

Отчет по задаче прогнозирования продаж.

Андрей Журавлев

С целью прогнозирования продаж были предоставлены данные по продажам в четырех городах и данные по потоку потребителей в данных городах. На основе этих данных была построена модель регрессии с помощью библиотеки CatBoost.

Основные результаты работы заключаются в следующем.

1. Были проанализированы продажи товаров в разных городах. Построена модель предсказания продаж. На отложенной тестовой выборке точность по метрике $\text{SMAPE} = 4,02\%$.
2. На основе полученной модели были сделаны предсказания продаж на три месяца вперед. Результат находится в папке result.
3. Анализ влияния цены на продажи показал, что изменение цены не приводит к значительному изменению продаж. Возможно, это связано с повышенным спросом в летний период. В результате при увеличении цены для каждого бренда приведёт к увеличению получаемой выручки.

Дальнейшие шаги по улучшению качества предсказания.

1. В целом наблюдается недостаточное количество данных для большого количества товаров. Необходимо увеличить данные как за представленные года, так и сделать агрегацию продаж по декадам месяца. Это позволит более корректно учесть сезонность (fbprophet) и влияние цены. Сейчас влияние цены оказалось достаточно низким в предсказании.
2. Увеличение данных также позволит разбить товары на группы с низкой и высокой сезонностью, что в целом благоприятно скажется на соответствующие модели.
3. Увеличение данных также позволит корректно построить модель. Сделать полноценный подбор параметров по валидационной выборке и оценить ошибку модели на отложенной тестовой выборке.
4. Проблему с высокой скоррелированностью групп покупателей можно решить с помощью PCA. Однако введение дополнительного гиперпараметра сейчас затруднительно из-за нехватки данных.
5. Поскольку для предсказания делали стандартизацию продаж, уместно подобную стандартизацию сделать и для покупателей для построения более надежных зависимостей.