

УДК 004.81

АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ПРОДАЖ НА ПЛОЩАДКАХ ЭЛЕКТРОННОЙ КОММЕРЦИИ

Авторы: Громов Е.Д., Арсениев А.Н., Рубан А.О.

Научный руководитель: к. т. н, доцент Гулева В.Ю. (Университет ИТМО, г. Санкт-Петербург)

Ключевые слова: прогнозирование, малые выборки, маркетплейсы, прогноз спроса, прогнозирование временных рядов, методы прогнозирования.

Актуальность заключается в выборе наиболее эффективного метода прогнозирования временных рядов в условиях малых выборок, но большого числа наборов данных. Это необходимо для точного планирования и распределения товарных запасов, что в свою очередь увеличивает экономическую эффективность дистрибуции.

В настоящее время существует огромное количество методов прогнозирования временных рядов, начиная от самых простейших, таких как метод скользящего среднего, метод экспоненциального сглаживания, метод Хольта, продолжая более сложными методами, такими как VAR, ARIMA с ее модификациями (VARIMA, SARIMA, SARIMAX), имитационным моделированием, как например метод Монте-Карло и заканчивая методом случайного леса и нейронными сетями, такими как LSTM и ETS. Однако у каждого метода прогнозирования есть свои ограничения и недостатки, которые препятствуют качественному прогнозированию, особенно в условиях малых выборок.

Цель: Выявление наиболее эффективного метода прогнозирования временных рядов в условиях малых выборок, но большого числа их наборов. Выявление корреляций количества продаж и других параметров, таких как цена, количество остатков на складе. Поиск точек совмещения различных методов прогнозирования для их объединения в новый метод. Выполнить наиболее точный прогноз на 30 дней, имея данные за 60 дней.

Наборы данных получены с крупнейшего сервиса аналитики площадок электронной коммерции MPSTATS.

В наборах данных содержатся данные о товарах за 60 дней: дата, количество продаж, цена, количество остатков на складе. Для апробации методов прогнозирования было собрано 10000 наборов данных.

Данные о продажах являются нестационарными временными рядами из-за особенностей спроса и предложения на площадках электронной коммерции, а именно из-за различных акций и распродаж, сезонности товаров, повышенного спроса перед праздниками и других факторов. Также

для каждого вида товара характерны различные виды распределения и спроса.

Ввиду вышеперечисленных особенностей простейшие методы прогнозирования, такие как метод скользящей средней, метод экспоненциального сглаживания или метод Хольта, совмещающий с себе 2 предшествующих метода, применять нецелесообразно, так как они следующее прогнозное значение получают из прошлого значения, это значит что прогноз на 2 и более периодов вперед будет основываться на уже спрогнозированных значения. Следовательно прогноз на 2 и более периодов вперед будет неточным.

При рассмотрении более сложных методов прогнозирования, таких как VAR, ARIMA результаты прогнозов улучшаются, однако эти методы не автоматизированы и требуют участия человека для эффективного прогноза различных товаров из различных категорий. При попытке применить AUTOARIMA результаты прогнозов ухудшаются.

Наилучшим из возможных выходов становится использование нейросетей и/или машинного обучения. Случайный лес и модель LSTM имеют свои особенности и ограничения, такие как необходимость обучения на большом количестве наборов данных и разбиение товаров на категории, однако эффективность прогноза у данных моделей самая высокая. Однако в условиях малых выборок и большого числа наборов данных обучение нейронной сети является оптимальным методом прогнозирования

Наиболее перспективными методами прогнозирования являются случайный лес и LSTM. При проверке всех методов используется ошибка RMSE. Для подтверждения эффективности необходимо проводить апробацию методов на различных категориях и большом количестве выборок.

Громов Е.Д. (автор):

Гулева В.Ю. (научный руководитель):