

Review

Тематическое приложение к газете **Коммерсантъ**

Think big data

Понедельник 26 сентября 2022 №177/II (7378 с момента возобновления издания)

kommersant.ru

15 Как Big Data помогает оценивать кредитоспособность заемщиков

15 За счет чего концепция Data Fusion расширяет знания компаний о потребителе

16 Зачем бизнесу технологии конфиденциальных вычислений

Узнать клиента из миллиона

Применение анализа больших данных становится все более актуальным во время текущего кризиса на рекламном рынке: на фоне сокращения бюджетов и инвентаря рекламодатели ищут новые эффективные способы продвижения. Инструменты, позволяющие находить целевую аудиторию исходя из ее интересов и привычек, уже применяются не только в интернете, но и на других площадках, в том числе на телевидении. Что за возможности сегодня предоставляет Big Data в рекламе, с какими барьерами она сталкивается и какое будущее ей прочат эксперты?

— реклама —

Российский рекламный рынок по итогам 2022 года изменится даже сильнее, чем во время пандемии коронавируса. Сокращаются как бюджеты, так и доступный инвентарь: иностранные бренды останавливают продвижение, и в то же время для российских рекламодателей стала недоступной закупка рекламы на крупных иностранных интернет-платформах. В результате общий объем рекламного рынка в России может упасть на десятки процентов — рекламная группа Group4Media (ранее GroupM), например, прогнозировала снижение на 38% по итогам года, NMI Group — более чем на 50%.

Одновременно с изменением состава доступных рекламных инструментов происходит перераспределение рекламных бюджетов. Приток пользователей получают не только альтернативные российские площадки: соцсети, видеосервисы и интернет-платформы. Компании, которые остались на рынке и получили возможность усилить свои позиции благодаря импортозамещению, ищут возможности повысить эффективность продвижения. С марта, в частности, растет интерес к рекламным инструментам, использующим анализ больших данных о клиентах, что позволяет легче доставлять рекламное сообщение нужной аудитории — как в интернете, так и даже на ТВ.

Реклама по интересам

Программатик-платформы позволяют рекламодателям покупать через автоматизированную систему рекламу, которая показывается только определенным категориям пользователей или зрителей. Отличием таких платформ от обычной автоматизированной покупки рекламы является возможность качественно таргетировать показы по профилям пользователей. Так, рекламодатели могут указать точные характеристики целевой аудитории — демографические и социальные, а также предпочтения и интересы и недавние события из жизни. При этом такой способ закупок позволяет оптимизировать рекламные кампании в режиме реального времени также по конкретным параметрам.

Найти нужную аудиторию позволяет массив обезличенных данных о клиентах. Например, программатик Platforma объединяет анонимизированные данные нескольких партнеров, в числе которых банки, мобильные и телеком-операторы, сервисы платного ТВ и т. п. Эти данные позволяют разделить пользователей более чем на 2,5 тыс. сегментов. В целом применение Big Data уже вышло за пределы закупок рекламы в интернете: компании объединяют данные о пользователях из самых разных сфер, чтобы продемонстрировать рекламу в подходящее время заинтересованным зрителям. На основе больших объемов Big Data строятся сегменты для кампаний не только в интернете, но и на ТВ. Например, можно показать рекламу детского питания по телевизору тем, кто искал в интернете детские товары, или баннер на сайте со спортивными товарами — тому, кто вчера смотрел по ТВ спортивные тренировки.

Как это работает

О кейсах применения таких технологий рассказали в Platforma. Среди примеров — персонализированная рекламная кампания развлекательного телеканала «Суббота!» от создателей телеканала «Пятница!», который запустился в 2021 году вместо телеканала «Супер». Одной из первоочередных задач компании было привлечение к телезрителю интернет-аудитории.

Аудиторные сегменты на данных банков и телеком-операторов показывают более высокую эффективность по сравнению с традиционными социально-демографическими

Platforma решила показать рекламу интернет-пользователям: они видели рекламу в тот момент, когда были дома, подключались к Wi-Fi и могли сразу включить телеканал, который им только что рекламировали. Для врезки баннерной рекламы в трафик привлекли Quiet Media — продавца рекламы на инвентаре операторов связи. Кампания проходила по всей России в сентябре — ок-



тябре 2021 года, целевой аудиторией были выбраны женщины от 18 до 44 лет с IPTV. Для рекламы использовали мультиформатные и fullscreen-баннеры с призывом «Посмотри прямо сейчас!».

Чтобы оценить эффективность, в Platforma сопоставили, сколько потенциальных зрителей находят-ся дома (пользуются домашним интернетом) и прямо в моменте гото-

вожила увеличить долю канала на 12%, по данным Mediascope с 1 ноября по 31 декабря 2021 года.

Анализ Big Data может не только помочь провести определенную рекламную кампанию в моменте, но и оценить качество уже состоявшихся, а также в целом изучить свою аудиторию, чтобы сформировать долгосрочную стратегию продвижения. Так, для сети Selgros Cash & Carry (работает в 6 российских городах и управляет 11 магазинами) было организовано исследование совместно с агентством Proximity Media, платформой TVID Technology, компаниями MediaHills и Platforma.

Партнеры составляли аудиторные сегменты не только по социально-демографическим признакам, но и с учетом поведенческих факторов: интересов, покупок и др. Затем аналитики построили полный путь пользователя, используя анонимизированные данные множества партнеров, в том числе «Ростелекома», ВТБ и видеосервиса Wink.

В итоге исследование показало, что число уникальных клиентов торговой сети во время ТВ-кампании увеличилось на 25%, а среднее количество транзакций в сутки выросло

на 10%. Наибольший отклик реклама получила у женской аудитории: среди тех, кто купил что-то в Selgros после просмотра ТВ-ролика, женщин было в два раза больше. Кроме того, без продвижения на ТВ лишь 5% исследуемой аудитории приходят за покупками в магазины сети в течение месяца, но с медийной поддержкой их число увеличивается вдвое. Оказалось также, что наиболее эффективными для торговой сети стали размещения на «Первом канале» в Волгограде, а в Москве наибольшую пользу принесли «Россия 1» и НТВ. На базе этой аналитики сеть сможет оптимизировать микс каналов в следующих рекламных кампаниях.

Трудности сегментирования

Один из вопросов, которые всегда поднимаются при обсуждении возможности применения Big Data, — это возможные изменения в законодательстве в сфере данных о пользователях. По словам руководителя практики по сопровождению IP- и IT-делок Semenov & Pevzner Ирины Остапчук, в мире есть четкий тренд на его ужесточение. Еще в 2018 году, напоминает она, вступил в силу общий регламент по защите данных в ЕС (GDPR), в 2020-м — закон Калифорнии о конфиденциальности потребителей (CCPA), который называют американским аналогом GDPR. Регуляторы по всему миру более тщательно подходят к соблюдению этих законов и накладывают большие штрафы на нарушителей. Так, в 2021 году Amazon была оштрафована на €746 млн: она без согласия пользователей использовала их профили для таргетированной рекламы. Недавно в Калифорнии Sephora оштрафовали на \$1,2 млн за нарушение CCPA: выяснилось, что компания продавала персональные данные покупателей для аналитических и маркетинговых технологий.

Ограничениям подвергаются и cookies. Это небольшой фрагмент данных, который отправляется веб-сервером и хранится на компьютере пользователя, в cookie находится информация о пользователе, например логин и пароль, местоположение, языковые настройки, сведения о товарах в корзине интернет-магазина. Согласно GDPR, перед тем, как уста-

новить cookies на компьютер пользователя, сайты должны обязательно запрашивать согласие. Поэтому с тех пор, как закон вступил в силу, чтобы начать пользоваться каким-то сайтом, надо сначала нажать на множество согласий в множестве окошек.

В рамках российского законодательства в настоящий момент нет четкого понимания того, относятся ли cookie-файлы к персональным данным пользователя, указывает руководитель направления «Разрешение IT & IP-споров» юридической фирмы «Рустам Курмаев и партнеры» Ярослав Шидле. Использование cookies не запрещено, но, чтобы обрабатывать такие данные законно, необходимо как минимум предусмотреть упоминание о том, что такие данные собираются, в документации, публикуемой на сайте, отмечает юрист BCP Litigation Анастасия Мырзина: «Наилучшей практикой является выведение отдельного окна с просьбой дать согласие на действия с cookies».

Что если не будет cookies

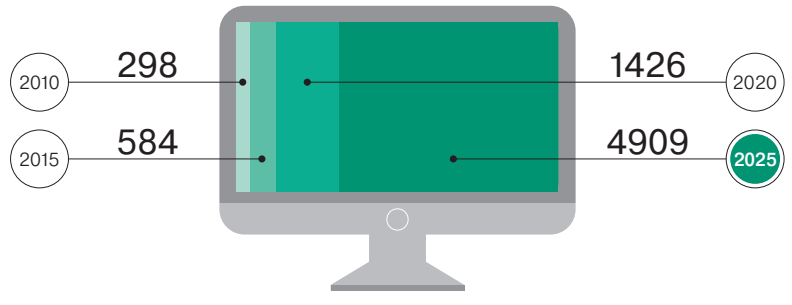
Пока российские законодатели не предлагали инициатив в отношении cookies. Однако юристы допускают, что рано или поздно российская практика пойдет по международному пути. «Я не исключаю, что в обозримом будущем могут появиться законодательные ограничения, связанные со сбором обезличенных данных в России зарубежными операторами и передачей уже сформированных массивов зарубежным компаниям для их использования в коммерческих и иных целях», — предполагает, например, партнер Enterprise Legal Solutions Юрий Федюкин.

Но даже если cookies будут так или иначе ограничены в будущем, на российском рынке уже есть инструменты, которые помогают обходиться совсем без них и при этом продолжать находить целевую аудиторию с помощью данных. Один из них — Stable ID. Это идентификатор, в обезличенном виде привязанный к какому-либо стабильному параметру, например почте, номеру устройства или мобильному телефону. Идентификатор работает исключительно с обезличенными агрегированными анонимизированными данными, которые невозможно соотнести с конкретными людьми.

с 15

ЧИСЛО ВЗАИМОДЕЙСТВИЙ С ЦИФРОВЫМИ ДАННЫМИ КАЖДОГО ИНТЕРНЕТ-ПОЛЬЗОВАТЕЛЯ В СУТКИ

ИСТОЧНИК: IDC И SEAGATE



Восторг и страх интернет-рекламы

— экспертное мнение —

Основатель компании Alfadart МИХАИЛ ГЕТМАНОВ — о персонализации рекламы и приватности данных.

Успехи рекламных технологий в интернете за 20 последних лет очевидны даже для неспециалиста. Как уже не удивляет, когда после того, как мы посмотрели в интернет-магазине условный «утюг», реклама этого «утюга» начинает преследовать нас на других сайтах. Такая персонализация рекламы стала для нас не просто привычной, но даже ожидаемой.

Несмотря на логичность и видимую простоту персональной рекламы, за каждым показом рекламного баннера или видеоролика стоят огромные вычислительные мощности и массивы данных, которые нужны для того, чтобы подобрать каждому пользователю интернета именно ту рекламу, которая его заинтересует.

За последние годы с помощью технологий Big Data рекламные платформы научились сохранять в своих базах данных и затем учитывать мельчайшие нюансы нашего поведения в интернете, которые могут повлиять на наш интерес к той или иной рекламе. При этом высшим пилотажем считается способность рекламной платформы «склеивать» данные одного и того же пользователя сразу на разных



устройствах, которые он использует для выхода в интернет: компьютеры дома и на работе, мобильный телефон, «умный» телевизор.

Пока что лишь немногие рекламные платформы умеют делать это эффективно. Однако если это удастся, то можно персонализировать рекламу не только на сайтах в интернете или в мобильных приложениях, но даже на цифровых уличных рекламных конструкциях. Так, крупнейший российский интернет-холдинг VK умеет показывать нужную рекламу именно в тот момент, когда доля целевой аудитории рекламодателя возле рекламного щита выше, например 50%. Для этого используются данные из про-

филей пользователей в соцсетях VK и «Одноклассники», а также история посещения сайтов в сети.

Все это напоминает известную сцену из фантастического боевика «Особое мнение», снятого по мотивам одноименного рассказа Филипа Дика. Главным героем, которого играет Том Круз, проходит по футуристическим коридорам станции метро и видит на экранах рекламные ролики, на которых к нему обращаются по имени и фамилии. Эта сцена вызывает у зрителя одновременно чувство восторга и страха, и это не просто талантливая работа режиссера фильма. Судя по всему, природа уже заложила в человека реакцию отторжения к тому, чтобы его персональные сведения использовались без разрешения.

Поэтому неудивительно, что в последние годы наряду с трендом персонализации интернет-рекламы появился противоположный ему тренд приватности. Дело в том, что возможность персонализации рекламы достигается путем передачи специальных идентификаторов пользователя между участниками рекламной экосистемы, то есть между сайтами и рекламными платформами. Для этого задействуется механизм так называемых файлов cookies. Сторонники ужесточения приватности в интернете добиваются того, чтобы ограничить передачу cookies, считая,

что таким образом интернет-компании, в том числе рекламные платформы, не смогут идентифицировать пользователей.

Главными действующими лицами тренда приватности являются браузеры, с помощью которых все мы путешествуем по интернету. Именно браузеры в состоянии разрешить или запретить передачу cookies и, следовательно, разрешить или запретить возможность идентификации пользователей. Первым ограничения на cookies ввел браузер Safari от компании Apple. Затем инициативу подхватил Firefox. Позднее ограничения на cookies ввел и российский «Яндекс. Браузер». Оставшийся (и при этом самый крупный по количеству пользователей) Google Chrome планирует ввести такие ограничения в 2024 году.

Участники рынка интернет-рекламы со страхом ожидают того момента, когда Google «отменит» cookies. Ведь это значит, что реклама не просто перестанет быть персонализированной, реклама также станет менее эффективной и более дорогой. Для того чтобы привлечь в интернете то же самое количество покупателей, рекламодателю придется заплатить за рекламу не только для пользователей из своей целевой аудитории, но также и для многих других пользователей, которым его товар уж точно не нужен. И все это из-за того, что нель-

зя будет идентифицировать нужных людей с помощью cookies.

Тем не менее участники рынка не падают духом и ищут способы, которые позволят персонализировать рекламу, несмотря на блокировку cookies. Некоторые компании планируют использовать email-адреса, чтобы узнавать пользователей на сайтах. Другие компании начинают использовать технологии машинного обучения для того, чтобы «угадывать» пользователя на основании его косвенных признаков, например разрешения экрана, настройки звука, установленные шрифты и другие настройки браузера. Такие технологии называются технологиями фотггерпринтинга (от англ. «fingerprint» — «отпечаток пальца»).

Google тоже хочет каким-то образом облегчить жизнь рекламным платформам, разрабатывая новые технологии для персонализации рекламы без использования cookies, однако на рынке есть большой скепсис по поводу этих усилий.

Неожиданная помощь может прийти от операторов связи, которые в последние годы стали очень активны на рекламном рынке. Дело в том, что операторы связи в любом случае идентифицируют, то есть распознают в сети каждого своего абонента независимо от того, каким устройством или браузером он пользуется. И

теперь операторы смотрят на трудности с персонализацией рекламы как на возможность предложить рынку новую уникальную услугу — идентификация пользователей.

Естественно, такая услуга должна соответствовать всем условиям закона «О защите персональных данных». Операторы не будут отдавать рынку данные о своих абонентах, но позволят рекламным платформам использовать их анонимные зашифрованные идентификаторы. Эти операторские идентификаторы помогут персонализировать рекламу так же эффективно, как старые cookies. И, судя по всему, нужный уровень приватности операторы тоже смогут обеспечить, ведь новый механизм не будет общедоступным, как cookies (так, Tele2 и Platforma в июне протестировали доставку аудиторных сегментов, построенных на стабильных анонимизированных идентификаторах, в рекламную технологическую платформу Yabbi. — **Review**).

Современные рекламные технологии уже способны реализовать то, что казалось фантастикой 20 лет назад, когда вышел фильм «Особое мнение». Но важно, чтобы сегодня рынок интернет-рекламы выбрал именно такие технологии персонализации, которые будут также вызывать восторг обычного пользователя и при этом не будут его пугать.

Review Think big data

Поделиться геолокацией

Бизнесу особенно важно находиться в нужное время в нужном месте: неудачное расположение офиса может стать причиной потерь от 30% до 50% первоначальных инвестиций в его создание, а в некоторых случаях даже привести к закрытию компании. Чтобы избежать такого сценария и предотвратить ущерб, компании прибегают к услугам геомаркетинга, который позволяет оценить перспективность локации для открытия бизнеса, выявить факторы, влияющие на спрос, и запланировать поставки. Геоаналитика широко используется в финансовой сфере, ритейле, сфере услуг и FMCG, при этом отдельными инструментами для проверки гипотез, например маркетплейсом геослоев, может воспользоваться даже малый бизнес.

— геоаналитика —

Маркетинг к месту

Среди проблем с ведением бизнеса 11% предпринимателей называют расположение бизнеса, следует из исследования Промсвязьбанка, «Опоры России» и агентства Magram Market Research, проведенного по итогам 2021 года. Невозвратные инвестиции из-за неудачной локации могут доходить до 30–50%, оценивает генеральный директор разработчика решений для бизнеса на основе больших данных Platforma Алексей Каштанов: «Например, если при открытии салона красоты вы тратите 1 млн руб., то можете потерять около 300 тыс. руб.».

Оценить перспективность места расположения торговой точки можно с помощью геомаркетинга. Этот инструмент позволяет проверить бизнес-гипотезы, а также выявить факторы, влияющие на спрос в конкретной локации. Применение продуктов геоаналитики при выборе оптимального размещения торговых точек и офисов продаж позволяет увеличить выручку до 20%, оценивали PwC и Strategy&. По оценке экспертов компании Platforma, использование геоаналитики у 20% компаний-клиентов позволило увеличить выручку на 30% и средний срок жизни — на 9–12 месяцев.

Это направление маркетинга стремительно развивается во всем мире: аналитики MarketsandMarkets ожидают, что объем глобального рынка геомаркетинга вырастет с \$10,7 млрд в 2020 году до \$32,5 млрд к 2025 году при совокупном годовом темпе роста (CAGR) на уровне 25% в течение прогнозируемого периода. В России, по данным ВТБ, рынок геоаналитики растет на 15–20% в год. По оценкам iKS-Consulting, в 2020 году рынок российских сервисов геоаналитики приближался к объему 850 млн руб.

Одним из драйверов роста востребованности сервисов геоаналитики стала пандемия. «Резкие изменения в бизнесе из-за пандемии COVID-19 сделали модели, основанные на большом количестве исторических данных, менее актуальными. Теперь принятие решений требует большего разнообразия данных для лучшего понимания ситуации», — отмечали аналитики Gartner, ука-



дании: в одних офисах произошел наплыв клиентов, тогда как другие пустовали.

Банк решил модернизировать сеть, причем при выборе локации и формата для офисов ему требовалось принять в расчет не только стандартные факторы вроде пола и возраста жителей, но и необычные — например, чтобы дорога пешком от метро занимала не больше 15 ми-

500х500 м. Для каждого сектора аналитики строят прогнозы продаж кредитных продуктов, привлечения депозитов и новых клиентов на операции, приносящие банку комиссионный доход. Результат визуализируется при помощи тепловых слоев: в тех местах, где показатель высокий, слой имеет более насыщенный цвет.

Такой метод позволил ВТБ выбрать оптимальные локации для новых отделений. Чтобы банк мог определиться с форматом офисов, аналитики с помощью алгоритмов машинного обучения рассчитали интегральный рейтинг интереса к кредитным продуктам в каждой локации. Результат показан в виде сине-красной шкалы, где в синюю зону попадают метрики с небольшим вкладом, а в красную — наиболее значимые. Например, если в красную зону одной из локаций попали два параметра из трех — продажи кредитных продуктов и депозитов, то система оценит это местоположение как благоприятное для открытия офиса, где состав команды продаж будет ориентирован в большей степени на первые два продукта.

В результате применения геоаналитики ВТБ оптимизировал сеть офисов в 68 городах России и за полгода сократил накладные расходы на аренду и содержание помещений на 10%, увеличил поток новых клиентов на 19%, а клиентскую базу — на 6%. Сегодня ВТБ, также, как и другие крупные российские банки, продолжает оптимизировать сеть с помощью геомаркетинга.

Геоаналитика на малых оборотах

Инструменты геомаркетинга широко применяются не только в финансовой сфере, но и в ритейле, предприятиях из сферы услуг и FMCG-компаниях. В этих областях расположение офисов имеет принципиальное значение с точки зрения конкурентных преимуществ. Удачное расположение важно бизнесу любого размера, однако бюджеты малого не всегда предусматривают услуги геомаркетологов.

Сделать геоаналитику доступнее для малого бизнеса, а также для компаний, которые планируют наружную рекламу, позволяет маркетплейс геослоев, который в сентябре запустила Platforma.

Преимущество маркетплейса состоит в том, что он позволяет купить уже готовый рецепт решения своей боли — предрасчитанный с помощью ИИ геослой, демонстрирующий оптимальные места для открытия торговых точек по 15 отраслям

(в октябре ожидается расширение до 30 отраслей). Есть и ручной режим работы с маркетплейсом, когда заказчик сам выбирает геослой для их пересечения. Так, можно находить зоны высокой концентрации целевой аудитории, смешивать со слоем сумм трат целевой аудитории на товары целевой категории товаров, а также учитывать как текущее расположение прямых конкурентов, так и наличие коммерческой недвижимости по рыночным ценам.

Стоимость геослоев на маркетплейсе — от 500 руб. для городов-миллионников до 10 тыс. руб. для Москвы. Предрасчитанный ИИ слой стоит от 2,5 тыс. руб. для городов-миллионников до 50 тыс. руб. для Москвы. Все слои приобретаются и остаются в доступе в личном кабинете заказчика.

До конца года GeoPlatforma намерена реализовать возможность загрузки в систему данных заказчика, преобразовав их в геослой, чтобы включить их в анализ или моне-

тизировать, продавая доступ другим клиентам платформы. При этом круг потенциальных покупателей можно ограничить, например исключить из него прямых конкурентов.

После выбора локаций заказчик может с помощью готовых математических моделей провести необходимый анализ данных — выбрать наиболее перспективные локации для развития бизнеса или планировать оптимизацию уже имеющейся сети с учетом особенностей каждой площадки. Таким образом, геоанализ на пяти слоях может обойтись компании в 2,5–50,0 тыс. руб., и он позволит «существенно сократить риск ошибки при выборе места для будущего бизнеса», подчеркивает Алексей Каштанов.

Мастерство предсказания

Большие данные из разных отраслей применяются также для прогнозирования спроса и планирования поставок. По оценкам McKinsey Digital, влияние искусственного интеллекта на планирование цепочки поставок составляет от \$1,2 трлн до \$2 трлн. Так, треть продаж американской Amazon приходится на персональные рекомендации, а международная служба доставки FedEx с помощью прогнозной аналитики с точностью от 65% до 90% определяет клиентов, готовых уйти к конкурентам, и делает им персональные предложения.

Многие российские компании также внедряют собственные системы прогнозирования спроса — например, предиктивная система торговой сети «Пятерочка» на основе чековой аналитики прогнозирует вероятность продажи товара с точностью до часа и рассчитывает заказы на пополнение полок, а маркетплейс Ozon благодаря системе прогнозирования перестал закупать товары у поставщиков с запасом и сэкономил на аренде складских помещений. Внедрение нейросетей для анализа спроса проводил и «Магнит»: по прогнозам аналитиков ритейлера, предиктивная система позволит увеличить выручку на 4 млрд руб. в год за счет снижения дефицита товаров и сэкономить 1 млрд руб. на сокращении списаний продукции.

Прогноз потребительских предпочтений может быть как составлен на основе собственных данных компании, так и собран на основе различных источников — в таком случае он позволит выйти за пределы поведения клиента в магазине и дать более широкую картину в связке с информационным полем. Например, на основе данных социальных сетей американская сеть Walmart выяснила, что пользователи в восторге от Cake Pops, и оперативно запустила этот десерт в своих магазинах.

Услугу по прогнозированию спроса на основе больших данных о клиентах от компаний различных отраслей в России предлагает компания Platforma. Обезличенные данные о чеках и транзакциях компаний-партнеров, в том числе банков, соцсетей, телеком-операторов, операторов фискальных данных и ритейлеров, попадают в хранилище Platforma. С помощью ML-моделей аналитики ищут точки пересечения между базами данных и выстраивают клиентский путь на разных площадках. В зависимости от запроса клиента Platforma формирует пользовательские сегменты, составляет портрет типичного представителя аудитории и прогнозирует его покупки.

Юлия Степанова

Применение продуктов геоаналитики при выборе оптимального размещения торговых точек и офисов продаж позволяет увеличить выручку до 20%

зывая на тренд перехода от больших данных к более широкому, собранному из различных источников. Такие данные лежат в основе инструментов геомаркетинга — например, геоаналитическая платформа GeoPlatforma содержит более 1 тыс. слоев данных для каждого квадрата геосетки 500х500 м по городам с населением более 500 тыс., получаемых из различных сфер бизнеса: IT, финансовой отрасли, телеком-операторов и ритейлеров.

Банк данных

Когда COVID-19 поменял маршруты и график перемещения людей, компании столкнулись с оттоком трафика и были вынуждены перестраивать бизнес с учетом новых реалий. Так, в ВТБ заметили изменение потоков посетителей во время пан-

нут, а клиенты в радиусе 500 м от локации хотя бы раз интересовались ипотекой, потребительским кредитованием, автокредитами за последний год. Чтобы решить эту задачу, ВТБ использовал геоаналитические инструменты: карту с геоданными для 68 городов России. Анонимизированные массивы данных собраны в облачном хранилище, где искусственный интеллект ищет закономерности между ними в зависимости от запроса клиента.

Для решения задачи ВТБ из массива данных была отобрана информация о населении, уровне доходов, интересе к банковским продуктам, кредитных историях, а также данные о расположении офисов конкурентов и доступной коммерческой недвижимости. Карты городов на платформе разделены на квадраты

«Эффективность геоданных зависит от способности искать нестандартные подходы»

— цена вопроса —

Руководитель по развитию геоаналитических сервисов Platforma АЛЕКСЕЙ АЛЕКСЕЕВ — о выборе геоаналитического сервиса.

Использование геоаналитики стало в последние годы не столько приятным дополнением, повышающим эффективность бизнеса, сколько реальной необходимостью. При этом для России эта услуга новая — пока на рынке геомаркетинга не более десяти игроков, что при отсутствии западных конкурентов дает прекрасную возможность отечественным стартапам проявить себя. Драйвером роста спроса на геоаналитику является B2G-сектор, тогда как в секторе среднего и малого бизнеса наблюдается низкий уровень знаний



о возможностях искусственного интеллекта (ИИ) и геоаналитики.

Для повышения осведомленности делается уже немало, в том чи-

сле отдельные инструменты геоаналитики предлагаются и бесплатно: в Москве действует госпрограмма по подбору помещений для бизнеса, на аналитическом портале Федеральной налоговой службы можно увидеть показатели в региональных срезах, портал Бизнес-навигатора МСП отмечает на карте всевозможные предприятия, ВП-портал с оперативной статистикой Росстата показывает актуальные данные.

Собственные сервисы геоаналитики создают и крупные компании на основе собственных баз больших данных. На реализацию подобных проектов уходит много времени, а стоимость их разработки — от 300 млн до 600 млн руб. Подобные платформы могут учитывать специфические требования заказчика, они адаптированы под внутренние

процессы и дают возможность строить математические модели с использованием чувствительной внутренней информации компании.

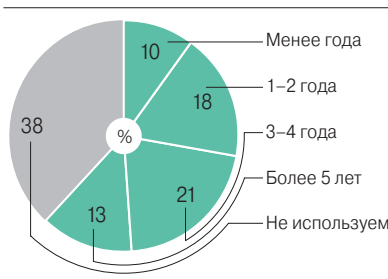
Оптимальным по соотношению стоимости и получаемого за эти деньги функционала решением может быть «коробочная» платформа геоаналитики. Подобная платформа не настолько гибкая, как собственные решения крупных компаний, однако ее стоимость для решения базовых задач малого и среднего бизнеса — от 50 тыс. до 200 тыс. руб. При ежемесячном использовании геоплатформы по двум-трем городам стоимость достигает нескольких миллионов в год, а для крупного ритейл-бизнеса в случае регулярного использования полного функционала геоплатформы цена составляет от 50 млн до 100 млн руб. в год.

Плюсом геоплатформ является также то, что они помогают транслировать опыт крупных населенных пунктов на более маленькие. Дело в том, что большинство данных, которые сейчас накоплены, содержат сведения именно о крупных городах. При этом в современных условиях центрами развития малого бизнеса становятся не крупные города, а небольшие населенные пункты — например, довольно прибыльные магазины открываются именно в малых городах, поскольку практически каждый из проживающих там проходит мимо подобных магазинов. Так как данные о маленьких населенных пунктах ограничены, задачи геоаналитики в них решаются во многом благодаря креативности продуктовых команд и нахождения новых нестандартных подходов. На-

пример, некоторые ритейлеры при заходе в новый регион анализируют ночные спутниковые снимки, чтобы оценить прирост числа освещенных улиц города. Скорость разрастания освещенности города косвенно свидетельствуют об экономическом росте конкретного города. Это актуально, например, для продавцов стройматериалов: они видят, что в регионе увеличилось количество жилых районов и спрос на отделочные материалы может также вырасти.

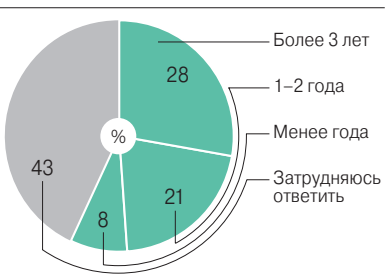
В целом при выборе геоаналитического сервиса я бы советовал сравнить удобство интерфейсов, оценить широту спектра геоданных, периодичность и регулярность их обновления, точность математических моделей и сопоставить это со стоимостью решения бизнес-задачи и гибкостью тарифов.

КАК ДАВНО ВЫ ИСПОЛЬЗУЕТЕ РЕШЕНИЯ ДЛЯ РАБОТЫ С BIG DATA?



ИСТОЧНИКИ: ИССЛЕДОВАНИЕ VK CLOUD И ARENADATA, ОПРОС ГЛАВ IT-ДЕПАРТАМЕНТОВ, СТО, СТО, СДО, СДО КРУПНЫХ РОССИЙСКИХ КОМПАНИЙ, СЕНТЯБРЬ 2022 ГОДА

СРОКИ ОКУПАЕМОСТИ ПРОЕКТОВ BIG DATA



Ускорение скоринга

Скорость и точность принимаемых решений — наиболее важные характеристики, когда речь идет о кредитном скоринге, то есть оценке платежеспособности лица, желающего получить кредит. Ошибка в оценке может стоить бизнесу миллионов потерь, а долгие сроки рассмотрения заявок уменьшают конкурентоспособность банка. Поэтому финансовые организации ищут все новые способы повысить эффективность скоринга. Расширенное число источников информации о потенциальном клиенте — один из них. Подобные услуги предлагают, например, скоринг-платформы.

— финсектор —

Интеллект в банках

В условиях ужесточающейся конкуренции необходимо максимально быстро выводить на рынок новые продукты, считают более чем в трети банков, следует из опроса «ЛАНИТ-Би Пи Эм» (входит в группу «Ланит»), проведенного в 2021 году. По мнению банков—участников опроса, для повышения качества и скорости принимаемых решений наиболее приоритетными технологиями являются Big Data и продукты на базе искусственного интеллекта (ИИ). В ближайшие два года треть крупных организаций будет использовать ИИ для принятия решений ради конкурентных преимуществ, прогнозировали аналитики Gartner.

в баллах и в зависимости от их количества принимает решение об одобрении или отказе в выдаче кредита.

Клиенты традиционных банков могут ждать одобрения заявки на кредит до недели, тогда как банки, использующие технологию ИИ, анализируют данные в режиме реального времени и могут принимать решения по кредитам гораздо быстрее, отмечали в McKinsey. В Сбербанке, например, еще в 2019 году ИИ принимал решение о выдаче 100% кредитных карт, более 90% потребкредитов и свыше 50% ипотечных кредитов, говорил зампред правления Сбербанка Александр Ведяхин.

Балльные танцы

Когда в начале прошлого века методика прогнозирования рисков только появилась, сотрудники банков



ном заемщике может быть и собственный опыт банка во взаимодействии с ним: некоторые банки с помощью нейросетей анализируют тематики и тональности коммуникаций с клиентами.

Вес имеет практически вся известная информация о заемщике, которую по закону может хранить кредитное бюро, поэтому полный алгоритм расчета кредитного скорингового балла никто из игроков на рынке не раскрывает, отметили в АО «Банк „Русский стандарт“». По их словам, ключевым показателем при расчете скорингового балла является платежная дисциплина заемщика, то есть насколько он аккуратно вносил платежи по предыдущему или текущему кредиту.

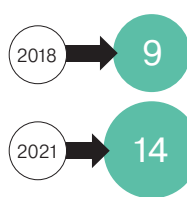
Однако скоринговый балл рассчитывают и для клиентов, которые никогда не брали кредит, на основании таких показателей, как доход, пол, возраст, количество детей, наличие семьи, город проживания и других характеристик клиента, которые отнесут его к конкретной сегментной группе. Далее анализируется платежная дисциплина подобной группы: «Например, женщины статистически лучше платят, чем мужчины, с возрастом возрастает ответственность заемщиков при погашении кредитов», — продолжают в банке «Русский стандарт».

На скоринговый балл также влияют доход человека и информация о том, как часто человек меняет места работы, как долго он работает на своем месте, добавляют в банке. Для принятия решения банки использу-

ют кредитный скоринговый балл, который получили от бюро, а также свою информацию о заемщике, поэтому скоринговый балл у каждого банка может отличаться от скорингового балла любого бюро, добавляют в банке.

Газпромбанк в дополнение к классическому скорингу внедряет более сложные модели прогнозирования, построенные на большом наборе данных по историческим заявкам с помощью технологий искусственного интеллекта и аналитики Big Data, рассказали в банке. По их словам, количество параметров при обучении моделей ИИ по каждому клиенту на два порядка больше, чем в классических моделях, что позволяет моделям при обучении выявлять более сложные и нелинейные закономерности в данных. «Мы видим, что за счет использования более сложных и точных по качеству алгоритмов ИИ, которые обучаются на больших наборах данных, можно как увеличивать портфель выдач по

ИНВЕСТИЦИИ ФИНАНСОВОГО СЕКТОРА В БОЛЬШИЕ ДАННЫЕ (\$ МЛРД)



ИСТОЧНИК: SNS TELECOM & IT.

ИНВЕСТИЦИИ ФИНАНСОВОГО СЕКТОРА В БОЛЬШИЕ ДАННЫЕ В РАЗВИКЕ ПО ТИПАМ ПРЕДОСТАВЛЯЕМЫХ ФИНАНСОВЫХ УСЛУГ (%)



ИСТОЧНИК: ВСЕМИРНЫЙ БАНК.

кредитным продуктам, так и уменьшать риски по ним. Это дает значительный экономический эффект для банка», — сообщили в Газпромбанке.

Цена ошибки

Однако доступных для оценки рисков данных не всегда бывает достаточно, чтобы банк мог принять правильное решение. Ошибки дорого обходятся финансовым организациям: объем просроченных кредитов физлиц в России в 2021 году превысил 1 трлн руб., проблемным оказался каждый десятый кредит.

Снизить вероятность ошибки позволяет концепция Data Fusion — подход, при котором объединяются данные из максимально возможного числа источников. Например, для оценки потенциального клиента можно привлечь данные в том числе мобильных операторов, ритейлеров, интернет-провайдеров. Подобный принцип лежит в основе скоринг-платформ, которые предлагают банкам свои базы для скоринга клиентов.

На основе данных интернет-трафика физлица можно выявить, например, интерес потенциального клиента к отелям в Южной Америке, дорогим шубам или конному спорту: «Наш опыт показывает, что это свидетельствует о низком риске просрочки по кредиту, тогда как посещение сайтов микрофинансовых организаций — о нестабильном финансовом положении», — рассказывают в разработчике скоринг-сервиса Platforma. По опыту компании при составлении скоринг-портрета юридических лиц могут учитываться такие факторы, как то, что организация работает с клиентами, но на ее телефоне нет ни входящих, ни исходящих звонков: «Это может быть фирма-однодневка с формальным телефоном». Также важен трафик сайта компании: «Интернет-магазин с десятком уникальных посетителей в день вряд ли можно считать перспективным бизнес-партнером», — добавляют в Platforma.

Использование данных из различ-

ных источников увеличивает точность скоринга как минимум на 10% и приносит миллионную прибыль, утверждают в Platforma. Так, качество работы скоринг-моделей оценивается с помощью коэффициента Джини — он показывает, насколько точно модель определяет принадлежность клиента к одной из групп по степени риска. Чем ближе показатель к единице или к 100%, тем сильнее распределяющая способность модели. Когда банк—клиент компании начал использовать сервис для скоринга Platforma, агрегирующий данные множества партнеров, его коэффициент Джини достиг 0,41. Банк стал одобрять на 10% больше заявок на кредиты, хотя их совокупное количество осталось на уровне 20 тыс. заявлений в месяц. В результате за счет более точной классификации клиентов бизнес увеличил прибыль на десятки миллионов рублей в месяц.

Платформенный подход

Когда финансовая организация обращается к скоринг-сервису, он в первую очередь ищет соответствия между данными заказчика и собственными. Скоринг-сервис Platforma, например, содержит обезличенные данные более 90 млн физических и 5 млн корпоративных клиентов. В получившейся выборке формируются переменные, которые содержат кредитные данные, информацию о финансовом статусе и социальных характеристиках клиентов. На их основе сервис сможет классифицировать клиентов по уровню дохода и оценить, кто из них выплатит кредит. Для оценки строится математическая модель — это может быть как уравнение, так и алгоритмы машинного обучения, если речь идет о десятке различных характеристик. В результате создается скоринговая карта, в которой каждой характеристике присвоен балл — чем он выше, тем значимее характеристика.

Подобная методика подходит не только для банков, желающих оценить кредитоспособность потенциальных клиентов, но и для страховых компаний, которым нужен прогноз вероятности наступления страхового случая, для транспортных компаний и интернет-магазинов, которые хотят проверить надежность партнеров. Также скоринг применяется при проведении государственных тендеров: он позволяет отсеять фирмы-однодневки и неблагонадежных участников на этапе сбора заявок. Кроме того, скоринг-модели могут дать рекомендации по общению с клиентами, например вычислить, в какое время суток им лучше звонить.

Юлия Степанова

Data Fusion: огромные перспективы больших данных

— экспертное мнение —

Вице-президент, заместитель руководителя департамента анализа данных и моделирования банка ВТБ СЕРГЕЙ ГОЛИЦИН — о концепции Data Fusion.

Понятие Data Fusion — сквозное объединение данных — появилось в аэрокосмической отрасли и робототехнике и первоначально подразумевало алгоритмы, которые объединяли и обрабатывали данные от различных датчиков для автоматического принятия наиболее эффективного решения. Теперь под Data Fusion мы понимаем не только одновременное использование разных и косвенно связанных массивов больших данных, но и применение для их обработки сразу нескольких технологий. Например, машинное обучение и георазметка данных или машинное обучение и графовая аналитика.



Сегодня бизнес-стратегии многих компаний, начиная от самых мелких и заканчивая транснациональными корпорациями из самых разных отраслей экономики, осно-

ваны на концепции Data Fusion. В частности, в финансовой сфере ее считают естественным шагом в развитии Data Science. Работа с большими данными — это минимум, без которого уже невозможно создавать современные конкурентоспособные продукты и сервисы, условие устойчивого развития бизнеса.

Данные применяются прежде всего для улучшения клиентского сервиса, управления рисками, персонализации предложения и противодействия мошенничеству. Однако компаниям мало только собственных данных, и без кросс-индустриальных данных о клиенте они рискуют потерять позиции. Концепция Data Fusion — это новое направление работы с данными, которое значительно расширяет знания бизнеса о потребителе.

Большие данные являются источником инсайтов о клиентском по-

ведении и позволяют вычленять закономерности и предпочтения разных категорий потребителей. Именно аналитика больших данных дает возможность предлагать клиентам персонализированные продукты или выстраивать таргетированные коммуникации. Например, в ритейле большие данные обеспечивают эффективную работу рекомендательных систем, а данные о геолокации, взятые у телеком-операторов, позволяют оптимизировать маршруты общественного транспорта.

На сегодняшний день все успешные финансовые организации используют в своей работе большие данные. Во многом именно от этого зависит конкурентоспособность организации на рынке. Благодаря Big Data и алгоритмам искусственного интеллекта в банковской сфере произошел важный поворот от продук-

тоцентричного к клиентоцентричному подходу: для повышения эффективности продаж услуги должны быть ненавязчивыми, удобными. Среди направлений, в которых банк ВТБ применяет объединение данных прямо сейчас, — кредитный скоринг, прогнозирование рисков, разработка новых продуктов, геоналитические сервисы для эффективного развития бизнеса и выстраивание маркетинговых коммуникаций на основе более узкого сегментирования клиентов по поведенческим группам.

Вопреки заблуждениям о том, что сбор персональных данных финансовых организациями ущемляет безопасность, аналитика больших данных помогает пресекать мошенничество. К тому же Big Data — это способ банка обезопасить себя и своих клиентов от чрезымерных рисков. Например, геонали-

тические модели, разрабатываемые банком на основании объединения внутренних и внешних данных о финансовой активности клиентов, позволяют компаниям более эффективно открывать точки продаж, одновременно минимизируя бизнес-риски и повышая свою доступность для потенциальных покупателей.

В настоящий момент потенциал уже собранных данных раскрыт не полностью. Данные всегда несут большой синергетический потенциал, однако ни одна из отраслей не способна создать полную картину цифровых следов своих клиентов без синергии с другими отраслями. Для того чтобы синергия данных несла новые возможности, а не риски для клиентов, она должна базироваться на качественно новых современных технологических концепциях, таких как Data Fusion.

Узнать клиента из миллиона

— реклама —

На основе такого идентификатора, например, уже действует рекламный сервис Bridge, созданный компаниями Redllama и Platforma. Технология разработана с помощью собственных моделей Platforma и строится на обработке агрегированных и потоковых данных партнеров: интернет-провайдеров, мобильных операторов и поставщиков рекламного инвентаря. Bridge обеспечивает связь арри веб-трафика и его обогащение единым идентификатором пользователя, который включает, например, знания об интересах, возможностях, потребностях и т.д. Это позволяет собирать кастомизированные аудиторные сегменты.

Барьеров стало больше

Текущий кризис на рекламном рынке добавил новые вызовы для специа-

листов, занимающихся развитием таких технологий в нашей стране. Директор по аналитике и информационным продуктам People & Screens (входит в группу dentsu Russia) Александр Горбачев говорит о том, что в целом в агентстве видят интерес к решениям на основе Big Data, которые позволяют планировать рекламу в условиях турбулентности, но многие компании взяли паузу в проектах, связанных с большими данными, из-за необходимости «приземления» облачной инфраструктуры в России.

«В частности, много аналитических систем было завязано на Google BigQuery, многие строили визуализацию в Tableau и PowerBI. Сейчас доступ к этим инструментам ограничен, а многим приходится еще и отказываться от Google Analytics 360. Поэтому для эффективного использования больших данных сейчас надо пересобрать системы сквозной аналитики», — рассуждает Алек-

сандр Горбачев. Сокращение бюджетов на маркетинг также бьет в первую очередь по инвестициям в исследования и дорогостоящим проектам с использованием больших данных, продолжает он: «Мы видим увеличение барьеров к использованию продвинутых технологий, бюджеты фокусируются на более консервативных инструментах с понятной простой механикой».

«Big Data — это для жирных времен, когда есть деньги, в противном случае это случайные всплески. А в кризис в будущее вкладывают единицы. Есть интерес, а есть платежеспособный спрос. В этом году интерес к теме Big Data продолжает расти, а платежеспособный спрос падает», — также отмечает директор диджитал-дивизиона группы «Игро-никс» Евгений Чувашов. С марта интерес перераспределился в сторону поиска решений у локальных игроков, так как часть поставщиков ушли

с рынка РФ или доступ к их продуктам ушел в серую зону, добавляет он.

Без технологий нет эффективности

Тем не менее, несмотря на новые трудности, анализ больших данных в рекламе в России будет развиваться, прогнозируют эксперты. Даже сейчас есть некоторый рост спроса на него, говорит директор по данным и технологиям ГК «Родная речь» Александр Феоктистов. «Связано это в первую очередь с тем, что с российского рынка ушли крупные рекламные площадки, которые предоставляли большое количество инструментов. Сейчас клиенты замещают инвентарь локальными, но у них не всегда есть нужные инструменты, поэтому часть задач приходится решать самим клиентам или рекламным агентствам», — говорит он.

Будет расти спрос на любые инструменты, направленные на повы-

шение эффективности рекламных компаний и рост продаж, считает господин Феоктистов. «Ситуация с ограничениями в области персональных данных будет влиять на собираемость данных, их полноту и объемы, поэтому инструменты уже меняются», — подчеркивает эксперт. — Но рекламные технологии продолжат развиваться, и такие сервисы будут востребованы все больше. Рекламодатель и агентства всегда будут стараться достичь большей эффективности закупок, а без технологий это невозможно».

Продукты на Big Data уже сейчас отвечают на важнейшие вопросы об эффективности рекламы и дают совершенно новые возможности по кроссканальной атрибуции и управлению рекламой, отмечает руководитель по развитию рекламных сервисов Platforma Анатолий Новожилов. По его мнению, аудиторные сегменты для таргетиро-

ванной рекламы на данных банков и телеком-операторов показывают более высокую эффективность по сравнению с традиционными социально-демографическими сегментами, а сквозная аналитика позволяет оценивать влияние ТВ-рекламы на продажи в различных сетях и интернете, оптимизировать выбор медиаканалов, медиавеса и инвестиции в рекламу. «StableID дает возможность строить сценарии коммуникации в интернете, ТВ и цифровой наружной рекламе с привязкой к конкретному человеку. И это только начало», — говорит он. — Бурное развитие будет с появлением крупных индустриальных игроков, которые предоставят доступные инструменты для рынка, не требующие развития сложной технологической инфраструктуры на стороне рекламодателей», — прогнозирует господин Новожилов.

Валерия Лебедева

Большие данные как повод для партнерства

Объем данных растет с каждым годом: к 2025 году в мире будет насчитываться 175 ЗБ данных, прогнозировали в IDC в 2019 году. Вместе с этим растет и ее ценность: анализ Big Data позволяет компаниям выстраивать стратегии работы, оптимизировать распределение бюджетов и усилий. В тренде сегодня совместная работа компаний с данными, то есть объединение клиентских баз для анализа рынка и создания партнерств. Однако необходимость защиты персональных данных, а также нежелание делиться информацией с конкурентами становятся для подобных партнерств настоящим препятствием. Так на рынке сформировался запрос на технологии конфиденциальных вычислений, которые облегчают совместную работу с Big Data. В России эти технологии используют как отдельные компании, так и платформы обмена данными.

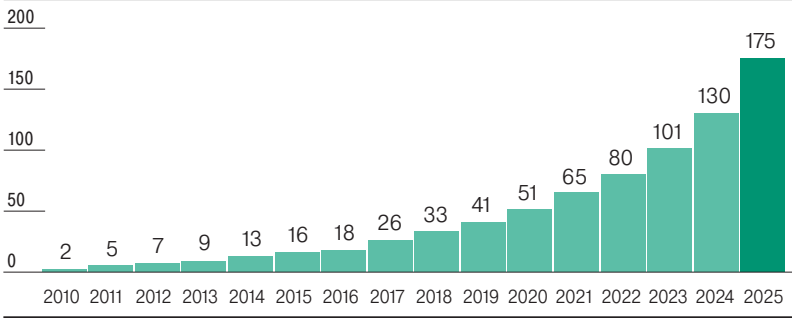
— перспективные технологии —

Большие заработки на больших данных

Работа с большими данными с каждым годом сулит бизнесу все большие заработки: в 2021 году объем мирового рынка Big Data и бизнес-аналитики составил \$215,7 млрд, тогда как в 2018-м он оценивался в \$168,8 млрд, следует из оценок Statista. По данным MarketsandMarkets, глобальные расходы компаний и госучреждений на работу с инструментами Big Data по итогам 2021 года достигли \$162,6 млрд. В России же, по оценке Ассоциации больших данных, рынок Big Data составлял от 10 млрд до 30 млрд руб. в 2019 году с перспективой роста до 300 млрд руб. к 2024 году.

КАК ВЫРАСТЕТ ОБЩЕМИРОВОЙ ОБЪЕМ ГЕНЕРИРУЕМЫХ ДАННЫХ (В ЗЕТТАБАЙТАХ; 1^о ЗБ — 10²¹ (СЕКСТИЛЛИОН) БАЙТ)

ИСТОЧНИК: IDC И SEAGATE.



Однако одним из камней преткновения для бизнеса на пути использования больших данных является законодательство по защите персональных данных. В Европе в рамках GDPR (General Data Protection Regulation) действуют оборотные штрафы за утечку информации, в результате чего ответственность компании, допустившей утечку, может достигать миллионов евро. В России обсуждается возможность введения аналогичной системы: Минцифры согласовало законопроект, предполагающий введение штрафа за утечку персональных данных в размере от 1% до 3% от годовой выручки компании (см. „Ъ“ от 30 мая).

«Помимо работы с развивающимся международным законодательством о конфиденциальности и защите данных IT-директора должны избегать любой потери доверия клиентов в результате инцидентов, связанных с конфиденциальностью», — подчеркивали аналитики Gartner в отчете, определяющем тенденции в сфере технологий в 2022 году. Одним из стратегически значимых трендов, по мнению Gartner, является работа с конфиденциальными вычислениями.

Аналитики ожидают, что к 2025 году 60% крупных организаций будут применять вычисления, повышающие конфиденциальность.

Что такое конфиденциальные вычисления

Конфиденциальные вычисления — это технология, которая облегчает совместную работу с Big Data и позволяет компаниям обмениваться информацией, не открывая прямого доступа к цифрам. Например, компании хотят выяснить объем общего рынка, но не хотят раскрывать друг другу свою выручку. В таком случае им потребуется третий участник, так как в паре любой результат будет обратимым. Такую возможность дает технология конфиденциальных

вычислений. Она представляет собой прямую связь между приложением и процессором, которые образуют доверенную систему, недоступную для других программ и систем. Большую ценность такие вычисления создают, к примеру, в сфере здравоохранения, которая оперирует наиболее чувствительными данными. Например, в начале пандемии IT-компании разрабатывали экспресс-системы диагностики COVID-19 на основе искусственного интеллекта, однако медицинские организации не имели права пере-

Мэтчинг с партнерами — следующий шаг для развития рынка работы с большими данными

давать сторонним компаниям снимки пациентов, необходимые для машинного анализа. Эту проблему можно было бы решить с помощью технологии конфиденциальных вычислений Intel Software Guard Extensions (Intel SGX), которая появилась в 2020 году и является наиболее зрелой на данный момент.



Развитие технологий конфиденциальных вычислений усложняет то, что нейронные сети требуют использования очень мощных серверов, а для построения модели необходимо большое количество времени. «Сложность в том, что построение моделей — интерактивный процесс. Часто приходится менять данные или исходные параметры анализа», — объясняет директор по анализу данных и моделированию Platforma Сергей Голицын. — А если модель обучается неделю, то даже простейший анализ может растянуться на месяцы». Однако, по его словам, уже есть стартапы, которые за три года существования этой технологии смогли оптимизировать ее до уровня, при котором модель обучается за один день.

Конфиденциальные вычисления реализуются при помощи процессоров, которые поддерживают этот протокол защиты, но из-за ограничения импорта в РФ подобные процессоры ограничено доступны, отмечает эксперт по информационной безопасности Ахених Евгений Качуров. По его мнению, чтобы исправить ситуацию, необходимо перейти на отечественные процессоры: «На фоне широкого распространения российских процессоров производители смогут инвестировать в развитие технологий, что приведет к распространению технологии в РФ».

Российский опыт

Конфиденциальные вычисления, по сути, новейшая технология для российского рынка: она появилась в фокусе отечественных интеграторов и производителей аппарат-

ной видят риски с точки зрения информационной безопасности, говорит господин Качуров.

Применение технологии позволило бы предотвратить масштабные утечки персональных данных, согласен ведущий инженер CorpSoft24 Михаил Сергеев. Однако, по его мнению, для небольших компаний ее работа «не очень рентабельна, так как расходы на внедрение и использование в разы больше, чем финансовые последствия от утечек».

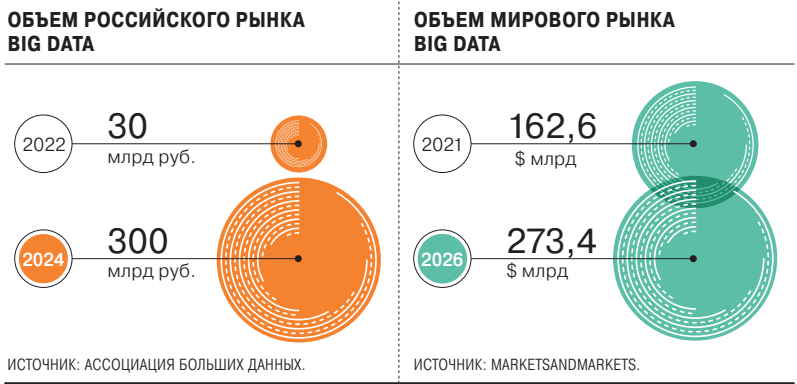
В России такие технологии более востребованы в сферах с повышенными регуляторными требованиями или строгими требованиями к конфиденциальности данных (финансовый сектор, ритейл, IT), рассуждает Евгений Качуров. Так, первую в России доверенную среду совместной работы с данными запустили в 2019 году ритейлер «Магнит» и компания по управлению цифровыми лицензиями Aggrepoint. Технология используется для управления цифровой рекламой: информация о клиентах «Магнита» проходит обработку с помощью Intel SGX, после чего персональные данные обезличиваются и с ними может работать софт рекламных групп — партнеров ритейлера, например сегментировать аудиторию по предпочтениям в выборе товаров. В результате «Магнит» на 20–30% сократил затраты, связанные с размещением рекламы в цифровых каналах. Ритейлер продолжает использовать технологию и работает над возможностью ее использования другими участниками проекта, рассказали в компании.

«Одной из самых востребованных и перспективных на рынке» считают технологию конфиденциальных вычислений и в ВТБ: в банке сейчас идет пилотный проект с ней. Однако существующие на рынке решения «не полностью отвечают» требованиям банка в части универсальности, скорости вычислений и обеспечения технологического суверенитета, поэтому ВТБ сейчас разрабатывает собственное решение.

Альфа-банк уже давно пробует конфиденциальные вычисления, для реализации которых нужны «весьма продвинутые технологии, а также надежные партнеры». «У банка есть

и то и другое, что позволяет уже сейчас получать первые значимые результаты, подтверждающие высокий потенциал этого подхода», — говорит руководитель департамента продвинутой аналитики Альфа-банка Алексей Каширин. По его мнению, для компаний, развивающих системы управления и процессы принятия решений на основе продвинутой аналитики, конфиденциальные вычисления — это «не просто перспективная опция, а практически безальтернативный путь».

В Газпромбанке для комплексного обеспечения инфорбезопасности автоматизированных систем технологию конфиденциальных вычислений пока не применяли, но не исключают такой возможности в будущих проектах, рассказали в пресс-службе.



ИСТОЧНИК: АССОЦИАЦИЯ БОЛЬШИХ ДАННЫХ.

ИСТОЧНИК: MARKETSANDMARKETS.

Кроме финансового сектора к вычислениям такого рода пристально присматривается и телеком. В 2021 году «Билайн Бизнес» (структурное подразделение ПАО «Вымпелком») начал создание децентрализованной платформы для обработки клиентских данных, которая позволит повысить точность подбора аудиторий для маркетинговых активностей и масштабировать текущий функционал. Предполагается, что на платформах будут размещаться данные крупнейших организаций, в том числе ритейлеров и фармкомпаний, которые участники смогут монетизировать, предоставляя доступ друг другу и создавая новые партнерства. В целом мэтчинг с партнерами — «следующий шаг для развития рынка работы с большими данными, так как крупные игроки уже научились работать с данными и получать от них

мой для обмена данными, которая обеспечивает безопасные условия и инфраструктуру. Например, базу для подобной платформы разработали компании Platforma и HFLabs. Технологию впервые протестировали в июле текущего года один из крупнейших в стране банков и ведущий телеком-оператор: компании объединили имеющиеся у них сведения, состоящие более чем из 250 млн записей, и выявили количество клиентов, которые пользуются услугами обеих организаций.

В перспективе платформа сможет стать банком данных, где партнеры хранят свои данные без доступа со стороны других участников, но с возможностью работать с базами, проводить анализ и создавать совместные сервисы и бизнес-продукты, рассчитывают в компании.

Юлия Степанова

«Многие слышали о конфиденциальных вычислениях, но почти никто не видел»

— экспертное мнение —

Технический директор HFLabs НИКИТА НАЗАРОВ — о проблемах с интерпретацией термина и выборе партнеров в работе с новой технологией.

Рынок конфиденциальных вычислений только формируется, но уже есть сложности с интерпретацией самого термина. Ситуация с конфиденциальными вычислениями в России можно сравнить с лохнесским чудовищем: многие о них слышали, но никто (или почти никто) не видел. Более того, за этим термином скрываются сразу три разных понятия: протокол конфиденциальных вычислений, аппаратная платформа и федеративное обучение.

Протокол конфиденциальных вычислений — это математика



найти средний доход клиента) без раскрытия входных данных каждого из них. Однако в разговорах о конфиденциальных вычислениях в 95% случаев речь идет не о протоколе, а об использовании технологий аппаратной защиты, например Intel SGX. В этом случае производитель предоставляет защищенную аппаратную среду, которая гарантирует безопасность загружаемых в нее данных. На практике это позволяет владельцам облачной инфраструктуры общаться своим клиентам полную защиту данных. И в этом аспекте уход с российского рынка Intel — однозначно грустный факт.

Наконец, иногда под конфиденциальными вычислениями подразумевают федеративное обучение. В общих чертах это процесс, который использует протокол конфи-

денциальных вычислений для обучения нейронных сетей. Естественно, что в этом случае также не происходит раскрытия данных источников. Отмечу, что ни для федеративного обучения, ни для реализации обмена с помощью протокола конфиденциальных вычислений технологии вроде Intel SGX применять необязательно.

Путаница в понятиях усугубляется еще и тем, что нередко в маркетинговых целях термином «конфиденциальные вычисления» называют то, что вообще к ним не имеет отношения. Поймать за руку в данном случае сложно, так как нужно иметь доступ не только к описанию технологии, но и к реализации конкретного алгоритма. Информация на сайтах компаний, представляющих платформы для конфиденциальных

вычислений, может изобиловать математическими деталями или, напротив, быть максимально абстрактной (раз — и магия!), поэтому всем, кто хотел бы использовать эту технологию, я бы рекомендовал с максимальной тщательностью подбирать партнеров.

Что касается интереса рынка к конфиденциальным вычислениям, то в первую очередь он направлен на федеративное обучение. Крупные компании собрали огромные массивы данных о своих клиентах и теперь хотят обмениваться ими друг с другом. Информация от друзей игроков позволяет выбрать из собственной песочницы и, например, учитывать сведения партнеров в скоринговых моделях. При этом важно защитить клиентскую информацию так, чтобы избежать утечки

данных и не нарушить закон. И если посмотреть на конфиденциальные вычисления с этой точки зрения, то они, безусловно, перспективны, так как сама технология обеспечивает безопасность обмена.

Другое дело, что штрафы за утечки персональных данных в последнее время не превышают 60 тыс. руб., тем не менее наметился отчетливый тренд по ужесточению ответственности в таких ситуациях — Минцифры уже разрабатывает законопроект, по которому наказание может составлять до 1–3% от оборота. На этом фоне конфиденциальные вычисления, несмотря на сложность в плане организации процесса и более длительное, чем при классических методах, обучение скоринговых моделей, становятся все более актуальными.