Мы владеем сетью магазинов, в которых продаются различные товары. Наши сотрудники собрали датасет из 8523 продаж товаров в различных магазинах нашей сети. Наша задача построить модель, предсказывающую продажи каждого продукта в конкретном магазине. Используя эту модель, попробовать выявить факторы, больше всего влияющие на увеличение продаж. Описание датасета:

|  |  |
| --- | --- |
| Variable | Description |
| Item\_Identifier | Id продукта |
| Item\_Weight | Вес продукта |
| Item\_Fat\_Content | Содержание жира в продукте |
| Item\_Visibility | %полок, отведенный под наш продукт в конкретном магазине |
| Item\_Type | Категория продукта |
| Item\_MRP | Максимальная цена продажи продукта |
| Outlet\_Identifier | Идентификатор магазина |
| Outlet\_Establishment\_Year | Год открытия магазина |
| Outlet\_Size | Площадь магазина |
| Outlet\_Location\_Type | Тип города, в котором расположен магазин |
| Outlet\_Type | Признак является ли магазин продуктовым или супермаркетом |
| Item\_Outlet\_Sales | Продажи продукта в конкретном магазине. Именно ее и надо предсказывать |

В результате работы должен получиться:

− Jupyter-ноутбук с моделью

- Признаки, влияющие больше всего на уровень продаж

− Датасет, если после ваших манипуляций он отличается от исходного;

− Документ с обоснованием решения и краткими результатами: какие техники и почему использовали, что получили, что можно улучшить (можно в рамках jupyter notebook’а)

## 

## Критерии оценки:

|  |  |
| --- | --- |
| **Баллы** | **Что надо сделать** |
| 10 | Провести EDA |
| 10 | Обработать категориальные признаки |
| 10 | Устранить пропущенные значения |
| 10 | Изучить корреляцию признаков с данными о продажах |
| 10 | Выбрать и обосновать метрику, на основе которой будем измерять качество полученной модели |
| 20 | Построить и подобрать оптимальные параметры для любой линейной модели |
| 20 | Построить и подобрать оптимальные параметры для любой нелинейной модели |
| 20 | Провести стекинг нескольких моделей |
| 10 | Оценить качество модели на отложенной выборке |
| 10 | Выбрать топ 3 признака больше всего влияющие на объемы продаж |

Все шаги решения должны сопровождаться подробным описанием полученных результатов и обоснованием выбора того или иного шага! За отсутствие описания так же будут снижаться баллы.

Максимальное доступное количество баллов - 130

Для получения зачета надо набрать минимум 80 баллов

Для получения зачета с отличием надо набрать минимум 120 баллов