

ОБЗОР МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ПРОДАЖ В ОБУВНОМ РИТЕЙЛЕ

Каипов И. К., Чигвинцев К. А.

*Белорусский государственный университет, Минск, Беларусь,
e-mail: kaipov1995@gmail.com*

Введение

Продажи в обувном ритейле характеризуется сезонностью и коротким жизненным циклом коллекций. Поэтому важно делать точные прогнозы продаж, остатки коллекций приводят к дополнительным логистическим расходам, потому что их нужно хранить до следующего сезона или уничтожить. В любом случае это приводит к потере прибыли. Большинство компаний хотят максимизировать свою прибыль. Все больше компаний пытаются использовать анализ данных для прогнозирования продаж, который помогает компаниям принимать правильные решения на основе данных. Ожидается, что глобальный рынок аналитики для розничной торговли вырастет с 4,3 млрд долларов в 2020 году до 11,1 млрд долларов к 2025 году при среднегодовом темпе роста (CAGR) 21,2%. [3]

В данном материале рассматривается обувная компания с развитой розничной сетью обувной продукции в России и Беларуси. В связи с прогрессом в области розничной аналитики конкурирующих компаний компания хочет разработать систему прогнозирования продаж для планирования распределения и пополнения обувной продукции. Пополнение запасов происходит не реже одного раза в 2 недели, поэтому горизонт прогноза равен 2 неделям. Желаемая точность прогнозов компании составляет максимум 10% погрешности для общего объема продаж. Компания также заинтересована в исследованиях по оценке потерянных продаж.

Методы прогнозирования. В данном материале рассматривается реализация трех методов прогнозирования продаж. Первый метод – это пример метода, используемого в текущих бизнес-планах. Второй метод является современным представителем методов временных рядов. И в-третьих, это один из самых популярных методов прогнозирования в сообществе специалистов по данным.

Первый метод носит название Like for Like (LFL). LFL – интуитивно понятный метод: поведение известного объекта используется для прогнозирования поведения аналогичного объекта. Обычно компании используют метод LFL для прогнозирования неизвестных событий. Например, компания В хочет оценить продажи в новом магазине. Он находит магазины со схожими характеристиками: население города, площадь магазина, регион, климат и т. д. Затем он оценивает продажи на основе аналогичных магазинов с некоторыми корректирующими параметрами. Открытие первого магазина в городе вызывает интерес и аналогично эффекту от новой коллекции, поэтому корректирующий параметр увеличит предполагаемые продажи.

Рассматриваемая компания также использует LFL для оценки продаж. Он работает следующим образом: для прогнозирования следующего месяца менеджеры по продажам проверяют продажи за аналогичный период прошлого года и корректируют его с учетом коэффициента роста бизнес-плана. На 2020 год коэффициент роста составлял 6%.

LFL – это простой метод, который игнорирует любой промежуточный анализ, специфичный для плановой экономики и стратегий продвижения. Однако прогноз LFL лежит в основе бизнес-плана рассматриваемой компании и будет сравниваться с другими методами прогнозирования в данном материале.

Второй метод Prophet – это библиотека с открытым исходным кодом для прогнозирования временных рядов, созданная Facebook и выпущенная в 2017 году. Это аддитивная регрессионная модель с четырьмя основными компонентами [4]:

- тренд;
- годовая сезонная составляющая, смоделированная с использованием рядов Фурье;
- еженедельная сезонная составляющая;
- праздничные и выходные дни.

Такие методы, как ARIMA и экспоненциальное сглаживание, включены в пакет Prophet с разным весом, с его сильными и слабыми сторонами, поэтому Prophet можно рассматривать как современного представителя методов прогнозирования временных рядов с разложением данных на основные компоненты.

Модель Prophet можно записать как формулу 1:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (1)$$

Тренд $g(t)$ – это линейная или логистическая функция для представления тренда. Самая важная особенность - автоматическое определение точек изменения в тренде или их установка вручную.

Годовые и недельные сезонные компоненты $s(t)$ отвечают за моделирование сезонных изменений. Годовая сезонность улавливает сезонные изменения, повторяющиеся из года в год. Еженедельная сезонность предназначена для отображения изменений за неделю.

Компонент $h(t)$ отвечает за отметку дней с аномалиями, например Пасха, Черная пятница, национальные праздники.

Ошибка ϵ_t представляет собой периодические изменения, которые не учитываются в модели, предполагается, что она нормально распределена.

Прогнозирование с использованием Prophet основано только на количестве продаж, поэтому модель не принимает во внимание такие характеристики, как характеристики продукта или погоду. Но Prophet хорош для построения прогнозов по умолчанию для категории плавного спроса с большим преимуществом в настройке поведения компонентов прогноза.

Градиентный бустинг – это метод машинного обучения для задач классификации и регрессии, который строит модель прогнозирования в виде ансамбля слабых предикторов (обычно деревьев решений). [1]

Оценка методов прогнозирования. Для экспериментов была выбрана история продаж 32 магазинов из 386 по следующим причинам. Во-первых, эти магазины были выбраны, чтобы избежать случайных событий, таких как задержка пополнения, работа этих магазинов стабильна и не содержит необъяснимого поведения в истории продаж. Также все магазины не новые, работали не меньше года. Во-вторых, эти магазины расположены в Санкт-Петербурге и Новосибирске, и эти города являются наглядными примерами разных сроков доставки: 1–2 дня и 1–2 недели соответственно. В-третьих, легче увидеть разницу в прогнозах при небольшом количестве магазинов.

Период прогнозирования составляет 2 недели с 18.10.2020 по 31.10.2020. Период в 2 недели выбран, поскольку это максимальное время выполнения заказа в сети, поэтому это максимальный период времени, необходимый отделу продаж прогноза для разработки политики распределения товара.

Абсолютная ошибка в процентах (APE) и средняя абсолютная ошибка в процентах (MAPE) выбраны в качестве метрик, поскольку необходимо знать, насколько точны методы прогнозирования, и не имеет значения, меньше или больше прогнозируемое значение настоящего значения продаж. [2]

Прогноз продаж осуществляется для каждого дня, затем все дневные продажи суммируются, чтобы представить продажи за период прогноза, который также является периодом между пополнениями.

Все программы для расчета прогнозов написаны на языке программирования Python. Общая оценка метрик, позволяющих судить об эффективности алгоритмов представлена в таблице 1.

Табл. 1. Сравнение метрик рассмотренных методов

	LFL	Prophet	LGBM_ date	LGBM_ features	LGBM_ lags_ tweedie	Ensemble_ best
APE	27,8%	6,8%	4,0%	8,0%	2,9%	0,5%
MAPE	31,8%	15,9%	10,6%	13,2%	11,5%	12,2%
MEDIAN	25,6%	16,2%	6,4%	12,2%	8,9%	8,7%

Для увеличения эффективности продаж не рекомендуется рассматривать какие-либо решения, занижающие продажи, потому что заниженный прогноз ведет к более низким продажам, чем завышенный. Учитывая вышенаписанную рекомендацию, остаются LFL, Prophet и Ensemble_best. Ensemble_best имеет лучшие, чем другие кандидаты, поэтому лучше использовать ансамбль из Prophet и LightGBM, что подтверждает практические рекомендации использовать ансамбли моделей, а не отдельные модели.

Работа выполнялась при поддержке проектов БРФФИ-РФФИ Ф20Р-134

Литература

1. Boosting Algorithms as Gradient Descent / L. Mason [et al], 2000, p.513-517
2. Mood, A. Introduction to the Theory of Statistics / . McGraw-Hill, 1974. p. 229
3. Retail Analytics Market by Application (Merchandising Analysis, Customer Analysis, Performance Analysis), Business Function (Finance, Sales, Marketing, Supply Chain and Store Operations), Component, End-User, Region & Forecast to 2025 [Electronic resource] / Markets and Markets – Mode of access: <https://www.marketsandmarkets.com/Market-Reports/retail-analytics-market-123460609.html> - Date of access: 26.03.2021.
4. Taylor, S.J. Forecasting at Scale / S.J. Taylor, B. Letham - The American Statistician, 2017 – 72(1), 35-42 p.