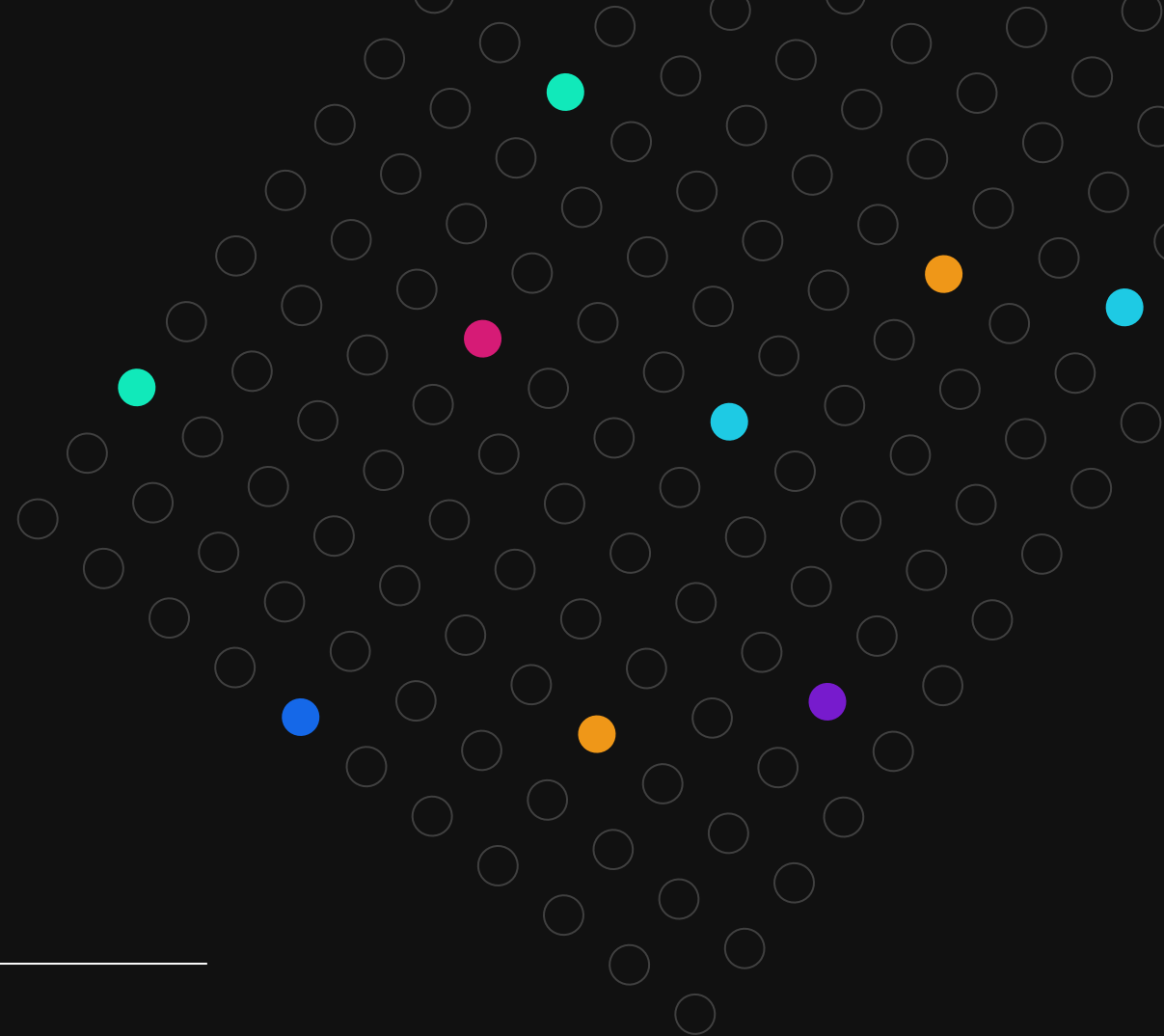


Прогнозирование инфляции

цифровой
прорыв

сезон: III



Постановка задачи и проблематика

Алгоритм прогнозирования месячного ИНЦ

На основе микроданных о потребительских ценах, собранных посредством веб-скрейпинга интернет-магазинов, разработать модель, позволяющую формировать краткосрочный прогноз месячного ИПЦ в режиме реального времени.



Веб-скрейпинг

Является технологией, реализация которой представляется наиболее предпочтительной относительно других технологий сбора данных о ценах



Проблема неровного края

Данная проблема возникает при работе с данными и может затруднять оперативное прогнозирование инфляции



Краткосрочные модели

Позволяют оценивать волатильные показатели, дополняя долгосрочный прогноз и делая его более точным

Шаги решения



Общий анализ данных

Изучение пропусков, корреляции, распределения цен



Подготовка данных

Отбор товаров, избавление от выбросов, заполнение пропусков, преобразование цен в индексы



Кластеризация данных

Разделение товаров на категории с помощью KMeans



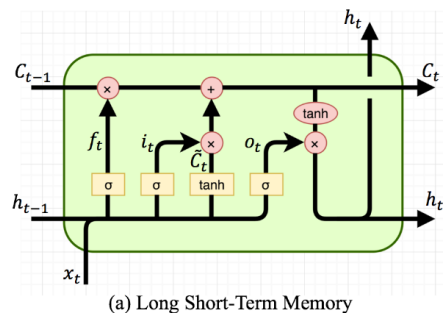
Обучение моделей

Обучение моделей на данных изменения цен относительно предыдущего месяца

Рабочий процесс

Не сработало:

- ARIMA
- Facebook Prophet
- DeepAREstimator (GluonTS)



Успешно сработало:

- Linear Regression
- CatBoost
- Метод кластеризации KMeans



CatBoost

Уникальность подхода:

- Кластеризация товаров по изменению цен
- Предсказание среднего изменения цены в группе
- Хорошая адаптируемость за счет использования простых моделей

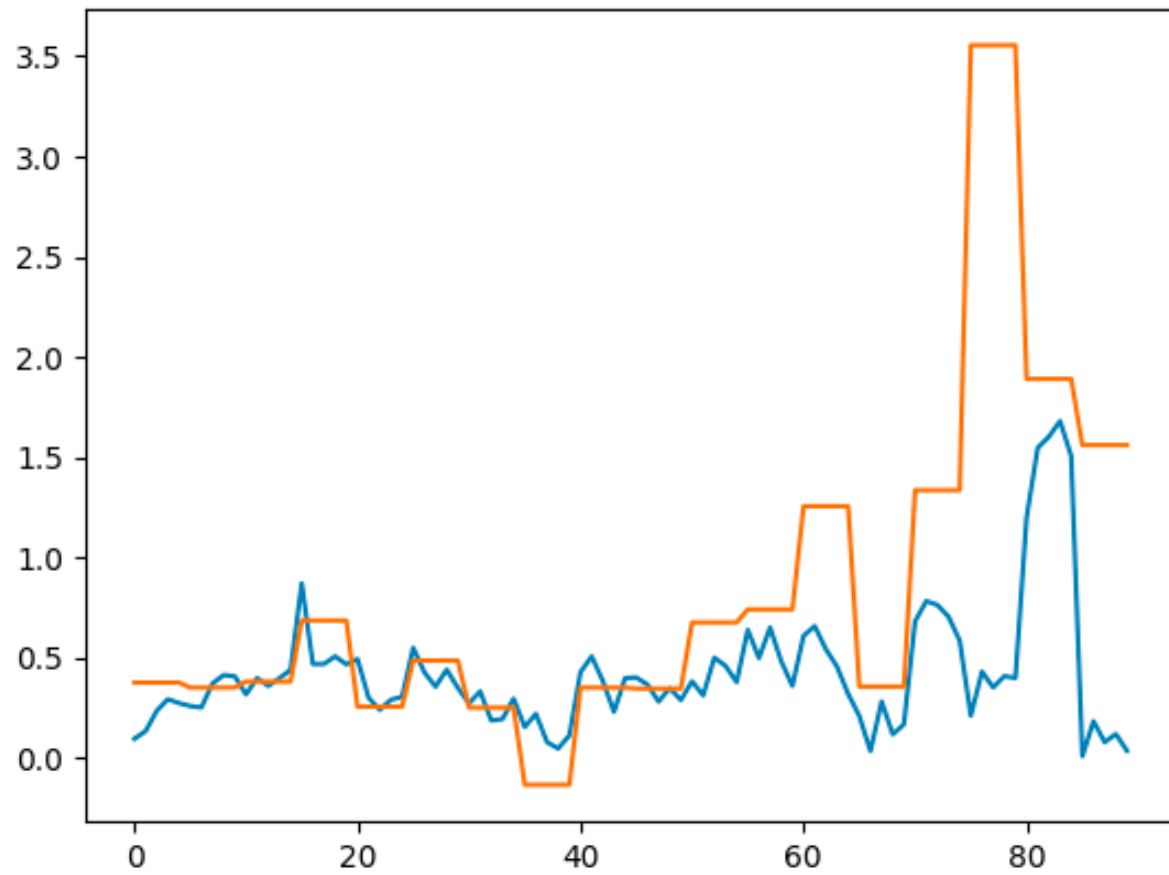
Вывод:

Для нейросетей данных слишком мало, более простые подходы работают лучше

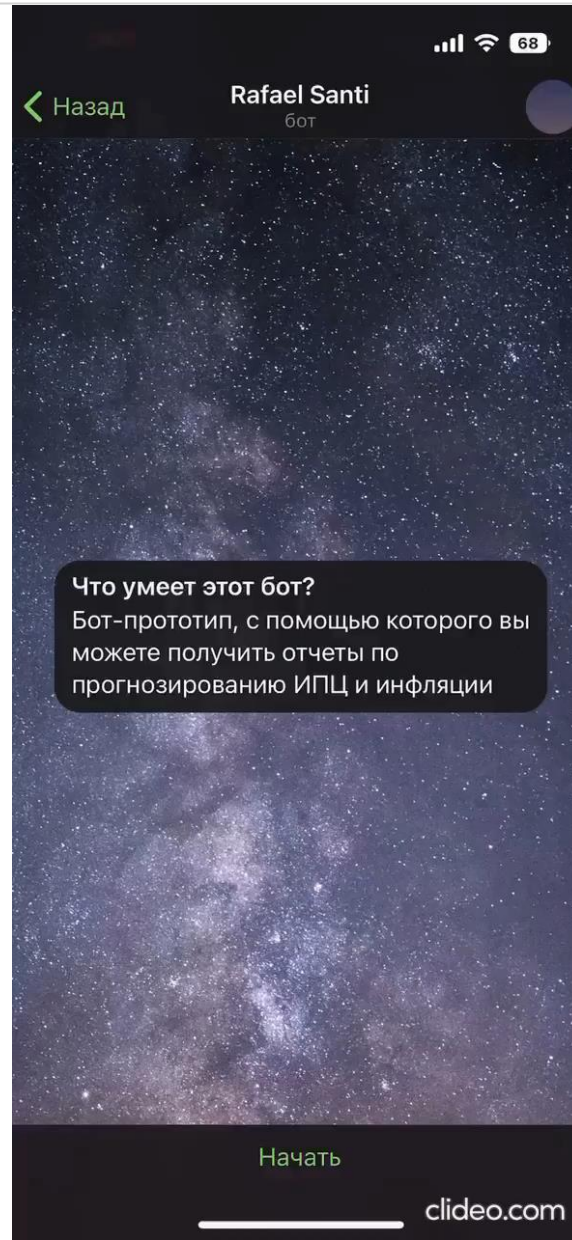


Результаты

Mean Absolute Error на валидационной выборке - 0.45



Как мы видим продакшн



Состав команды



Дмитрий Куценко

@IT_shnik

Data Scientist



Никита Романов

@KARTASAR

Data Scientist



Давид Джалаев

@RomanovDavid

Backend developer