

Лекция 5. Большие данные в рейтинговой системе

Дилемма: «умные» технологии

Предполагаемую полезность больших данных в наши дни легко продают конечным пользователям, хотя на практике гораздо больше пользы получается для компании. Примером может служить «умный» электросчетчик.

Традиционный электросчетчик — это простой прибор, который измеряет, сколько электричества было использовано. Он установлен у вас в доме, вы сами сообщаете в электросбытовую компанию показатели счетчика или к вам каждый месяц приходит представитель компании, записывает эти показания, а потом на их основании выставляется счет. В любом случае сотрудники компании приходят к вам с проверками.

Но в настоящее время во все большем количестве домохозяйств в развитых странах устанавливают электросчетчики нового поколения. Например, в Великобритании запущена широкомасштабная программа по установке «умных» счетчиков, и планируется установить их в 26 миллионах домохозяйств к 2020 году. Продажи идут хорошо. Разработчики утверждают, что эти «умные» электросчетчики помогут нам снизить плату за электричество, потому что они показывают, сколько электроэнергии используется и сколько это стоит. Видя эту информацию, человек начнет экономить. Но дисплей может показывать вам, что стоимость в ту минуту, когда вы на него смотрите, низкая (вы же не будете к нему бегать каждый час?). Я очень сомневаюсь, что подобные дисплеи помогут существенно снизить потребление электроэнергии. Более того, нам говорят, что «умные» электросчетчики знают, как пользоваться специальными тарифами и, соответственно, поставлять гражданам более дешевую энергию в разное время дня. Они же умные!

Но разве электросбытовыми компаниям выгодно поставлять нам более дешевую энергию? Счетчики должны быть выгодны, раз компании так активно выступают за их установку. Дело в том, что после установки таких счетчиков электросбытовыми компаниям больше не потребуется нанимать людей, которые ходят по домам и записывают показания счетчиков. Это очень серьезное сокращение расходов. А меняющиеся тарифы вполне могут запутать обычного потребителя, особенно пожилого возраста, и вместо того чтобы экономить, придется платить по повышенной ставке за пользование электроэнергией, например, в пиковые часы.

Конечно, нельзя утверждать, что «умная» технология опасна для потребителей, так как на самом деле обманет доверчивых граждан. Она может очень облегчать жизнь людей. Для примера приведу «умный» термостат (умный с точки зрения потребителя), которым можно управлять со смартфона. Такой термостат определяет, когда в доме никого нет, и автоматически снижает температуру в помещении. И вы платите меньше. И вы можете послать сигнал термостату о том, что приближаетесь к дому.

«Умные» электросчетчики — это примеры использования технологии *Big Data* на практике с фактически односторонней пользой (для компаний) и сомнительной для потребителя, которому еще нужно «отбить» установку счетчика. Никто их бесплатно не ставит, а данных о реальном снижении платы за электроэнергию мало, хотя бы потому, что «умные» счетчики появились не так давно.

Большой босс

Мы смело можем утверждать, что большие данные начинают править нами в этом мире. Хотя только в США порядка 800 000 человек получают деньги благодаря

компаниям, работающим с большими данными. Но технологии *Big Data* также контролируют найм на работу и то, как люди, нанятые компаниями, выполняют свою работу. В компании *liber*, которая уже упоминалась, по всему миру работает более миллиона водителей, ими управляют большие данные. Эта компания ориентируется на потребителей, как и масса других. В магазинах в субботу работает больше людей, чем в понедельник. В ресторанах больше персонала в пятницу и субботу вечером, и в любом случае больше в 20:00, чем утром. И вообще утром многие рестораны закрыты. Они ориентируются на обеденный перерыв и ужин.

Технологии *Big Data* позволяют оптимизировать использование и расстановку персонала, часто даже поминутно! Это очень выгодно компании, но может иметь печальные последствия для сотрудников — неудобные часы работы, отсутствие стабильного графика, сокращение доходов.

Если у вас стабильный график, вы можете планировать вашу жизнь вне работы. Но если рабочий график строится с использованием технологий *Big Data*, вы должны быть готовы к выходу на работу по звонку каждый день. Например, обещают плохую погоду, а в плохую погоду в ваш ресторан приходит больше народа. Или, наоборот, приходит меньше. Выпускаемый вашей компанией товар в дождь лучше покупают или к вам в магазин в дождь не приходят вообще. Это все знают большие данные и выдают рекомендации вашему работодателю. А страдаете вы. Вы не сможете ничего планировать. Если в вашем контракте не прописана минимальная заработная плата, минимальное количество рабочих часов, то может оказаться, что в некоторые месяцы вы даже не сможете заплатить за квартиру и при этом где-то подработать, потому что вынуждены находиться в состоянии боевой готовности.

В Японии, а теперь и ряде других стран у некоторых производителей популярен так называемый *JIT*-подход, или «точно в срок». Это система оперативного управления запасами. Все ресурсы, включая персонал, материалы, оборудование используются таким образом, чтобы обеспечить непрерывный производственный процесс. Здесь предполагается минимизация времени изделия (детали, материала) в пути и минимальные запасы сырья, материалов, комплектующих, соответственно нет расходов на их хранение, огромных складских помещений. Все поступает в производство «с колес». Система упрощает учет и контроль запасов, способствует рациональному использованию производственных мощностей и сокращению затрат. То есть за хранение комплектующих и материалов платит их поставщик, а не фирма-изготовитель. Но для непрерывно-поточного производства необходима согласованная работа всех участников процесса. Такое возможно в Японии, в которой впервые применили этот подход, но, например, невозможно в странах Латинской Америки.

Такой подход очень упростился с применением технологий *Big Data*. Но компании, работающие таким образом, часто избавляются от постоянных сотрудников, чтобы не оплачивать им больничные и отпускные. Люди становятся фрилансерами — и компания приглашает их только тогда, когда есть конкретный заказ. Подвозят комплектующие, материалы, работа распределяется между людьми, работающими по вызову, а не постоянно, и этим людям платят или за отработанные часы, или за количество штук произведенного товара. Так выгодно компании. Но не людям.

Против компании *Uber* было подано несколько исков, а в Великобритании даже прошла протестная демонстрация курьеров, нанятых подразделением *UberEats*, которые развозят заказы из ресторанов. Изначально курьерам предложили очень неплохую ставку — 20 фунтов стерлингов в час, а потом алгоритм стал гораздо сложнее. За каждую доставку полагалась небольшая сумма, а потом предлагались доплаты за каждую милю, за доставку в час пик и т. д., но были и вычеты, например проценты компании за предоставление работы. В США потребовали относиться к

водителям как к постоянным работникам с соответствующими выплатами — платить фиксированную ставку за час независимо от того, поступил вызов или нет. Но решения зависят от штатов и силы организаций, отстаивающих права работников. Где-то решения принимаются в пользу компании, где-то в пользу работников. Но компания после этого вполне может уйти из этих штатов, и сотрудники вообще лишатся работы.

Теперь машины используют и для найма на работу, иногда только на первом этапе, отсеивая резюме. Например, у вас великолепная квалификация, отличные рекомендации, но вас ни разу не пригласили на интервью, несмотря на то, что вы отсылали резюме в несколько организаций. Дело может заключаться в том, что у вас пенсионный возраст, и машина отсеяла вас еще на первом этапе. На подробную информацию о вас никто не смотрел! Машины просматривают профили в социальных сетях. Например, машина оценила вас как интроверта и одиночку — и все. Система решает, что вы компании не подходите, потому что работа подразумевает общение с клиентами. Но вполне может быть, что вы прекрасно работаете с людьми, а вне работы любите побыть в одиночестве. Однако алгоритм уже классифицировал вас соответствующим образом — и вы автоматически исключаетесь из кандидатов на должность.

К сожалению, подобных случаев в наше время много, никакой прозрачности оценок нет. Для человека поиск работы всегда стресс, но руководителям до этого нет дела. Наоборот, создается впечатление, что в современном мире менеджеры стараются этот стресс еще усилить. Людям предлагаются многочисленные тесты, результаты вводятся в машину, и она принимает решение. Руководство готово положиться на оценку системой, просто принимает выданный машиной результат, не желая тратить время и углубляться в изучение поданных резюме. Зачем с кем-то разговаривать, если есть умные машины? Но в некоторых видах бизнеса ни одна машина не может заменить личное общение с человеком. На людей надо смотреть, на человеческом уровне оценивать их умение говорить, общаться. Машина не может понять, какое впечатление производит человек.

В любом случае я советую вам очень внимательно и вдумчиво подходить к тому, что вы публикуете и загружаете на *Facebook*, *Twitter*, *Instagram*, *LinkedIn*. Даже если в компании отбор кандидатов на должность проводят люди, они будут смотреть ваши аккаунты в социальных сетях. Что-то сыграет вам на пользу — например, потенциальный работодатель увидит, что вы активно занимаетесь спортом и даже купаетесь зимой в проруби. Значит, будете меньше болеть и пропускать работу. Отлично! Но если на ваших страничках то и дело встречаются пьяные вечеринки, серьезная компания может отменить вашу кандидатуру.

Потенциальных работодателей могут заинтересовать и ваши друзья. С кем вы общаетесь в Интернете? Как часто вы публикуете посты? Сколько комментариев к ним получаете? Сколько у вас подписчиков в *Twitter*? На основе собранной в социальных сетях информации системы анализа *Big Data* могут сделать вывод о вашей способности работать в команде, умении устанавливать связи, об общительности. То есть в современном мире доступные данные о вас позволяют дать вам оценку, определить ваш тип личности. Отсутствие данных — это тоже характеристика, и некоторые компании могут насторожиться как раз это.

Но, как правило, нет механизма для проверки эффективности данных, использованных для предсказания результата. В данном случае — определения наилучшего кандидата на должность. Даже если найдено совпадение характеристик кандидата и требований к работе, не факт, что он подойдет идеально. Отбор вообще может происходить по прихоти создателя системы. Не подошел одной компании — может подойти другой.

Рейтинг

С большими данными связан и рейтинг организаций и людей. И этот рейтинг может в буквальном смысле решить судьбу. Он может значительно увеличить количество клиентов и, соответственно, принести дополнительную прибыль, а может навсегда испортить репутацию. К настоящему времени мы уже привыкли ставить оценки онлайн — мы оцениваем гостиницы, в которых останавливались, рестораны, в которых ужинали, купленные товары самого разного рода и назначения, уровень сервиса и конкретных работников.

Например, масса путешественников по всему миру учитывает рейтинг *TripAdvisor* — американского сайта путешествий, созданного в 2000 году. Он считается крупнейшим в мире, самым узнаваемым и заслуживающим доверия сайтом путешествий. С помощью этого сайта вы можете спланировать поездку в любую страну мира. Для пользователей услуги сайта бесплатны, а их очень много — 315 миллионов уникальных посетителей каждый месяц! Сайт работает в 45 странах на 25 языках. Большую часть контента создают сами пользователи — оставляют отзывы, загружают фотографии.

Этот сайт можно назвать силой в мире путешествий. Эту силу дают ему опубликованные рейтинги. Сайт становится фильтром для доступа к организации. Возможно, вы также пользовались или, по крайней мере, слышали про онлайн-платформу *Airbnb* — это площадка для размещения предложений, поиска и аренды частного жилья на недолгий срок по всему миру (в 191 стране и 65 000 городов). Можно сдать и снять жилье целиком, можно частично. Сайт помогает хозяину и гостю установить контакт. Профиль в *Airbnb* можно связать с аккаунтом в *Facebook*, и в таком случае вы сможете увидеть, где останавливались ваши друзья, и попросить их поделиться впечатлениями (если они этого еще не сделали). С 2008 года, когда эта платформа появилась, до 2017 года с ее помощью нашли жилье более 150 миллионов человек.

Заполнение профиля в *Airbnb* является обязательным. Профили пользователей (как хозяев, так и гостей) содержат рейтинги, отзывы и рекомендации, оставленные другими пользователями. Сайт предоставляет возможность приватной переписки хозяина и потенциального гостя. Оплата производится через сайт. После выезда гостя и гость и хозяин оставляют отзывы друг о друге и выставляют оценки по определенному набору критериев. На основании их формируется рейтинг обоих.

В компании по предоставлению услуг такси *Uber* клиенты выставляют оценки водителям, а водители выставляют оценки пассажирам. Соответствует ли этот рейтинг действительности? Нет. И этот рейтинг приводит к таким последствиям, о которых те, кто оценивал, и подумать не могли. Они не хотели создавать проблем водителю. Люди просто не понимали, к каким последствиям могут привести их оценки. Но они не виноваты — виноваты те, кто придумывал критерии, например в Великобритании. Они просто не знают англичан. Англичанин никогда не будет впадать в крайности, такова природа нации. Англичанин постарается избежать и самой лучшей, и самой худшей оценки. Всегда. Он поставит четыре звезды, если ему понравилось. Они будут означать «хорошо» или даже «очень хорошо». Но, по мнению компании *Uber*, приемлемый рейтинг для водителя — 4,6. И те водители, которые в среднем получают ниже, лишаются места, то есть их удаляют из базы данных. И здесь могут играть роль личные предубеждения.

Рассмотрим еще один пример рейтингов, на этот раз в системе образования. Эта система была придумана Мишель Рири, управляющей школами Вашингтона, округ

Колумбия. Она решила оценить качество преподавания. Но если все делать правильно, будет очень дорого, а это деньги налогоплательщиков. Мишель Рири решила не тратить на многие часы мониторинга экспертами, а просто ввести в машину оценки учащихся. Эти данные есть во всех школах, не нужно напрягаться, их собирая.

А ведь это не прямые показатели работы учителей. Оценки учеников совсем не обязательно свидетельствуют о том, насколько хорош или плох тот или иной учитель. Четкой и прямой связи нет. Но так оценивать легко и дешево. В первый год использования этой системы оценки учителей уволили 2% учителей, оказавшихся внизу рейтинга. На следующий год работу потеряли уже 5%. В общей сложности — свыше 200 профессиональных преподавателей.

Таким образом, можно сделать вывод, что системы должны базироваться на параметрах, которые отражают квалификацию, эффективность, качество работы и т.д. И нельзя верить в то, что компьютер всегда прав, потому что беспристрастен. Он хорош настолько, насколько хорошо написан алгоритм, насколько высоко квалифицирован человек, который его писал.

Кредитный рейтинг

Мы уже говорили про то, что вас могут не взять на работу, потому что так решила машина. Но вы также можете не получить кредит. Раньше вы общались с банковским менеджером, который знал вас лично, он составлял впечатление о вас и решал, рискует ли банк, выдавая вам кредит — и оправдан ли риск. Но теперь решение о выдаче вам кредита принимает алгоритм, действующий на основе больших данных.

Система собирает данные о кредитах, которые вы брали, о том, как вы их выплачивали, о том, были ли задержки и пропуски платежей, как вы платили штрафы. Конечно, учитывается информация о ваших доходах, об имеющемся имуществе и его стоимости, членах семьи, как работающих, так и иждивенцах. Алгоритм интерпретирует все эти данные и выносит вердикт, причем делает это быстро, в особенности в сравнении с тем временем, которое требовалось на одобрение кредита в те времена, когда технологий *Big Data* еще не существовало. И если личный банковский менеджер вам что-то объяснял, то теперь у вас нет никакой возможности узнать, почему система оценивала вас таким образом. И вы не можете указать на ошибку, как при общении с личным банковским менеджером. Например, вы один раз задержали платеж не по своей вине, а по вине банка, компьютеры которого вышли из строя. Вам не был начислен штраф, потому что это не ваша вина, но в системе имеется информация о дате платежа, и алгоритм рассматривает ее как задержку платежа. А вам из-за этой задержки платежа по вине банка предлагают менее выгодные условия по кредиту, потому что к такому выводу пришла система.

Самая известная в Великобритании микрофинансовая организация *Wonga* является не только лидером в кредитовании «до зарплаты», но и имеет наименьший процент невозвращенных кредитов. Они очень рано поняли, насколько полезны технологии *Big Data*, и ими пользуются. *Wonga* собирает всю информацию об обратившихся в компанию людях — все, что только можно найти в Интернете, и не скрывает этого. Они активно изучают социальные сети, друзей, их кредитную историю, технологию, которой вы воспользовались для обращения в *Wonga* (это онлайн-платформа), местность, где вы живете физически. Алгоритм, принимающий решение о выдаче краткосрочного кредита, действует по не совсем понятным правилам. Но это срабатывает! Например, изначально компания использовала предположение, что человек, друзья которого обычно возвращают кредиты, тоже вернет кредит. Скорее да, чем нет. Человек, который живет в той местности, где люди

обычно возвращают кредиты, тоже вернет кредит. Скорее да, чем нет. Система настраивалась с течением времени и становилась все более и более эффективной. Иногда критерии кажутся странными, иногда включение их в систему кажется безумием, но, если это срабатывает, значит, компания работает правильно. Люди физически не смогли бы проверить все эти критерии, их связь с возвратом или невозвратом кредита, да, наверное, и не стали бы, если бы даже могли. А современные технологии позволяют это сделать. Компания успешна, пусть и действует на основании безумных (на первый взгляд) предположений. А это — самое главное.

Но большие данные также испортили массу кредитных историй, поставили на людей черные метки, которые смыть невозможно. Они осложнили людям жизнь сейчас и в будущем, потому что информация навсегда останется где-то в дебрях Интернета, и другой алгоритм, через десять и двадцать лет, сможет вытянуть ее на свет и осложнить человеку проведение финансовых операций. Внести исправление крайне сложно, иногда невозможно. Так почему мы допускаем, что какой-то скрытый алгоритм может разрушить человеческую жизнь, повлиять на судьбу?

Был ли предыдущий подход лучше? Трудно сказать. Во всем есть свои плюсы и минусы. Я считаю, что использование технологий *Big Data* все-таки лучше, чем личный банковский менеджер, но при одном условии — открытые алгоритмы. Мы должны иметь возможность понять, почему получили именно такой кредитный рейтинг и как можно быстро исправить ошибки.

Белое или черное?

Рейтинговые агентства сейчас работают по всему миру и используют большие данные. Это организации, присваивающие кредитные рейтинги, которые показывают способность должника вернуть долг, своевременно уплатить проценты, а также вероятность дефолта. Существуют не только кредитные рейтинги, но и рейтинг качества управления, надежности компании и т. д. О ценности кредитных рейтингов для ценных бумаг уже давно идут споры по всему миру. «Большая тройка» международных рейтинговых агентств (*Moody's, Standard & Poors, Fitch Ratings*) контролирует примерно 95% рейтингового бизнеса.

В последние годы в США и Великобритании возникли самые разные рейтинговые агентства, основатели которых поняли, что появилось много потенциальных клиентов, желающих узнать как рейтинги организаций, так и частных лиц. Упомянутая выше «большая тройка» никогда не интересовалась частными лицами. А новые агентства продают частным лицам доступ к их собственным данным. Люди не могут его получить сами, так как не имеют доступа к технологиям *Big Data*. В ряде стран вопрос уже обсуждается на законодательном уровне — мы должны иметь право доступа к данным о нас самих.

Но рейтинговые агентства пока продолжают торговлю данными. Например, некоторые организации обращаются к ним за кредитным рейтингом людей, пытающихся устроиться к ним на работу. Но ведь если человек просрочил свой кредит, это совсем не значит, что он будет плохим администратором. Он может быть очень щепетилен в плане чужих денег. Или недавний студент устраивается на работу, а у него не выплачен кредит на образование. Как он мог его выплатить, если еще не работал? Но у него уже плохой кредитный рейтинг.

В целом трудность при работе с большими данными заключается в том, что кому-то они приносят большую пользу, а кому-то — вред и могут даже сломать жизнь. Мы часто осуждаем использование технологий *Big Data*, если пользу получает только государство или компания и при этом не получают отдельные люди. Есть алгоритмы, в

которых никто не видит смысла. В любом случае нужно взвешивать пользу технологий *Big Data* для большинства, учитывая последствия для меньшинства или даже отдельных людей.

Мне кажется, нельзя говорить, что большие данные — это очень хорошо или что это очень плохо. Это не белое и черное. У больших данных много оттенков, и мы должны максимально использовать возможности, которые у нас появились благодаря им, и корректировать недостатки.

Список использованных источников:

1. Просто Big Data. — СПб.: Страта, 2019. — 148 с.