «Альпина Паблишер», 2014

Все права защищены. Никакая часть электронной версии этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами, включая размещение в сети Интернет и в корпоративных сетях, для частного и публичного использования без письменного разрешения владельца авторских прав.

* * *

«Просчитать будущее» предлагает массу замечательных примеров того, как организации в разных сферах деятельности извлекают ценные практические знания из данных. И новичок, и эксперт найдут эту книгу интересной и узнают что-то новое.
Крис Пулиот, директор по аналитике и разработке алгоритмов в Netflix

Четкое и убедительное объяснение могущества прогнозной аналитики и того, как она может трансформировать компании и даже целые отрасли.
Энтони Голдблум, основатель и генеральный директор Kaggle.com

Обязательная к прочтению, эта книга открывает нам глаза на технологии прогнозного моделирования, предсказывающие наше поведение и ежедневно влияющие на наши жизни.
Чжоу Ю, аналитик сервиса Online-to-Store в компании Google

Наконец-то появилась полноценная книга об этой отрасли. Доктору Сигелю удалось сделать то, чего до него никто даже не пытался сделать, – написать доступную и увлекательную книгу о прогнозной аналитике, которую должны прочитать все, кого интересует связанный с ней потенциал – и риски.
Марк Берри, вице-президент People Insights, подразделения ConAgra Foods

Эрик Сигель предлагает нам глубокое понимание этого нового мира больших данных, машинного обучения и интеллектуального анализа данных.
Марк Перриш, вице-президент по работе с подписчиками в Barnes & Noble

Захватывающе и увлекательно – читается как триллер! Прогнозная аналитика все больше проникает в повседневную жизнь людей и незаметно оказывает влияние на то, что мы делаем. Эрик Сигель рассказывает не только о возможностях, но и об угрозах, которые прогнозная аналитика несет с собой в реальный мир.
Марианна Дизик, статистик в Google

Подробный рассказ о том, как можно справиться с непредсказуемостью этого мира. Эрик четко объясняет, почему одни варианты действий более прибыльны, чем другие, – и я полностью с ним согласен!
Деннис Мортенсен, генеральный директор Visual Revenue, бывший директор по анализу данных в Yahoo!

Эта книга посвящается, с огромной любовью, моей матери, Лайзе Шамберг, и моему отцу, Эндрю Сигелю

Предисловие

Эта книга рассказывает о количественных методах прогнозирования человеческого поведения. Первые попытки в этом направлении были предприняты еще во время Второй мировой войны. В 1940 году отец кибернетики Норберт Винер занялся предсказанием поведения немецких летчиков с вполне конкретной целью – сбивать в небе их самолеты. Его метод использовал в качестве входного параметра наблюдаемую траекторию движения самолета, учитывал наиболее вероятные маневры уклонения и выдавал прогноз, где скорее всего окажется самолет, чтобы его можно было поразить выпущенным снарядом. К сожалению, Винер мог предсказать траекторию движения самолета только на одну секунду вперед, тогда как, чтобы сбить его, требовалось предугадать его положение в небе через 20 секунд.

Из книги Эрика Сигеля вы узнаете о множестве куда более успешных попыток подобного прогнозирования. Со времен Винера несравнимо выросла как производительность компьютеров, так и массивы накопленных данных. В результате банки, розничные торговцы, организаторы политических кампаний, медицинские учреждения и многие другие организации научились весьма успешно прогнозировать поведение конкретных людей. Цель этих усилий – привлечение новых клиентов, победа на выборах, борьба с заболеваниями и т. д.

По моему убеждению – которое, судя по всему, разделяет и Сигель, – эта интеллектуальная деятельность в целом полезна для человечества. В таких сферах, как здравоохранение, борьба с преступностью и терроризмом, она позволяет спасти жизни. Использование прогнозной аналитики в рекламе повышает эффективность последней, позволяет экономить время и внимание ее получателей и даже способствует защите окружающей среды, сохраняя деревья благодаря уменьшению объемов почтовых рассылок и издаваемых каталогов. В политике успех также, кажется, сопутствует кандидатам, следующим этому научному подходу (хотя некоторые могут не согласиться, я вижу в этом определенный положительный момент).

Однако, как отмечает Сигель в самом начале своей замечательной книги, эти методы могут быть использованы и с неблаговидными целями. «С большей силой приходит бóльшая ответственность», – цитирует он Человека-паука, подразумевая, что как общество мы должны быть осторожны при использовании таких прогнозных моделей – или придется ограничивать их применение, а значит, и выгоды от них. Как и другие мощные технологии или разрушительные инновации, прогнозная аналитика по своей сути находится вне морали и может быть использована как во благо, так и во зло. Но, чтобы избежать последнего, вам необходимо понимать, на что в принципе способен этот научный подход, и вы узнаете об этом, если продолжите читать книгу.

Прогнозная, или предиктивная, аналитика – не единственный вид аналитики, но, безусловно, наиболее интересный и важный из всех. Не думаю, что нам нужны новые книги, посвященные чисто описательной аналитике, которая показывает прошлое, но не дает понимания того, почему это произошло. В своих работах я также часто ссылаюсь на третий тип – «предписывающую» аналитику, которая объясняет, что делать, с помощью контролируемого эксперимента или оптимизации. Однако эти количественные методы намного менее популярны, чем прогнозный подход.

Книга и лежащие в ее основе идеи служат хорошим противовесом теории Нассима Николаса Талеба. В своих книгах, включая знаменитый труд «Черный лебедь», Талеб утверждает, что многие попытки прогнозирования обречены на неудачу вследствие случайной и непредсказуемой природы сложных событий. Без сомнения, он прав в том, что некоторые события действительно относятся к разряду «черных лебедей» и не поддаются прогнозированию, но дело в том, что во многих случаях человеческое поведение вполне стандартно и предсказуемо.

Многочисленные примеры успешного прогнозирования, приводимые Сигелем, напоминают нам о том, что большинство лебедей – белые.

Сигель не входит в число приверженцев идеи «больших данных». Разумеется, некоторые из приведенных им примеров попадают в эту категорию, описывая ситуации с наличием слишком большого или неструктурированного объема данных, которым невозможно легко управлять при помощи обычных реляционных баз данных. Но качество прогнозной аналитики зависит не от относительного размера имеющихся у вас данных, а от того, что вы с ними делаете. Я обнаружил, что зачастую «чем больше данных, тем ничтожнее результаты», и многие приверженцы больших данных довольствуются их использованием для создания какого-нибудь визуально привлекательного аналитического продукта. Но это далеко не так ценно, как создание прогнозной модели.

Из-под пера Сигеля вышла книга одновременно сложная и доступная для понимания даже неискушенного читателя. В ней вы найдете множество увлекательных историй, красочных иллюстраций и занимательное повествование. Я бы рекомендовал ее к прочтению даже далеким от этой темы людям, поскольку вряд ли можно сомневаться в том, что на протяжении жизни их поведение подвергалось и еще не раз будет подвергаться внимательному анализу и прогнозированию. Кроме того, в своей профессиональной деятельности большинство людей будет все чаще сталкиваться с использованием прогнозных моделей, поэтому будет нелишним научиться их учитывать и оценивать и действовать в соответствии с ними.

Короче говоря, мы живем в «прогнозируемом» обществе. И лучший способ преуспеть в нем – понимать цели, методы и ограничения предиктивных моделей. А лучший способ понять их – прочитать эту книгу.

Томас Дэвенпорт,
профессор Гарвардской школы бизнеса, заслуженный профессор Колледжа Бабсон, один из основателей Международного института аналитики, соавтор книги «Аналитика как конкурентное преимущество» и нескольких других книг из этой области

Вступление

Прошлое – забыто. Будущее – закрыто. Настоящее – даровано, потому его и зовут настоящим.
Приписывается Алану Александру Милну, Биллу Кину и мастеру Угвею, мудрой черепахе из мультипликационного фильма «Кунг-фу панда»

Когда я говорю людям, чем занимаюсь, они порой удивленно пожимают плечами. Что ж, таковы издержки моей профессиональной деятельности.

Информационный век страдает от одного вопиющего пробела в информации. Это заявление может удивить многих, учитывая, сколь скрупулезно мы записываем «все, что происходит в этом мире». Не довольствуясь тщательным документированием всех важных исторических событий, мы разработали системы, которые регистрируют каждый клик, платеж, звонок, аварию, преступление или случай заболевания. Кажется, любители данных должны быть удовлетворены, если не избалованы до крайности таким объемом информации.

Но эта кажущаяся бесконечность информации исключает события, знания о которых наиболее ценны для нас, – а именно *события, которые еще не произошли*.

Способность предугадать будущее – заветная мечта любого человека, и мы, каждый в отдельности и как общество в целом, буквально одержимы предсказаниями. Мы поклоняемся божествам ясновидения. Мы охотно платим хиромантам, внимательно изучаем гороскопы, обожаем астрологию и печенья с предсказаниями.

К сожалению, многие из тех, кто слепо верит экстрасенсам и иже с ними, напрочь отвергают науку. Как правило, они считают научный подход либо слишком трудным для понимания, либо слишком скучным. Или, быть может, они уверены в том, что предсказания по своей природе невозможны без содействия сверхъестественных сил.

В связи с этим мне нравится приводить в пример один телесериал – комедийный детектив «Ясновидец» (Psych), где главный герой – молодой консультант полиции, современная реинкарнация Шерлока Холмса – обладает настолько развитой наблюдательностью и неординарными дедуктивными способностями, что ему удается убедить полицейских в том, что он ясновидящий. Используя свои уникальные навыки, он помогает полиции расследовать самые сложные уголовные дела, при этом делая вид, что руководствуется не реальными уликами, которые находит, а неким наитием и голосами из потустороннего мира.

Я испытываю аналогичный порыв, когда у меня спрашивают мнение по поводу астрологических знаков. Но, как правило, я отделываюсь шуткой, заявляя, что «я – Скорпион, а Скорпионы не верят в астрологию».

Самый распространенный вопрос, который мне задают на вечеринках, – чем я зарабатываю на жизнь. И, честно говоря, я уже готов встретить удивленный взгляд, как только произнесу два слова – *прогнозная аналитика*. Большинство людей могут позволить себе роскошь описать свою работу одним словом: врач, юрист, официант, бухгалтер или актер. Но мое заявление всякий раз выливается в пространную беседу. Любая попытка быть краткими терпит провал:

Я консультант в области одной из бизнес-технологий. Как правило, за этим следует вопрос: «Какой именно из технологий?»

Я разрабатываю компьютерные модели, позволяющие предсказать, что люди будут делать в тех или иных ситуациях. Недоумение вкупе с полным недоверием и отчасти страхом.

Я учу компьютеры анализировать данные, чтобы прогнозировать индивидуальное человеческое поведение. Недоумение плюс нежелание продолжать разговор о том, что именно это за данные.

Я анализирую данные с целью поиска закономерностей. Еще более непонимающие взгляды, неловкие паузы, перемежающиеся с вопросами невпопад.

Я помогаю маркетологам определить, какие клиенты будут покупать, а какие нет. Кажется, на этот раз меня поняли, но такое объяснение существенно упрощает и сужает сущность того, чем я занимаюсь.

Я предсказываю поведение клиентов так же, как экспресс-тест на беременность позволяет определить, беременны вы или нет. Все делают шаг назад.

Потому я и написал эту книгу, чтобы объяснить вам суть прогнозной аналитики и продемонстрировать, почему она является интуитивным, мощным и порой внушающим благоговение инструментом.

У меня есть хорошая новость: *в деле прогнозирования многое достигается малым*. Я называю это «эффектом прогнозирования», и эта тема красной нитью проходит через всю книгу. Ценность прогнозов очевидна – разумеется, если они не сводятся к простым догадкам. От нас не требуется делать невозможное и заниматься сомнительным ясновидением. Суть прогнозной аналитики не менее захватывающая, но куда более внушающая доверие: умение даже совсем немного приподнять завесу тумана над завтрашним днем приносит щедрое вознаграждение. Другими словами, прогнозная аналитика помогает нам противодействовать финансовым рискам, укрепить сферу здравоохранения, избавиться от спама, эффективнее бороться с преступностью и увеличить продажи.

Кто вы по духу – ученый или бизнесмен? Вас воодушевляет сама идея прогнозирования или та польза, которую это умение может принести миру?

Что касается лично меня, то меня поразила сама возможность *познать непознаваемое*. Прогнозная аналитика, кажется, бросает вызов фундаментальному закону природы, согласно которому вы не можете увидеть будущее, потому что оно еще не настало. Однако мы нашли обходной путь, создав машины, способные учиться на прошлом опыте. Эта регламентированная дисциплина использует то, что мы *уже знаем* – в виде данных, – для составления все более точных сценариев того, что произойдет дальше. Соединяя передовые математические подходы и технологии, мы, ученые, проводим бесчисленные тесты и вносим корректировки, пока в конце концов не получаем систему, способную проникнуть через ранее считавшийся непроницаемым барьер между сегодняшним и завтрашним днем. Мы смело шагаем туда, где раньше не ступала нога человека!

Кто-то занимается продажами, кто-то политикой. Я занимаюсь предсказаниями, и это потрясающе.

prediction
data
predictive analytics
computer
decisions
things
predictive
technology
business
People
organizations
insurance
marketing
companies
science
based
like
machine learning
work
less
make
better
may
likely
within
knee
risk
future
possible
even
Researchers
book
computers
take
order
company
credit
well
human
best
power
across
others
customers
detect
spam
put
name
used
first
just
customer
organization
knowledge
accuracy
learning
mortgage
experience
drive
Effects
value
machine
something
system
surgery
respond
researchers
learns
turn
death
developed
often
makes
drives
accordingly
life
hand
card
Effect
behavior
fraudulent
government
derived
process
team
general
time
claims
literally
answer
next
examples
see
common
predicting
million
fact
number
mind
brain
another
job
crime
year
medical
completely
mail
way
learn
new
hospital
times
individual
bad
called
every
think
yet
target
Learning
learned
methods
following
humanity
call
demand
many
years
public
never
debtors
thing
topic
health
Data
case
annually
estimated
growing
person
email
people
day
healthcare
benefit
know
almost

Введение
Эффект прогнозирования

Как и любому другому человеку, порой мне сопутствует удача, порой нет. Черные и белые полосы в моей жизни чередуются точно так же, как и в вашей. Но мне всегда хочется узнать – не могло ли все произойти иначе? Для начала позвольте мне рассказать вам шесть коротких историй.

1. В 2009 году я почти уничтожил свое правое колено, катаясь на горных лыжах в Юте. Прыгнул я хорошо, а вот приземлился плохо. Для **операции** на колене мне нужно было выбрать место, откуда будет взята ткань для восстановления моей порванной передней крестообразной связки (главной связки коленного сустава). Выбор был сложным, поскольку от него могло зависеть, как я буду жить дальше – с хорошим коленным суставом или плохим. Я выбрал мышцы задней поверхности бедра. *Не могли ли врачи в больнице сделать лучший с медицинской точки зрения выбор?*

2. Несмотря на все мои страдания, дороже всего это обошлось моей **страховой** компании – операции на колене стоят дорого. *Могла ли компания лучше предвидеть этот риск, заключая договор страхования с сумасшедшим любителем горных лыж, и назначить соответствующую страховую премию?*

3. В далеком 1995 году я стал жертвой еще одного инцидента, хотя тот и причинил мне гораздо меньше страданий. У меня **украли персональные данные**, что в итоге стоило мне десятков часов, потраченных на бюрократические формальности и оформление различных документов, чтобы восстановить мой испорченный кредитный рейтинг. *Не могли ли кредиторы предотвратить эти неприятности, каким-либо образом определив, что заявки на новые кредиты на мое имя поступают от мошенников?*

4. Очистив свое доброе имя, я взял **ипотечный кредит** для покупки квартиры. Было ли это хорошим решением или же *мой финансовый консультант должен был предупредить меня, что в скором времени размер моей задолженности может превысить стоимость приобретенной недвижимости?*

5. Моя профессиональная жизнь также не отличается стабильностью. Хотя сам **бизнес** процветает, моя компания всегда сталкивается с риском изменения экономических условий и растущей конкуренцией. *Можем ли мы защитить рентабельность, спрогнозировав, какие маркетинговые мероприятия и другие инвестиции принесут отдачу, а какие будут пустой тратой денег?*

6. Каждый день я, как и вы, переживаю множество мелких событий, которые определяют мою жизнь. Хороший **спам**-фильтр оказывает существенное влияние почти на каждый рабочий час. Мы сильно зависим от эффективного **поиска** в Интернете при решении вопросов, связанных с работой, здоровьем (например, чтобы разобраться со спецификой операций на коленном суставе), домашним ремонтом и т. д. Мы полагаемся на персонализированные **рекомендации музыки и фильмов** от Pandora и Netflix. Мне интересно, почему после стольких лет все эти компании не изучили меня достаточно хорошо, чтобы посылать мне меньше **рекламной макулатуры** (и таким образом спасти от вырубки немало деревьев)?

Все эти ситуации важны. От них может зависеть, будет ли успешным или неудачным ваш день, год или вся жизнь. Но что между ними общего?

Все эти проблемы – и многие другие, им подобные, – лучше всего решаются через *прогнозирование*. Будет ли результат хирургической операции положительным для пациента? Не может ли обращающийся за кредитом человек оказаться мошенником? Не может ли ипотечный кредит стать непосильным бременем для домовладельца? Отреагирует ли клиент на

отправленный ему рекламный буклет? Через прогнозирование подобных вещей можно существенно улучшить качество медицинского обслуживания, снизить финансовые риски, свести на нет спам, повысить эффективность борьбы с преступностью и сократить расходы.

Прогнозирование в большом бизнесе – судьба активов

Есть и другой аспект. Помимо очевидной выгоды для нас с вами как для потребителей прогнозирование полезно и для организаций, давая им в руки совершенно новый вид оружия в конкурентной борьбе. Компании набрасываются на прогнозы, как коршуны, – в положительном смысле.

В середине 1990-х годов ученый-предприниматель по имени Дэн Стейнберг пришел в штаб-квартиру крупнейшего американского банка Chase, чтобы предложить свою систему прогнозирования для управления миллионами ипотечных кредитов. Этот банковский колосс поверил в технологию прогнозирования, предложенную Стейнбергом, и взял ее за основу для принятия решений, касающихся операций со своим гигантским ипотечным портфелем. Что написал этот парень в своем резюме?

Прогнозирование – это сила. Крупный бизнес обеспечивает себе непоколебимое конкурентное преимущество, прогнозируя будущую судьбу и стоимость отдельных активов. В вышеупомянутом случае это означало, что, принимая решения по ипотечным кредитам на основе прогнозирования будущей платежеспособности домовладельцев, Chase уменьшал риск и получал дополнительную прибыль, которой в противном случае у него не было бы.

Позвольте вам представить… ясновидящий компьютер

Стремительно развиваясь и проникая во все новые сферы, сегодня предиктивные технологии стали обыденным явлением и затрагивают каждого из нас каждый день. Они влияют на ваш опыт незаметным образом – когда вы ведете машину, покупаете, учитесь, голосуете, обращаетесь к врачу, общаетесь, смотрите телевизор, зарабатываете деньги, берете в долг или даже крадете.

Эта книга рассказывает о наиболее значимых достижениях в области компьютерного прогнозирования и о двух силах, стоящих за ними: об увлеченных людях и увлекательнейшей науке, двигающих этот прогресс.

Прогнозирование – в высшей степени непростая задача. Каждый прогноз зависит от множества факторов: широкого разнообразия известных параметров, характеризующих каждого пациента, домовладельца или сообщение по электронной почте, которое может оказаться спамом. Как справиться с этой сложнейшей задачей – соединить вместе все части головоломки для составления конкретного прогноза?

Идея проста, хотя воплотить ее на практике не так просто. Проблема решается путем применения систематического научного подхода для развития и постоянного совершенствования наших умений в области прогнозирования. Другими словами, нам необходимо в буквальном смысле *научиться* прогнозировать.

Решение кроется в *машинном обучении* – компьютеры автоматически приобретают новые знания и способности, жадно поглощая самый ценный и самый мощный *неприродный* ресурс современного общества: данные.

«Накормите меня!» – пища для размышлений для компьютеров

Данные – это новая нефть.
Меглена Кунева, еврокомиссар по защите прав потребителей

Единственным источником знаний является опыт.
Альберт Эйнштейн

Богу мы верим, все остальные должны предоставлять данные.
Уильям Эдвардс Деминг (американский ученый, известный своей теорией управления качеством)

Большинство людей не испытывают никакого интереса к данным. Что может быть скучнее, чем эти бесконечные массивы сухих цифр и фактов, порой столь банальных, как пост в Twitter типа «Я купил себе новые кроссовки!». Это бесполезный побочный продукт, который в огромных количествах образуется в процессе ведения любого бизнеса.

Вы ошибаетесь! Правда в том, что данные представляют собой бесценное собрание опыта, на котором можно учиться. Каждая медицинская процедура, кредитная заявка, публикация на Facebook, рекомендация фильма, акт мошенничества, спам-сообщение или покупка – каждый положительный или отрицательный результат, каждая успешная или неудачная попытка продажи, каждый инцидент, событие и транзакция – кодируются как данные и сохраняется в базе данных. По оценкам, объемы данных увеличиваются на 2,5 квинтиллиона байтов в день (это единица с 18 нулями). Вот где произошел настоящий Большой взрыв, породив безграничные потоки сырых, необработанных данных, с которыми могут справиться только компьютеры. При правильном использовании компьютеры жадно поглощают этот океан данных – и учатся на них.

Иногда погоня за данными превращается в настоящую золотую лихорадку. Но данные – это не золото. Повторяю, необработанные данные – это сырье. Золото – то, что можно из них добыть.

Процесс машинного обучения на основе данных раскрывает всю мощь этого все возрастающего ресурса. Он позволяет выявить, что движет людьми и их поступками, что цепляет нас за душу и как устроен мир. Получение таких знаний и делает прогнозирование возможным.

Например, благодаря машинному обучению мы получили такие ценные сведения, как[1]:

• ранний выход на пенсию уменьшает ожидаемую продолжительность жизни;

• люди, которых на сайтах знакомств чаще отмечают как привлекательных, вызывают *меньше* интереса;

• большинство фанатов Рианны по своим политическим убеждениям – демократы;

• вегетарианцы реже пропускают авиарейсы;

• количество преступлений на местном уровне увеличивается после публичных спортивных мероприятий.

Машинное обучение опирается на подобные знания, чтобы совершенствовать прогнозные возможности систем через процесс обработки больших объемов данных по методу проб и ошибок, уходящий корнями в статистику и компьютерную науку.

[1] Более подробно об этих примерах читайте в главе 3.

Я знал, что вы это сделаете

Располагая такими возможностями, что мы хотели бы спрогнозировать? 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*. Давайте рассмотрим некоторые примеры[2].

Потребление

• Голливудские киностудии, принимая решение о производстве фильмов, прогнозируют их успех.

• Американский сервис Netflix заплатил $1 млн группе ученых, которым удалось лучше других усовершенствовать способность его системы рекомендаций прогнозировать, какие фильмы должны понравиться его пользователям.

• Австралийская энергетическая компания Energex прогнозирует спрос на электроэнергию для принятия решений о том, где строить собственные электросети, а компания Con Edison – возможные сбои системы в случае повышения уровня энергопотребления.

• Уолл-стрит прогнозирует цены акций, наблюдая за их движением под влиянием динамики спроса. Такие фирмы, как AlphaGenius и Derwent Capital, управляют торговыми операциями своих хедж-фондов, отслеживая тренды и настроения широкой общественности через посты на Twitter.

• Компании – от гиганта U. S. Bank до небольших фирм, таких как Harbor Sweets (производитель сладостей) и Vermont Country Store («классические продукты высокого качества, которые трудно найти»), – прогнозируют, какие клиенты будут покупать их продукцию, чтобы нацелить на них свои маркетинговые усилия. Эти прогнозы диктуют распределение драгоценных маркетинговых бюджетов. Некоторые компании в буквальном смысле прогнозируют даже то, как наилучшим образом повлиять на вас, чтобы заставить покупать еще больше (эта тема рассматривается в главе 7).

• Прогнозирование определяет и то, какие купоны вы получаете в кассе супермаркетов. Британский розничный гигант Tesco, третья по величине сеть продуктовых магазинов в мире, прогнозирует эффективность этого целевого маркетинга и ежегодно распределяет более 100 млн персонализированных скидочных купонов в 13 странах мира. Благодаря прогнозированию уровень использования купонов вырос в 3,6 раза по сравнению с предыдущими программами. Аналогичным образом поступают Kmart, Kroger, Ralph's, Safeway, Stop & Shop, Target и Winn-Dixie.

• Прогнозирование вероятности кликов приносит щедрое вознаграждение. Поскольку веб-сайтам часто платят за каждый клик по размещенным на них рекламным объявлениям, то им важно спрогнозировать, какие объявления скорее всего привлекут ваше внимание, чтобы немедленно показать их вам. Использование прогнозирования для правильного подбора рекламы обеспечивает многомиллионные доходы.

Любовь, работа, дети и разводы

• Ведущая социальная сеть для делового общения LinkedIn прогнозирует ваши профессиональные навыки.

---

[2] Больше примеров и дополнительных деталей вы найдете в таблицах в приложении D.

• Популярные сайты знакомств Match.com, OkCupid и eHarmony прогнозируют, какая из красоток на вашем экране подходит вам больше всего.

• Ретейлер Target прогнозирует, кто из покупателей ждет ребенка, чтобы рекламировать соответствующие товары. Ничто не предсказывает будущие потребительские нужды лучше, чем ожидаемое рождение нового потребителя.

• Клинические исследователи прогнозируют вероятность супружеской неверности и развода. Существуют даже сайты, на которых вы можете самостоятельно оценить ваши шансы на долгий и прочный брак (например, www.divorce360.com). По слухам, компании, эмитирующие кредитные карты, делают то же самое.

Мышление и принятие решений

• В 2012 году Обама был переизбран президентом благодаря прогнозированию поведения избирателей. Его предвыборная кампания строилась на прогнозировании того, какие избиратели положительно отреагируют на контакт (через телефонный звонок, личный визит, листовку или телевизионную рекламу), а каких такого рода контакт непроизвольно подтолкнет проголосовать против. Эта технология была использована для принятия решений о методах предвыборной агитации для миллионов избирателей в колеблющихся штатах и помогла успешно склонить на сторону Обамы намного больше избирателей, чем позволило бы сделать традиционное таргетирование.

• «Что вы имеете в виду?» Системы научились определять намерение, стоящее за письменным словом. Citibank и PayPal выявляют отношение клиентов к их продуктам, а разработанная одним исследователем программа может сказать, какие отзывы о книгах на Amazon.com носят саркастический характер.

• Разработаны компьютерные программы по автоматическому оцениванию школьных сочинений, которые также работают на основе прогнозирования. Они оценивают сочинения так же точно, как и преподаватели.

• Созданный компанией IBM суперкомпьютер Watson принял участие в популярном в США интеллектуальном телешоу Jeopardy! (российский аналог – «Своя игра») и победил двух самых именитых чемпионов за всю его историю. Опираясь на технологию прогнозирования, эта машина умеет понимать заданные в свободной форме вопросы на самые разные темы и находить на них ответы в базе данных.

• Компьютеры буквально умеют читать ваши мысли. Исследователи научили системы декодировать сканы головного мозга и определять, о какого рода вещах вы сейчас думаете – таких как определенные инструменты, здания или еда, – с точностью более 80 %. В 2011 году IBM высказала прогноз, что через пять лет технологии чтения мыслей станут преобладающим направлением исследований.

Уход сотрудников и клиентов

• Компания Hewlett-Packard оценивает каждого из своих более чем 330 000 сотрудников по всему миру по шкале «Риск ухода», показывающей вероятность того, что этот человек покинет компанию, – чтобы менеджеры могли заранее вмешаться там, где это возможно, и изменить ситуацию.

• Вы когда-нибудь были недовольны услугами сотовой связи? Ваш оператор хочет об этом знать. Все крупные сотовые операторы занимаются прогнозированием вероятности того, что вы решите уйти к конкуренту (возможно, еще до того, как у вас в голове сформируется такая мысль), используя такие показатели, как сброс вызовов, интенсивность использования

телефона, информация о счетах и уход к другим операторам абонентов из вашего списка контактов.

• FedEx стабильно сохраняет за собой позиции лидера рынка, прогнозируя с точностью от 65 до 90 %, какие клиенты могут уйти к конкурентам.

• Американская система государственных университетов (APUS) прогнозирует, какие студенты могут бросить учебное заведение, и использует эти прогнозы. Университеты штатов Алабама, Аризона, Айова, Оклахома и нидерландский Технический университет в Эйндховене также прогнозируют выбытие студентов.

• «Википедия» прогнозирует, кто из ее редакторов, работающих бесплатно из любви к своему делу и ради развития этого бесценного онлайн-актива, собирается прекратить оказывать ей добровольную помощь.

• Исследователи из Гарвардской медицинской школы прогнозируют, что, если ваши друзья бросают курить, у вас больше шансов сделать то же самое. Отказ от курения заразителен.

Проблемы

• Страховые компании прогнозируют, кто из страхователей вероятнее всего попадет в аварию или сломает ногу, катаясь на горных лыжах. Allstate прогнозирует ответственность за причинение телесных повреждений при автомобильных авариях на основе характеристик застрахованного транспортного средства, что позволяет компании экономить около $40 млн в год. Другой ведущий страховщик сообщил о том, что экономит почти $50 млн в год благодаря применению передовых предиктивных моделей в своей практике актуарных расчетов.

• Компания Ford также использует прогнозный анализ данных, так что ее автомобили могут определить, когда у водителя снижается концентрация внимания из-за каких-либо отвлекающих факторов, усталости или опьянения, и принять меры, например подать звуковой сигнал.

• Используя данные Национального совета по безопасности на транспорте (США), исследователи установили, какие происшествия на воздушном транспорте имеют в пять раз более высокую, чем в среднем, вероятность привести к фатальному исходу.

• Все крупные банки и эмитенты кредитных карт прогнозируют, какие заемщики вероятнее всего могут нарушить обязательства по возврату кредита. Коллекторские агентства строят работу с каждым проблемным должником на основании прогнозов о том, какая тактика позволит им взыскать с него максимальную сумму.

Болезни и смерть

Я не боюсь смерти, я просто не хочу при этом присутствовать.
Вуди Аллен

• В 2013 году калифорнийская организация Heritage Provider Network пообещала $3 млн тому, кто создаст лучший алгоритм для прогнозирования поступления пациентов в ее больницы. Благодаря этим прогнозам можно будет предпринимать профилактические меры, что позволит сэкономить десятки миллиардов долларов, которые ежегодно тратятся на ненужную госпитализацию. Аналогичным образом Медицинский центр при Университете Питтсбурга прогнозирует вероятность скорой повторной госпитализации, чтобы избежать преждевременной выписки пациентов из больницы.

• Исследователи Стэнфордского университета научили компьютеры диагностировать рак молочной железы точнее, чем это делают врачи. Для этого был разработан инновационный метод, учитывающий большее число факторов при анализе образцов ткани.

• Исследователи из Университета Бригама Янга и Университета штата Юта научились правильно прогнозировать примерно 80 % преждевременных родов (и примерно 80 % своевременных родов) на основе пептидных биомаркеров, которые появляются в крови уже на 24-й неделе беременности.

• Исследователи разработали метод диагностики шизофрении на основе одного только анализа устной речи.

• Все большее число компаний, занимающихся страхованием жизни, выходят за рамки обычных актуарных таблиц и используют методы прогнозной аналитики для оценки риска наступления смерти. Хотя официально это не называется *страхованием на случай смерти*, на деле компании пытаются спрогнозировать, когда вы умрете.

• Этим занимаются не только компании по страхованию жизни. Одна из ведущих медицинских страховых компаний прогнозирует вероятность того, что пожилые держатели страховых полисов перейдут в мир иной в течение ближайших полутора лет, на основании определенных клинических показателей, содержащихся в последних заявлениях о выплате страхового возмещения за предоставление медицинских услуг. Не бойтесь – это делается с благими целями.

• Исследователи разработали методики прогнозирования риска смерти при хирургическом вмешательстве на основе индивидуальных характеристик пациента и характеристик его состояния, чтобы помочь врачам принимать объективные решения.

• Врачи регулярно – хотя и непреднамеренно – жертвуют одними пациентам ради того, чтобы помочь другим, и эта сложившаяся практика почти не вызывает споров. Но ситуацию можно существенно улучшить, если выйти за рамки прогнозирования диагноза или результата и прогнозировать также *воздействие* медицинской помощи (прогнозирование воздействия – тема главы 7).

Ложь, мошенничество, кражи и убийства

• Большинство средних и крупных банков используют технологии прогнозирования, чтобы противодействовать попыткам использования поддельных чеков, краж денег с кредитных карт и другим мошенническим действиям. Благодаря внедрению специальной системы Citizens Bank удалось на 20 % уменьшить убытки, вызванные мошенничеством с чеками. Hewlett-Packard сэкономила $66 млн, научившись выявлять мошеннические претензии по гарантии.

• Компьютерные прогнозные модели позволяют определить, кто должен находиться в тюрьме. Сотрудники судебно-пенитенциарной системы в штатах Орегон и Пенсильвания используют такие программы для оценки риска повторного совершения преступления и учитывают эти прогнозы при принятии решений об условно-досрочном освобождении и вынесении приговоров.

• Считается, что убийства в целом не поддаются прогнозированию со сколь-нибудь значимой степенью точности, но в рамках определенных групп риска предиктивные методики могут быть эффективны. В штате Мэриленд используется аналитическая система, выдающая прогнозы относительно того, кто из находящихся под надзором лиц может убить, а кто может быть убит. Университетские исследователи совместно с правоохранительными органами разработали прогнозные модели, способные предсказать вероятность рецидива среди тех, кто ранее был осужден за убийство.

• Один эксперт по борьбе с мошенничеством в крупном британском банке распространил свои разработки на сферу борьбы с терроризмом, выявив небольшую группу лиц, подозреваемых в преступных связях, на основе их банковских операций.

• В Чикаго, Мемфисе и Ричмонде (штат Вирджиния) полицейские усиленно патрулируют районы, где, по прогнозам, может произойти скачок уличной преступности.

• Вдохновленные детективным телесериалом «Обмани меня», где специалисты, изучая движение тела и мимику человека с целью распознать ложь, расследовали самые сложные преступления, исследователи из Университета Буффало создали компьютерную программу, способную выявить ложь с точностью 82 % на основе одного только движения глаз.

• В конце 1990-х годов, когда я работал преподавателем в Колумбийском университете, у меня была команда ассистентов, которые использовали компьютерную программу для проверки сотен домашних заданий по программированию на предмет выявления плагиата.

• Налоговое управление США прогнозирует, с какой вероятностью вы можете его обманывать.

Пределы и возможности прогнозирования

Экономист – это специалист, который назавтра узнает, почему не произошло то, что он предсказывал вчера.
Эрл Уилсон

Почему вы никогда не увидите заголовок «Экстрасенс выиграл в лотерею»?
Джей Лено

Все из перечисленных в предыдущем разделе достижений стали возможны благодаря прогнозированию, которое в свою очередь является результатом машинного обучения. Между всеми этими разнообразными примерами и научной фантастикой есть одно ключевое различие: они не вымышлены. И эти примеры – лишь скромный срез существующей ныне реальности. Можно с уверенностью сказать, что сила прогнозирования отныне всегда пребудет с нами.

Но не является ли такое утверждение чересчур смелым? В свое время датский физик Нильс Бор сказал: «Очень трудно сделать точный прогноз, особенно о будущем». В конце концов, возможно ли прогнозирование в принципе? Будущее покрыто мраком неизвестности, и неопределенность – единственное, в чем мы можем быть уверены.

Воспроизводится с разрешения правообладателя

Позвольте же мне немного рассеять перед вами туман в этой области. В принципе, точное прогнозирование невозможно. Даже погода прогнозируется всего лишь с 50 %-ной точностью, а предсказать поведение людей, будь то пациентов, клиентов или преступников, ничуть не проще.

Но есть и хорошая новость! Прогноз 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 «не реагирующих» людей из списка рассылки. А эти люди, в свою очередь, выиграют оттого, что получат по почте меньше макулатуры.

Прогнозирование: человек, который видит рекламную брошюру сегодня, покупает завтра

Таким образом, бизнес уже давно использует эту игру с цифрами для массового маркетинга, деликатно, но весомо склоняя чашу весов на свою сторону, – и делает это *без* высокоточных прогнозов. На самом деле, чтобы прогнозирование имело практическую ценность, довольно и низкой точности. Если в среднем ответная маркетинговая реакция составляет 1 %, то в выделенной группе доля потенциальных покупателей увеличивается до 3 %. В данном случае мы не можем с уверенностью предсказать, отреагирует или нет каждый отдельно взятый адресат на рекламную рассылку. Но стоимость создается благодаря выявлению группы людей, которые – *в совокупности* – склонны вести себя определенным образом.

Это демонстрирует в общих чертах то, что я называю эффектом прогнозирования. 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.

Эффект прогнозирования: малым достигается многое.

Это первый из пяти эффектов, о которых рассказывается в этой книге. Вероятно, вы уже слышали об эффекте бабочки, эффекте Доплера и эффекте плацебо. Оставайтесь с нами, и вы узнаете также об *эффекте данных, эффекте индукции, эффекте ансамбля и эффекте воздействия*. Рассказ о каждом из них включает любопытные сведения из области науки и технологий: интуитивный взгляд, раскрывающий перед вами, как это работает и почему позволяет достигать успешных результатов.

Поле чудес

Люди… действуют на основе своих убеждений и предубеждений. Если вы сможете устранить то и другое и заменить их данными, вы получите реальное преимущество.
Майкл Льюис, из книги «Moneyball. Как математика изменила самую популярную спортивную игру в мире»

О какой области знаний или отрасли науки мы здесь говорим? Обучение тому, как прогнозировать на основе данных, иногда называют *машинным обучением* – но это преимущественно научный термин, который используется в исследовательских лабораториях, на научных конференциях и в университетах (например, в конце 1990-х я несколько раз вел курс по машинному обучению в Колумбийском университете). Хотя именно в этих академических кругах куются новые знания, это не то место, где шины соприкасаются с дорогой. Там, где машинное обучение находит реальное практическое применение – в коммерческой, промышленной и государственной сферах, – его называют иначе:

> ***Прогнозная аналитика*** (***ПА***, англ. predictive analytics) – технология, опирающаяся на опыт (данные) для прогнозирования будущего поведения людей с целью принятия оптимальных решений.

Построенная на фундаменте компьютерных наук и статистики и активно развиваемая благодаря научно-исследовательским программам, прогнозная аналитика превратилась в самостоятельную дисциплину. Но ПА шагнула далеко за пределы теоретической науки и стала мощным практическим инструментом, оказывающим непосредственное влияние на нашу 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?*

Прогнозы определяют, как организации обращаются с конкретным индивидом и обслуживают его во всех сферах, составляющих функциональное общество

Таким образом, ПА кардинально отличается от стандартного прогнозирования (которое в английском языке называется словом *forecasting*). Последнее производит совокупные прогнозные оценки на макроскопическом уровне. Как будет развиваться экономика? Какой кандидат в президенты наберет больше голосов в Огайо? В то время как совокупная прогнозная оценка скажет вам, сколько стаканчиков мороженого будет куплено в штате Небраска в следующем месяце, ПА позволит узнать, какие именно жители Небраски 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. Но как только вы вступаете в эту область, вы тут же сталкиваетесь с массой 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. Как однажды выразился Майк Лукидес, вице-президент инновационного издательства O'Reilly: «Наука о данных похожа на порнографию – когда видишь, понимаешь, что это». 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.

Организационное обучение

Ведущие компании в эпоху Интернета, в том числе Google и Amazon… имеют бизнес-модели, которые опираются на предиктивные модели, основанные на машинном обучении.
Профессор Васант Дхар из Школы бизнеса Стерна при Нью-Йоркском университете

Если рассматривать организацию как своего рода «мегачеловека», не означает ли это, что она нуждается в «мегаобучении»? Люди объединяются в группу – будь то компания, правительство, больница, университет, благотворительная организация и т. п., – чтобы служить интересам ее членов и тех, кого она обслуживает. Будучи сформированной, группа выигрывает от разделения труда, взаимодополняющих навыков и эффекта масштаба. Возможности группы как целого намного превосходят сумму возможностей отдельных ее членов. Коллективное обучение является следующим логическим шагом для организации, позволяющим еще больше увеличить ее коллективный потенциал. Точно так же, как торговый агент со временем учится на своем положительном и отрицательном опыте взаимодействия с потенциальными клиентами, на своих успехах и неудачах, ПА дает в руки организации инструмент, посредством которого она может учиться на опыте, приобретаемом ею через отдельных ее членов и компьютерные системы. На самом деле организация, которая не использует получаемые ею данные таким образом, похожа на человека с фотографической памятью, который не умеет с пользой применять этот дар.

За редкими исключениями именно организации, а не отдельные люди, максимально выигрывают от использования ПА. Организации принимают огромное количество операционных решений, но, поскольку по природе своей они неэффективны и расточительны, существует значительный простор для оптимизации и улучшений. Маркетинговые службы делают массовые почтовые рассылки, но значительная часть рекламного материала, для изготовления которого тратится немало денег и вырубается немало деревьев, попадает прямиком в мусорную корзину. По оценкам, 80 % всех сообщений по электронной почте является спамом. Рискованным заемщикам выдается слишком много кредитов. Заявления на предоставление государственных пособий накапливаются в огромных количествах и не рассматриваются в срок. И это при том, что организации располагают изобилием данных, которые могут быть использованы для прогнозирования и соответствующего улучшения операций.

В коммерческом секторе прибыль является движущей силой. Только представьте себе, какие вырисовываются многообещающие перспективы, если повседневные рутинные операции станут более эффективными, целевой маркетинг – более точным, если будет предотвращаться больше попыток мошенничества, выдаваться меньше кредитов ненадежным заемщикам и привлекаться больше онлайн-клиентов. Позволяя оптимизировать критически важные операции, ПА существенно увеличивает коллективные возможности организации и ее эффективность в целом.

Новая модная профессия: аналитик данных

Самой привлекательной профессией в ближайшие десять лет будет профессия статистика.
Хэл Вэриен, главный экономист Google и профессор Калифорнийского университета в Беркли, 2009 год

Альтернатива прогнозированию будущего – анализ прошлого… а для этого нужно всего лишь иметь хорошую память.
Шелдон Купер, физик-теоретик, главный персонаж телесериала «Теория Большого взрыва»

Но прибыль – не единственный мотиватор. Источник энергии, главная сила, движущая эту махину вперед, – это «Сила умников»! Я имею в виду специалистов-практиков и их энтузиазм. По правде говоря, моя страсть к прогнозной аналитике проистекает вовсе не из ее ценности для организаций. Я занимаюсь этим ради собственного удовольствия. Идея, что машины действительно могут учиться, завораживает меня, и гораздо больше меня интересует то, что происходит внутри черного ящика, чем полезность происходящего для внешнего мира. Возможно, именно этот движущий мотив и отличает настоящего «умника» от других людей. Мы любим технологии; мы одержимы ими. Показательный пример: ведущее программное обеспечение с открытым исходным кодом, используемое в прогнозной аналитике, название которой состоит из одной буквы R (умники любят такие странные названия), имеет быстро расширяющуюся базу пользователей и добровольцев-разработчиков, которые совершенствуют его функциональные возможности и обеспечивают поддержку. Огромное число профессионалов и любителей стекаются на публичные конкурсы в сфере ПА, для которых характерен дух не столько состязательности, сколько сотрудничества. Мы работаем в организациях или консультируем их. Мы – востребованные специалисты, поэтому много летаем. И летаем высшим классом.

Искусство обучения

Что вы делаете с вашим ЦП [центральным процессором],
Чтобы в полной мере реализовать его потенциал?
Используйте ваш мозг с умом,
Чтобы его производительность росла
в геометрической прогрессии.
Если вы хотите сделать свой компьютер умнее,
Есть один только путь —
Саморазвитие путем проб и ошибок.

Из стихотворения автора этой книги «Научитесь этому!»

Много лет назад человечество создало Совершенную универсальную машину и, следуя необъяснимому порыву преуменьшить гениальность своего творения, дало ему имя «компьютер» (первоначально в английском языке это слово означало человека, производящего вычисления вручную). Эта машина могла выполнять любой бесконечный набор сложнейших инструкций без единой ошибки и жалобы, и за несколько десятилетий ее скорость и возможности возросли настолько, что человечество могло только восклицать: «Черт возьми, неужели мы это сделали?!» Этому фантастическому устройству гораздо больше подошло бы величественное название *La Machine*, но по иронии судьбы через несколько десятков лет это имя было даровано кухонному комбайну (я не шучу). Какая жалость. «Что мы должны делать с компьютером? Каков его истинный потенциал, и как нам использовать его в полной мере?» – в изумлении спрашивало себя человечество.

Между компьютером и нашим головным мозгом есть нечто общее, что, с одной стороны, наделяет их тайной, а с другой – делает в наших глазах чем-то абсолютно естественным и само собой разумеющимся. Если, размышляя над этим предложением, вы слышали, как мимо вас пролетела муха, вы понимаете, о чем я говорю. И мозг, и компьютер работают в тишине. Их механизмы не издают ни звука. Да, у компьютеров есть дисководы и вентиляторы, которые могут издавать некоторый шум, – так же как ваша голова, в которой заключен ваш мозг, может издавать сопение, чихи и храпы, – но вся основная работа осуществляется ими без участия «движущихся частей», так что эти усилия происходят абсолютно бесшумно и незаметно. В результате вывод информации на монитор или появление гениальных идей в вашем уме могут показаться настоящим чудом[3].

И наш головной мозг, и компьютер обладают поистине дьявольской мощностью. Так можно ли запрограммировать компьютер таким образом, чтобы научить его думать и чувствовать или сделать по-настоящему разумным? Кто знает? В лучшем случае это можно рассматривать как стимулирующие философские вопросы, на которые трудно дать ответ; в худшем – сделать их мерой успеха, который вследствие своей субъективности всегда будет оставаться недосягаемым. К счастью, в одном из вопросов у нас есть полная ясность: компьютеры обладают одной поистине впечатляющей, ключевой способностью, присущей человеку, – они умеют учиться.

---

[3] Полная тишина характерна для твердотельных электронных приборов, но компьютеры необязательно должны иметь такую конструкцию. Концепция универсальной машины, выполняющей инструкции, является абстрактной и не привязана к феномену электричества. Вполне возможно создать компьютер, который состоит из шестерней, колес и рычагов и приводится в движение паром или бензином. Конечно, я бы не советовал это делать, но вы можете попытаться. Такой компьютер будет очень медленным, громоздким и громким, так что никто не станет его покупать.

Но как? Оказывается, что обучение – умение делать общие выводы на основе частных наблюдений – непростая задача. Это глубокая философская проблема. Задача машинного обучения состоит в том, чтобы находить закономерности, которые проявляются не только в имеющихся данных, но и в общем, за их пределами, чтобы результат обучения был верен и в новых ситуациях, которые ранее никогда не возникали. По сути, именно способность к выявлению таких закономерностей и обобщению является волшебной пилюлей ПА. Разработка методов машинного обучения – настоящее искусство. Чуть дальше мы более подробно остановимся на этой теме, а сейчас я дам вам одну подсказку. Компьютер узнает о том, как *вы* вероятнее всего поведете себя в той или иной конкретной ситуации, изучая *других людей*, а не *вас*.

Продолжая засыпать вас головоломками, я хочу загадать еще одну загадку: *что часто происходит с нами, чего нельзя наблюдать и в отношении чего нельзя быть уверенным в том, что это произойдет с нами снова, – но что может быть спрогнозировано заранее?* Ответ на загадку вы найдете в последней главе этой книги.

Извлечение информации из данных с целью прогнозирования – это только первый шаг. Сделать следующий шаг и начать *действовать на основе прогнозов* – вот где требуется настоящая смелость. Как вы узнаете из захватывающей истории, которую я расскажу вам в главе 1, применение прогнозной аналитики в реальной деятельности по остроте ощущений сопоставимо с запуском ракеты в космос.

Глава 1
Ракета прогнозной аналитики: три, два, один, старт!

Сколько смелости нужно для того, чтобы внедрить прогнозную аналитику в реальную практическую деятельность? И какое вознаграждение это сулит? Что происходит, когда человек ставит все сбережения на разработанную им лично предикативную модель для торговли на фондовом рынке? Внедрить прогнозную аналитику – значит не только составлять прогнозы, но и действовать на их основе, применяя на практике извлеченные из данных знания. Это решительный шаг – но вы не сможете выиграть, если не будете играть.

В середине 1990-х годов один амбициозный молодой исследователь понял, что он не может больше ждать. Посоветовавшись с женой, он поставил все семейные сбережения на торговую систему, основанную на хитроумной модели прогнозирования динамики фондового рынка, которую он разработал в свободное от основной работы время. Подобно доктору Генри Джекилу, смело выпившему намешанное им непроверенное зелье, доктор Джон Элдер решительно сказал: «Поехали!»

Всегда страшно пробовать что-то новое, а тем более испытывать новую технологию. Запуск космического корабля может быть символом технологического величия и предметом национальной гордости, но за кадром остается небольшая группа жен астронавтов, которые в этот момент испытывают сильнейший страх за своих мужей. Астронавты же, по сути, добровольно обрекают себя на роль подопытных кроликов в колоссальном по размаху и важности эксперименте; они готовы пожертвовать собой ради того, чтобы стать частью истории.

Великие задачи рождают великие достижения. Мы уже совершили прогулку по Луне, а недавно первой негосударственной организации была вручена премия в размере $10 млн за разработку пилотируемого космического корабля многоразового использования. Мы вступаем в эпоху беспилотных автомобилей с системами автоматического управления – «Мам, посмотри, я не держусь за руль!» Сжигая в своих двигателях миллионы долларов всевозможных грантов и премий, эти машины самостоятельно рулят по улицам кампусов Google и BMW.

Замените рев ракеты на ряды данных, и перед вами откроется не менее головокружительная перспектива, чем возможность проникнуть за пределы Земли в космическое пространство, – а именно возможность проникнуть за пределы настоящего и заглянуть в будущее. Преодоление границы между настоящим и будущим – занятие ничуть не менее захватывающее, чем выход в космос, но менее опасное (космос – это вакуум, а вакуум убивает). Миллионы призовых долларов выделяются на то, чтобы научиться избегать ненужной госпитализации каждого отдельно взятого пациента или точно прогнозировать уникальные предпочтения конкретного потребителя. Телевикторина Jeopardy! (российский аналог – «Своя игра») вручила приз в $1,5 млн создателям компьютера, победившего в поединке между человеком и машиной, что стало возможным благодаря значительному прогрессу в способности машины прогнозировать правильные ответы на вопросы. (Разумеется, IBM вложила гораздо больше денег для достижения этой победы, но более подробно об этом вы узнаете в главе 6.) Прибегая к помощи прогнозных аналитиков, организации удерживают детей в школах, обеспечивают приток новых клиентов и предотвращают совершение преступлений. Без преувеличения можно сказать, что благодаря прогнозной аналитике одерживаются победы на политических выборах, чемпионатах по бейсболу и… я еще не упоминал об управлении финансовым портфелем?

Биржевая торговля при помощи «черных ящиков», т. е. механических торговых систем, автоматически принимающих решения о проведении операций на рынке, – это святой Грааль принятия решений на основе данных. Черный ящик – это компьютерная программа, которая анализирует поступающие в нее потоки текущих финансовых данных и выдает приказы (или

рекомендации) по покупке, продаже или удержанию бумаг. Он называется черным, поскольку до тех пор, пока он принимает правильные решения, вас не волнует, что происходит внутри него. Будучи работоспособным, он затмевает собой любой другой бизнес, возможный в этом мире: ваш компьютер становится устройством, превращающим электричество в деньги.

Для Джона Элдера запуск собственной торговой системы требовал немалого мужества. Даже если технология прогнозирования изменений на фондовом рынке будет огромным шагом вперед для человечества, для самого Джона этот шаг был совсем не маленьким. Для этого момента можно было подобрать множество метафор. Идя ва-банк и складывая все яйца в одну аналитическую корзину, Джон принимал большую дозу изобретенного им же лекарства.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, купив полную легальную версию на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.