

02 февраля 2022

Как перестать устанавливать неверные цены на товары и опираться на неверные данные

В науке о данных есть принцип GIGO — garbage in, garbage out, мусор на входе — мусор на выходе. Он объясняет, что неправильные данные, используемые в работе, формируют ошибочный результат. Верные мастер-данные — гарантия точного результата. О том, какие данные необходимо использовать для получения верных рекомендаций по оптимизации ценообразования, рассказывает Павел Серебренников, эксперт департамента аналитических решений ГК «КОРУС Консалтинг».

В ритейле один из наиболее прибыльных способов использования машинного обучения – оптимизация ценообразования. Оно позволяет предлагать цены на товары на основе рекомендаций искусственного интеллекта, который эффективнее человека учитывает влияющие на спрос факторы, и благодаря этому помогает достигать поставленных целей по повышению финансовой эффективности бизнеса.

Ценообразование, использующее в своей работе алгоритмы на базе машинного обучения (ML), называется оптимизационным ценообразованием (Price Optimizaion).



Качество работы алгоритмов машинного обучения напрямую зависит от качества данных, которыми они оперируют. Для расчета оптимальной цены системе на входе нужна верная информация — в противном случае есть риск некорректного результата. Например, если алгоритм опирался на ошибочные данные по ценовому мониторингу конкурентов, то сформированные им рекомендации грозят недополучением прибыли или даже ухудшением ценового восприятия у клиентов.

Три кита ценообразования: какие данные важны

Есть три вида данных, которые необходимо собирать для работы систем. Их использование в полном объеме повышает шанс найти и установить взаимосвязи между факторами, которые влияют на спрос и, в итоге, выполнить оптимальный расчет цен.

■ Исторические данные

Первый и критически важный вид данных – история продаж (транзакций на кассах) за период не менее, чем два года. Они обязательно должны содержать сведения об уникальном номере чека, дату продажи, ID товара, ID магазина, количество купленного товара и фактическую цену покупки. Желательно, чтобы в чеке присутствовали сведения и о величине скидки.

Для оптимизационного ценообразования исторические данные крайне важны, так как именно на их основе моделируются и рассчитываются основные факторы и показатели, необходимые для работы алгоритма. Например – эластичность спроса или прогноз продаж. В свою очередь на

основании этой информации алгоритм определяет цену, которая будет соответствовать указанным целям и стратегии компании.

Естественно, у каждого бизнеса есть свои особенности учетной политики. Поэтому далеко не всегда эти данные могут быть полными.

Второй по очереди, но не по важности вид исторических данных – история изменения регулярных цен товаров на полке за аналогичный период. В сочетании с историческими данными по продажам, она позволяет оценить, при какой цене покупатель отказался от покупки, посчитав ее слишком дорогой.

Также важны данные по остаткам товаров и их себестоимости на каждый день периода. Благодаря этим данным алгоритм определит, что по одним продуктам не было продаж, т.к. их просто не было на остатках магазина, а по другим снижение или отсутствие продаж обусловлено другими факторами. Например, ценой. А данные по себестоимости продуктов алгоритм будет учитывать в расчете конечной цены продукта на полке.

Еще один важный вид исторических данных – информация о проведенных промоакциях. Она используется алгоритмом для определения степени влияния промо на спрос и достижение целевых показателей, а также для расчета и рекомендаций оптимальных промо-цен. Данные должны содержать календарь проведения акций с привязкой к конкретному товару.

Как повысить качество таких данных? Прежде, чем использовать их в работе алгоритмов, важно проводить дополнительные исследования данных на наличие аномалий и проверять на достоверность, качество и полноту дата-

сета.. Верны ли выгруженные данные? Есть ли в них вся необходимая информация? Если есть ошибки, необходимо выстраивать систему верификации. Обычно в основе таких практик лежат определенные сценарии проверок. Реже – решения, использующие в своей работе ML-алгоритмы. Такие системы мониторят, например, полноту передачи данных с кассовых терминалов, появление пробелов данных по истории цен или промо. Реагируют на появление каких-либо логических ошибок. Например, что в учетной системе дата старта акции наступает позже даты его окончания и т.д.

Проведение бизнесом мониторинга качества данных позволяет не только быть уверенным, что их качество не повлияло на правильность принятия того или иного бизнес-решения, но и позволяет предвосхитить появление системных ошибок в учете, которые могут сказаться на качестве работы сервисов и инструментов, которыми планирует пользоваться бизнес.

Если данных не хватает — искать способы извлечения недостающей информации или дополнения существующей базы из других источников. Например, остатки можно рассчитать на основании документов, в которых зафиксированы все поступления и списание продукции с учета, а историю цен часто возможно восстановить на основании данных по продажам.

■ **Справочники и характеристики объектов**

При формировании рекомендаций алгоритмам требуется максимум доступной информации, описывающей товары и их характеристики. Среди них — справочники номенклатуры, категорий, брендов, магазинов, данные о продуктах и точках продаж.

К характеристикам продукта, например, можно отнести цвет, вес, материал, из которого он изготовлен, срок годности, кратность отпускной тары и тому подобное. К данным о магазинах — величину их торговых площадей, количество касс, посетителей, наличие или отсутствие парковки и её площадь. Оказать влияние может даже расстояние до ближайшей станции метро.

К сожалению, далеко не всегда компании располагают всеми этими данными. Это может быть связано как с особенностями учетной политики, так и со сложившейся ИТ-архитектурой. Но всегда необходимо помнить: чем больше видов и типов данных использует алгоритм, тем больше будет установлено факторов и их взаимосвязей, влияющих на выбор и поведение покупателя.

Что делать, если данных не хватает? Всегда есть вероятность, что необходимую информацию можно извлечь и очистить из уже имеющихся дата-сетов, получить из внешних источников, запросить у поставщиков или дать сотрудникам задачу на формирование нового справочника.

■ **Информация о сопутствующих факторах**

К такому типу данных относятся сведения о факторах, прямо или опосредованно влияющих на решение покупателя о покупке продукта. К этому же типу данных можно отнести те факторы, которые учитываются бизнесом при расчете оптимальной цены.

В первую очередь, о данных по ценовому мониторингу конкурентов. Не все компании сами собирают эту информацию — чаще всего для получения требуется обращение к внешним источникам – компаниям и сервисам,

предоставляющим услуги по проведению мониторинга цен как интернет-сайтов конкурентов, так и цен на полках непосредственно в самих торговых точках. Также стоит учитывать сведения о запуске программ кредитования покупателей. Немаловажны фактор — изменение политики мотивации персонала: личная презентация консультанта и помощь в поиске товара также влияет на совершение покупки.

Дополнительными влияющими факторами могут выступать такие сведения, как статистика просмотра карточек товара на сайте, его рейтинг, данные о затраченных средствах на маркетинг и способы продвижения продаж, себестоимость товаров в пути и тому подобное. У каждого ритейлера есть свои особенности, из-за которых список используемых факторов и показателей может быть расширен.

Окончательный перечень факторов, которые будут имплементированы в работу алгоритма, определяется исходя из особенностей бизнеса заказчика и по итогам проведения бизнес-аудита, дата-аудита и после проведения масштабных работ по исследованию имеющихся данных.

Как прийти к качественному результату

Для успешного применения алгоритмов важно не просто собирать данные, но и следить за их качеством. Чтобы построенные прогнозы были верны, информация на входе должна отвечать как минимум следующим критериям:

- 1. Полнота.** В данных не должно быть пропусков.

2. Ссылочная (индексная) целостность. (Полная взаимосвязь между всеми объектами, присутствующими в данных)
3. Корректность. Источники данных должны обладать высоким уровнем доверия.

Получить на 100% качественные данные почти невозможно — тем более, если они охватывают довольно большой исторический период. Опыт внедрения подобных систем показывает, что, например, в данных о продажах (по транзакциям на кассах) допускается отклонение не более, чем +/- 3% за весь период — иначе есть риск снижения качества работы алгоритма.

Поэтому важно не только собирать данные, но и предварительно их проверять. Соответствуют ли они вышеописанным требованиям, нет ли в них аномалий, выбросов, логических нестыковок, ошибок учета, дублей или потерянных кусков. Такие проверки осуществляются специализированными инструментами, в том числе и при помощи различных ML-алгоритмов. В своих проектах мы проверяем более чем 400 различных показателей, чтобы убедиться, что данные соответствуют требованиям.

ML-алгоритмы не допускают ошибок, они лишь опираются на имеющиеся данные. ML — это «волшебная черная коробка», результат которой напрямую зависит от того, какие именно данные ты в нее вложишь.

Обучать алгоритм можно лишь после того, как вы получите полный объем верных данных из надежных источников. Только так можно быть уверенным, что рекомендации алгоритма будут полезны и не навредят вашему бизнесу.