

САРАФАН – все магазины в одной корзине

Подтема

Модель машинного обучения для предсказания категорий товаров





Марина Дружинина

Data Science в медицине, год поступления в GeekBrains 2020

Немного о себе. Краткое описание
в несколько строк:

- Проживаю в городе Уфа
- Увлекаюсь машинным обучением, искусственным интеллектом
- Люблю путешествовать



Поставленная задача

Сарафан – это сервис в котором можно сравнивать продукты питания из разных магазинов и заказать домой или в офис в одном заказе (<https://sarawan.ru/>)

Основная цель — это возможность заказать товары (прежде всего продукты) в одной корзине из разных крупных сетевых магазинов с доставкой на дом.

Задача – разработать модель машинного обучения для определения категорий товаров. Категории в разных магазинах могут немного отличаться, и целью является унифицировать названия категорий.



Стадии выполнения проекта

1

Проанализировать данные из разных магазинов

2

Выбрать подходящий для обучения модели датасет

3

Предобработка данных и создание новых признаков

4

Обучить несколько моделей

5

Оценить полученные результаты

6

Выбрать лучшую модель

7

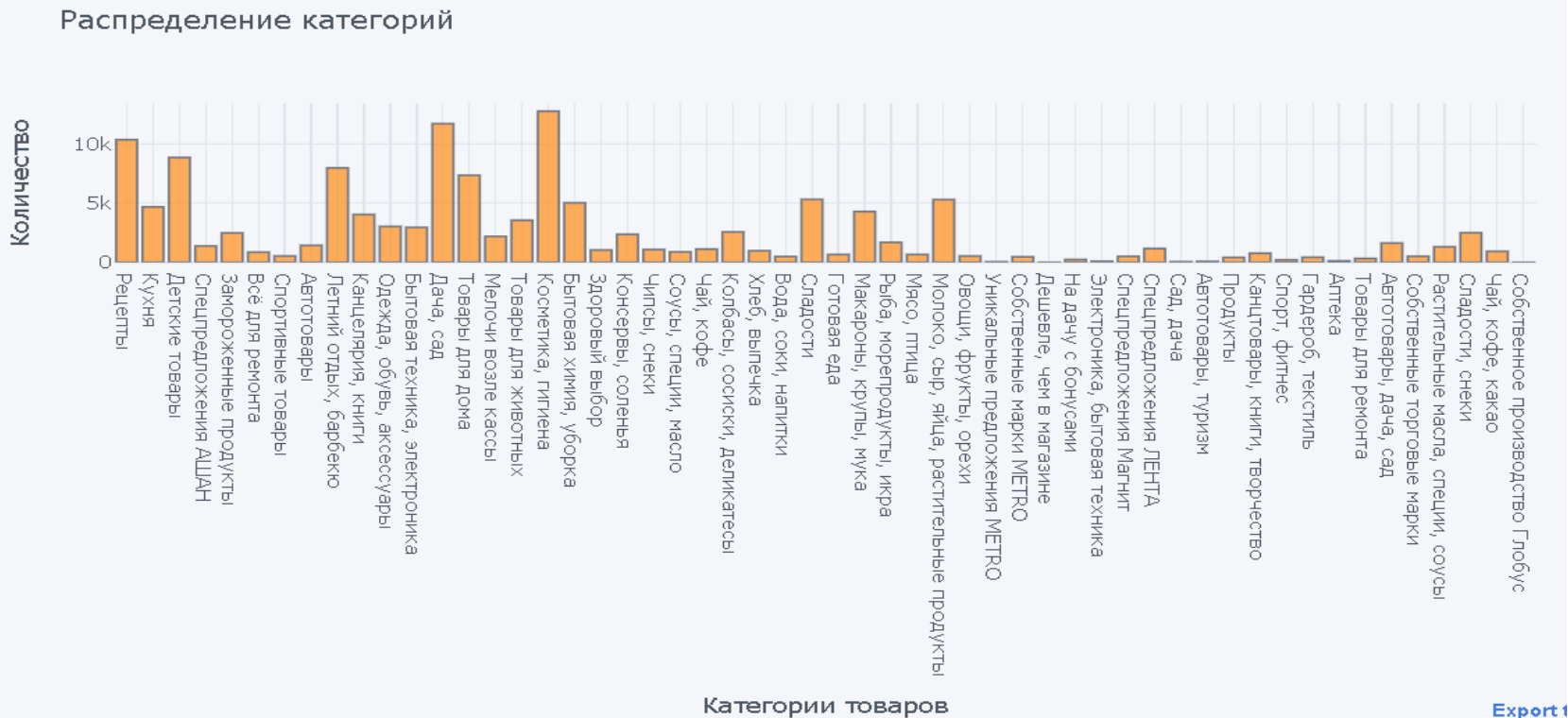
Определение категорий товаров с помощью модели

8

Тестирование на сайте



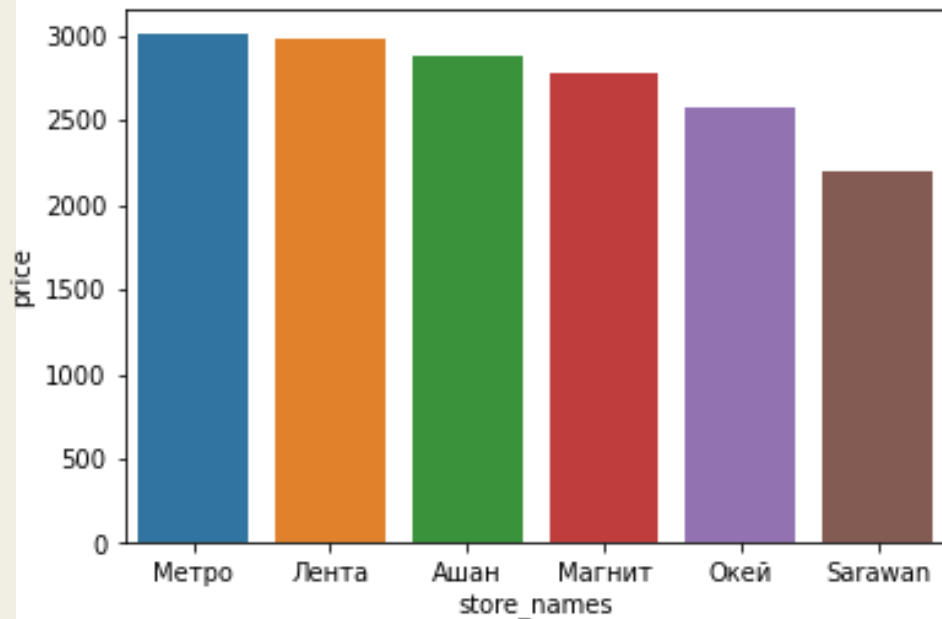
Магазины представлены следующими категориями товаров





Стоимость корзины продуктов в разных магазинах

Было смоделировано 1000 случайных корзин и подсчитано, что экономия с Сарафаном составляет в 15-20% от минимальной цены .





Итоговая модель



Модель машинного обучения

CatBoostClassifier.



Признаки для обучения

Было выбрано 3 категориальных признака:

- Первое слово в названии продукта
- Бренд
- Свойство (слово) из информации о продукте



Метрики

- TotalF1
- precision
- recall

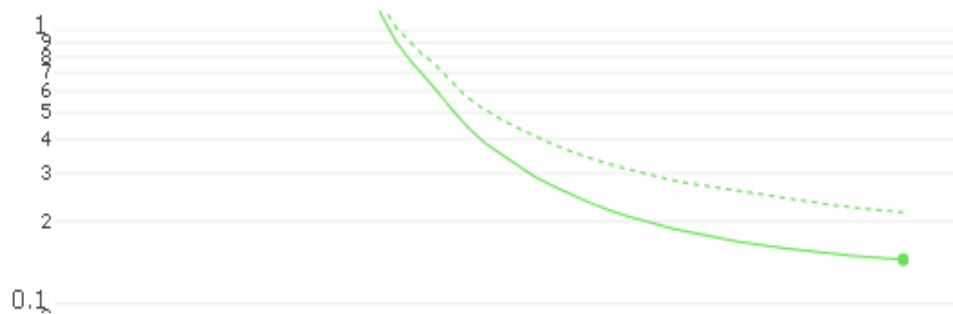


Функция потерь

MultiClass

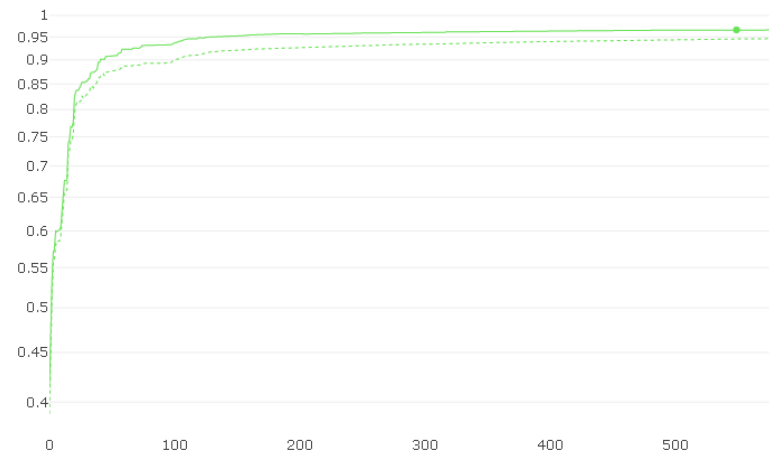


Графики обучения модели



Функция потерь: MultiClass

--- train
— test



F1 score

--- train
— test



Пример предсказания категорий моделью

```
In [118]: new_df.loc[new_df['name_store'] == 'Ашан', ['product_name', 'main_category_name', 'main_category_name_pred']].head(50)
```

Out[118]:

	product_name	main_category_name	main_category_name_pred
0	Петрушка	Рецепты	Овощи, фрукты, зелень, орехи
1	Петрушка в горшочке 120 г	Рецепты	Овощи, фрукты, зелень, орехи
2	Петрушка 100 г	Рецепты	Овощи, фрукты, зелень, орехи
3	Плавленный сыр Hochland Чизбургер 8 ломтиков 45...	Рецепты	Молочные продукты, яйца, растительные продукты
4	Помидоры черри Мое Лето желтые круглые без вет...	Рецепты	Овощи, фрукты, зелень, орехи
5	Оливковое масло Colavita Extra Virgin Greek 50...	Рецепты	Бакалея
6	Тарелка обеденная 25 см желтая	Кухня	Кухня
8	Тарелка десертная Luminarc Лайнз Q1898 белая с...	Кухня	Кухня
9	Оливковое масло Colavita Extra Virgin Mediterr...	Рецепты	Бакалея
10	Тарелка суповая 20,8 см желтая	Кухня	Кухня
11	Тарелка десертная Pasabahce Water Lily 19,5 см...	Кухня	Кухня
13	Тарелка для вторых блюд Actuel Алиса 25 см белая	Кухня	Кухня
16	Книга В гостях у синего трактора. Синий тракто...	Детские товары	Канцелярия, книги
17	Салатник 15 см ассортименте	Кухня	Кухня
18	Салатник Каждый день средний 15,8 см	Кухня	Кухня
19	Книга Первая сказка. Слушай и играй. Колобок	Детские товары	Канцелярия, книги
21	Книга Мой атлас с наклейками. Динозавры Урбано...	Детские товары	Канцелярия, книги
22	Салатник Pasabahce Invitation Bronze 12,5 см	Кухня	Кухня



Достигнутые цели \ Итоги

1. Разработана модель машинного обучения с применением CatBoost.
2. Сложности при выполнении проекта:
 - дисбаланс классов (какие-то категории представлены большим количеством товаров, а какие-то - малым)
 - оценка качества модели, так как в каждом магазине продукты объединены в категории в соответствии со своей бизнес логикой, которая иногда не совпадает.
3. Следующий этап – тестирование результатов на сайте.



Идеи на будущее

- В настоящий момент проект разрабатывается командой разработчиков из разных областей на общественных началах. Это хорошая возможность попробовать свои силы и получить опыт работы над реальным проектом.
- Разработан сайт, а также мобильные приложения Андроид и iOS, в котором можно сравнивать продукты питания из разных магазинов и заказать домой или в офис в одном заказе. Также разрабатывается приложение для курьеров.
- Проект находится в стадии тестирования, а в дальнейшем планируется запустить полноценный сервис.
- Также в будущем планируется разработать бота-помощника с использованием нейросетей.