

ПРОГНОЗИРОВАНИЕ СОВОКУПНОГО ЭФФЕКТА ПРОМО-АКЦИЙ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

FORECASTING THE CUMULATIVE EFFECT OF PROMOTIONAL CAMPAIGNS USING MACHINE LEARNING METHODS

A. Bystrov

Summary. This work aims to improve the accuracy of forecasting the final promotional effect in the fast-food industry. The study proposes an innovative approach that involves isolating the sales increase driven by promotional campaigns into a separate time series and subsequently modeling this value using machine learning methods. To solve this problem, the gradient boosting algorithm (CatBoost) is employed, which effectively handles categorical and numerical features, including discount depth, promotion mechanics, product characteristics, and regional specifics. The methodology includes constructing a base forecast using exponential smoothing to estimate «non-promo» sales, followed by determining the difference between actual sales and the baseline. This difference serves as the target variable for training the model, allowing the avoidance of information loss characteristic of traditional time series models. The experimental part, based on data from 760 unique promotions, demonstrates that the proposed approach leads to an average improvement of 8 % in the 1 — WAPE metric compared to the classical approach. The results confirm the possibility of more accurate demand forecasting, contributing to optimized inventory management and enhanced operational efficiency.

Keywords: demand forecasting, promotions, gradient boosting, promotional effect, fast-food industry.

Быстров Александр Игоревич

Аспирант, ФГБОУВО «Российская академия
народного хозяйства и государственной службы
при Президенте Российской Федерации», г. Москва
alexandr.jri.bystrov@yandex.ru

Аннотация. В данной работе ставится цель повышения точности прогнозирования итогового промо-эффекта в индустрии быстрого питания. Исследование предлагает инновационный подход, заключающийся в выделении прироста продаж, обусловленного промо-акциями, в отдельную временную последовательность и последующем моделировании этой величины с помощью методов машинного обучения. Для решения задачи используется алгоритм градиентного бустинга (CatBoost), который позволяет эффективно работать с категориальными и числовыми признаками, включая глубину скидки, механику акции, характеристики продукта и региональные особенности. Методология включает построение базового прогноза с использованием экспоненциального сглаживания для оценки «безпромо» продаж, а затем — определение разницы между фактическими продажами и базовой линией. Полученная разность служит целевой переменной для обучения модели, что позволяет избежать информационных потерь, характерных для традиционных моделей временных рядов. Экспериментальная часть, основанная на данных 760 уникальных акций, демонстрирует, что предложенный подход приводит к улучшению метрики 1 — WAPE в среднем на 8 % по сравнению с классическим подходом. Результаты подтверждают возможность более точного прогнозирования спроса, что способствует оптимизации управления запасами и повышению операционной эффективности.

Ключевые слова: прогнозирование спроса, промо-акции, градиентный бустинг, промо-эффект, индустрия быстрого питания.

Введение

В сфере быстрого питания (Quick Service Restaurant, QSR) промо-акции занимают центральное место в маркетинговой стратегии. Управленческая практика показывает, что мгновенный прирост объемов продаж при запуске акции может вводить в заблуждение классические модели временных рядов, так как всплески и «провалы» после акций искажают процесс обучения. Из-за этого существенно повышаются ошибки при попытке планировать выручку и объемы закупки.

Настоящая статья продолжает логику, изложенную в предыдущей работе, где было показано, что устранение промо-периодов при обучении «базового» ряда и выделение «промо-остатка» как отдельной переменной повышает точность регулярных прогнозов. Здесь производится следующий шаг: рассматривается задача

прогнозирования итогового (суммарного) промо-эффекта с помощью методов машинного обучения и показывается, что подобный подход точнее, чем классическое прогнозирование ежедневных продаж с учетом промо в одном общем процессе.

В данном эксперименте участвуют 760 уникальных акций, образующие 2322 уникальные пары «акция–товар», которые проходили в течение года.

Обзор литературы и методологических подходов

В ряде работ подчеркивается, что традиционный дневной подход не позволяет четко идентифицировать вклад промо-фактора. Это осложняет оценку реальной отдачи от кампании и корректное планирование запасов. Между тем, по мере проникновения машинных методов (Random Forest, градиентный бустинг и др.) в эко-

номические задачи, становится возможным отдельно моделировать эффект акции, что дает более высокую точность и гибкость.

В частности, описано, что машинное обучение (ML) в экономике имеет свою специфику (проблемы вычислительной эффективности, переобучения), но при корректном учете показателей оптимизации дает серьезные преимущества [1], а также находит все более широкое применение в разных отраслях [2]. Отмечается растущая востребованность ML-алгоритмов в цифровом ритейле и сфере быстрого питания [3]. Подобную эффективность градиентного бустинга при наличии расширенного набора факторов отмечали, в частности, при разработке маркетинговой стратегии российских розничных сетей [4].

Материалы и методы исследования

Сущность задачи и анализ признаков

Основная идея — выделить прирост продаж, обусловленный акцией, как отдельную целевую переменную. Если имеется «baseline»-прогноз (полученный, например, методом экспоненциального сглаживания без учета промо-дней), то можно вычислить (fact — baseline) и строить по этим значениям самостоятельную модель ML. При наличии детальных сведений о продукте, глубине скидки, механике акции (Discount, Set, Pairs, Coupon и т.д.), а также о канале реализации, региональной специфике и прочем, можно научить модель предсказывать итоговое отклонение от базовых продаж. Благодаря богатому набору признаков (features) градиентный бустинг на деревьях способен улавливать сложные нелинейные зависимости, которые неочевидны в традиционных авторегрессионных моделях временных рядов. Стоит отметить, что по сравнению с классическими предсказаниями по дням можно напрямую учитывать влияние каждой акции на товар, рассматривая ее как отдельную строку в обучающем наборе. В классической же схеме при одновременном действии нескольких акций на один и тот же товар приходится строить агрегированные признаки, несущие потерю в информации для модели.

Опираясь на сведения, собранные в процессе исторических промо-кампаний, было выделено несколько ключевых групп факторов:

- Ценовые факторы, включая абсолютную и процентную разницу между обычной и промо-стоимостью, а также совокупную стоимость набора (Set, где итоговая цена ниже суммы одиночных позиций).
- Характеристики продукта, которые учитывают группировку по ABC-категориям, тип продукта и его роль в меню.
- Сведения о продолжительности промо и механике акции (прямая скидка, купон на несколько то-

варов, программа «2+1», «подарочный» продукт и т.п.), а также возможность комбинировать несколько акций одновременно.

- Региональные и сезонные признаки, связанные с локацией ресторана (город, проходимость, расположение вблизи ТРЦ), временем года, праздничными периодами и т.д.

В имеющемся наборе механик преобладают прямые скидки, сетов и «пары». При этом для каждой механики важно корректно рассчитывать совокупную стоимость: если в случае прямых скидок цена уменьшается на заданную величину сразу, то при использовании сетов или «пар» итоговая скидка распределяется косвенно (через общую стоимость набора), что усложняет расчеты.

Для решения задачи применим алгоритм CatBoost — градиентный бустинг на деревьях, который хорошо работает с категориальными признаками и разнородными данными. Подобную эффективность CatBoost, особенно при наличии дополнительного набора факторов, отмечали и в сфере рекомендательных систем e-commerce [4].

На рисунке 1 показано распределение вкладов (SHAP values) ключевых факторов, влияющих на итоговый промо-эффект. Горизонтальная ось отражает относительную силу влияния каждого признака (слева направо: отрицательный вклад, справа налево: положительный вклад), а вертикальный список перечисляет факторы, сгруппированные по возрастанию или убыванию средней значимости. Каждая точка на диаграмме соответствует одной акции (или наблюдению) в выборке: насыщенность цвета (от фиолетового к красному) указывает на величину самого фактора. Анализ рисунка показывает, что самым существенным драйвером промо-эффекта для исследуемых данных являются `rmix_price` и `std_before_promo`: при высоких значениях этих признаков вклад в увеличение эффекта выражен сильнее (наблюдается облако точек справа от центральной вертикальной линии).

Заметим также значительную важность признаков, характеризующих промо-механику (например, `mechanics_name`) и тип продукта (`product_group`). Наличие разнонаправленного рассеяния точек указывает на то, что вклад большинства факторов зависит не только от их абсолютного значения, но и от взаимодействия с другими признаками (особенно с `mean_last_14_days` и `std_last_14_days`), что подтверждает целесообразность использования машинных методов (градиентного бустинга) вместо упрощенных линейных регрессий.

Дополнительно видно, что такие признаки, как `percent_discount` или `absolute_discount`, не всегда играют доминирующую роль при краткосрочных акциях (об этом говорит более низкая плотность «теплых» точек), однако при определенных условиях они могут су-

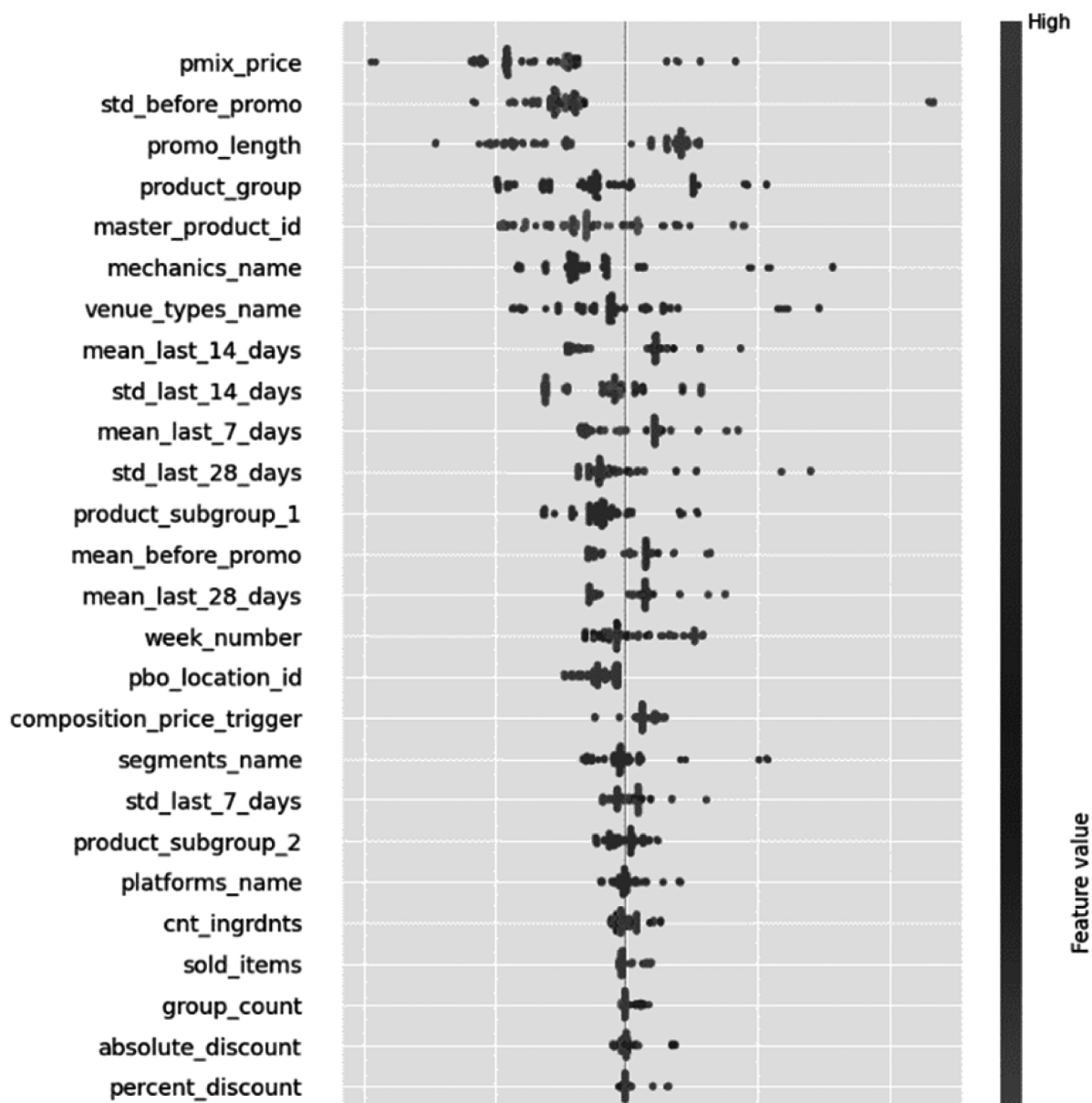


Рис. 1. Распределение вкладов (SHAP values) ключевых факторов, влияющих на итоговый промо-эффект

щественно повышать итоговый промо-эффект. Это согласуется с гипотезой о важности скидочных механик, но в то же время подчеркивает, что для полного объяснения поведения продаж необходимо учитывать и другие факторы — конкурентное окружение, сезонные колебания, исходные объемы продаж и т. д.

Методика и экспериментальная часть

Для обучения модели было сформировано множество записей вида «(акция, ресторан, продукт) → прирост». В каждой записи фиксировались целевая переменная (итоговый объем продаж, превышающий базовую линию) и набор факторов (признаки, описанные выше). Базовую линию (продажи без акции) получали, удаляя промо-периоды из обучающего ряда и используя

методы экспоненциального сглаживания по аналогии с предыдущей работой. Исторический фактический промо-эффект (прирост) определялся как разница между реальными продажами и данным базовым прогнозом.

Для валидации и тестирования данные разделялись таким образом, чтобы на тест не попадали фрагменты одной и той же акции (дабы исключить утечку информации). Настройка гиперпараметров CatBoost (глубина деревьев 6, количество итераций 2000, $\text{learning_rate}=0.005$) осуществлялась перебором по сетке. Как показывают кейсы российских розничных сетей [4], точная подстройка ML-моделей по специальной гиперпараметрической сетке дает максимальный прирост точности и минимизирует риск «недопрогноза», критичного в периоды всплеска.

Сравнение велось по двум ключевым сценариям. В первом («Baseline») применялся стандартный градиентный бустинг, где сведения о проводимых акциях учитывались через набор бинарных переменных, а прогноз строился на подневной основе — полностью для каждого дня недели. Во втором («Proposed») в модель вводились расширенные экзогенные факторы (такие как механика акции, глубина скидки и т. д.), а само обучение велось непосредственно на итоговом приросте (промо-эффекте).

Для оценки качества прогноза использовался показатель 1 — Weighted Absolute Percentage Error (WAPE). Данная имплементация (1-WAPE) обусловлена более доходчивым пониманием для бизнеса, четко показывая, что «чем больше — тем лучше». Данная метрика наиболее уместна при большом ассортименте, в котором одни товары продаются в тысячах единиц, а другие — в считанных десятках, причем стоимость товаров может незначительно отличаться. «Взвешенная» формула WAPE компенсирует различие масштабов продаж и позволяет корректнее сравнивать ошибки между позициями.

На основе одного исходного массива сформировали два датасета для обучения: первый (Baseline) представляет продажи в разрезе дней, тогда как второй (Proposed) строится сразу для итогового эффекта каждой акции. Хотя физический смысл многих признаков совпадает, в дневном подходе (Baseline) при параллельном проведении нескольких акций приходится использовать набор бинарных переменных, что усложняет структуру. В новом решении (Proposed) каждая акция считается отдельной сущностью, а пересечения нескольких промо в один день не вызывают конфликта признаков.

Тестовый период охватывает последние восемь недель датасета. В результате появились два обучающих набора: один для подневной модели (Baseline), второй для итогового эффекта (Proposed). Несмотря на схожий набор фичей, в первом случае (Baseline) пришлось вводить дополнительные бинарные столбцы, если в один день действовали несколько механик, тогда как во втором случае данная проблема решается архитектурно за счет акционной структуры.

Ниже приведена сводная таблица, демонстрирующая качество прогноза (в терминах 1 — WAPE, чем ближе

Таблица 1.

Качество прогноза для различных категорий товаров

Группа / Локация	Локация 1 (New → Baseline)	Локация 2 (New → Baseline)
A	72.3 % → 67.2 %	76.1 % → 70.0 %
B	74.2 % → 65.1 %	75.9 % → 68.1 %
C	64.0 % → 51.5 %	62.7 % → 52.7 %

к единице, тем лучше) для различных категорий товаров (A, B, C) в двух локациях.

Здесь товары разделены по перцентилям объема продаж следующим образом: группа A включает наиболее продаваемые позиции (выше 60-го перцентиля), группа B — позиции со средними объемами (от 30-го до 60-го перцентиля), а группа C — наименее продаваемые (ниже 30-го перцентиля). Видно, что «Proposed» решение улучшает метрику (1 — WAPE) практически во всех сочетаниях «Группа-Локация». У товаров группы C разница становится особенно заметной, так как ошибка в Baseline более подвержена резким колебаниям.

Суммарный показатель 1-WAPE для нового подхода оказался равен 72.44, в то время как у традиционного решения — 64.28. Такая разница свидетельствует о существенном преимуществе предлагаемого подхода при прогнозировании промо-эффекта.

Результаты и обсуждение

Эксперименты на данных сети быстрого питания продемонстрировали, что модель машинного обучения, работающая напрямую с выделенным промо-эффектом, в среднем улучшает точность прогноза (по WAPE) на 8% по сравнению с базовыми методами. Особенно заметной оказалась разница в случае сложных механик (например, «сеты»), где классические схемы ежедневно прогнозировали продажи, не учитывая взаимодействия между SKU. Градиентный бустинг же улавливает эти взаимодействия, используя факторы «состав набора», «соотношение цен» и т.п.

Отдельно следует отметить, что совмещение прогноза «базового» объема и «промо-эффекта» потенциально может приводить к некоторому завышению общего прогноза. Однако в условиях, когда «недопрогноз» хуже для бизнеса (вызывает дефицит продукта и недополученную прибыль), подобный сдвиг вверх оказывается оправданным: дополнительная подстраховка по запасам менее критична, чем риск потерь при неожиданном всплеске спроса. Для дополнительной иллюстрации на рисунке 2 приведены ящичные диаграммы (boxplots) ошибок прогнозов в новом и старом подходах, где четко видно, что медиана и верхние квартили распределения ошибок в новом решении (слева) оказываются заметно выше (ближе к нулю), чем в базовом сценарии (справа), где гораздо выше доля отрицательной ошибки.

Таким образом, можно заключить, что предложенный подход позволяет более точно оценивать планируемый суммарный спрос на период акции еще до ее запуска, что дает бизнесу возможность разумнее формировать запасы и управлять персоналом.

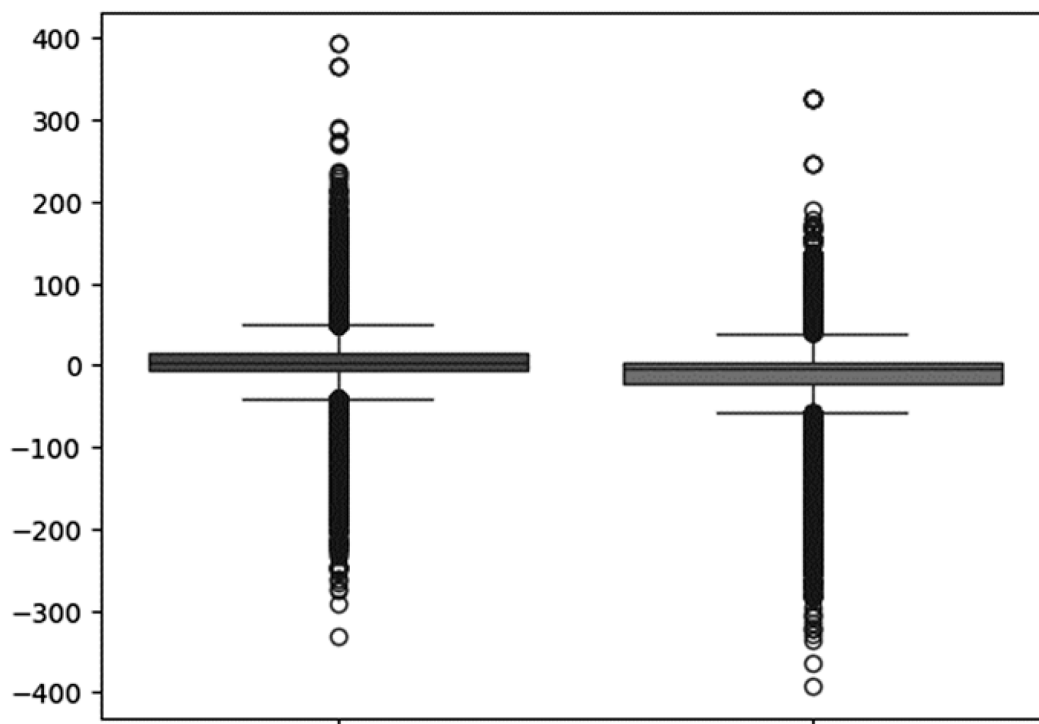


Рис. 2. Boxplot-анализ распределения ошибки прогнозов нового и старого подхода

Заключение

В настоящем исследовании показано, что выделение промо-эффекта из исторического ряда продаж в отдельную временную структуру и его последующее моделирование методами машинного обучения повышает точность прогнозирования. Алгоритмы градиентного бустинга, учитывая многомерные факторные признаки, эффективнее предсказывают итоговый прирост, чем классические модели, пытающиеся работать с промо в рамках единого временного ряда. Подтверждение по-

лучено на реальных данных сети ресторанов быстрого питания.

В перспективе целесообразно изучить экономический эффект от такого подхода (например, сокращение недопроданных или, напротив, избыточных запасов) и рассмотреть способы более тонкого распределения промо-эффекта по датам акции, включая учет внешних факторов конкуренции и применение гибридных моделей.

ЛИТЕРАТУРА

1. Ледовская Е.В. Методы машинного обучения для оптимизации экономических задач // Научный аспект. 2024. Т. 46. № 6. С. 5804–5820.
2. Парфенов П.А., Тимофеева А.А., Сологуб Г.Б., Алексейчук А.С. Прогнозирование покупки товара, показанного клиенту рекомендательной системой // Моделирование и анализ данных. 2020. Т. 10. № 4. С. 17–30.
3. Парамонов Е.С., Иванцов К.А., Мирончук В.А. Исследование востребованности применения методов машинного обучения в цифровой экономике // Beneficium. 2024. № 1 (50). С. 22–30.
4. Михайлов Д.В. Анализ практики применения алгоритмов машинного обучения в российских розничных сетях для развития операционной деятельности и разработки маркетинговой стратегии предприятия // Fundamental and applied approaches to solving scientific problems: сб. науч. ст. Уфа: НИЦ «Вестник науки», 2023. С. 122–141.

© Быстров Александр Игоревич (alexandrjri.bystrov@yandex.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»