**ОПИСАНИЕ ДАТАСЕТА**

В файле представлены чековые продажи трёх SKU в крупной ритейл-сети за период 2019-01-01 – 2020-01-07 включительно.

Имеющиеся атрибуты:

* ID\_check – Идентификационный номер чека
* ID\_store – Идентификационный номер магазина
* Date\_buy – Дата приобретения товара
* ID\_SKU – ID товара
* SKU\_QTY – Количество единиц товара в чеке
* SKU\_cost – Стоимость товара в чеке без скидки
* Discount – Размер скидки в долях

**ЗАДАЧА**

Необходимо спрогнозировать агрегированные продажи каждого товара для всей сети по неделям. Разрез агрегации Календарная неделя – SKU – Тотал сеть. Горизонт и лаг прогнозирования 1 неделя – например, для прогнозирования недели 1 можно использовать только фактические значения недели 0 и более ранних периодов. При этом модель не должна засасывать данные из будущего и учитывать неделю 1 и более поздние. Исключением является поле Discount – это информация из промо-календаря, который известен заранее, поэтому мы можем знать наличие и уровень скидки в будущую неделю продаж.

Тестирование должно проводиться на последних 10 календарных неделях, эти недели не должны участвовать в обучении моделей (включая кросс-валидацию)

Для выполнения задачи необходимо выполнить следующие пункты:

1. Сгенерировать минимум 5 фичей (предикторов) для прогнозирования. Создание фичей необходимо реализовать на любом диалекте SQL (лучше всего T-SQL).
2. Сформировать корректное целевое событие: Количество проданных единиц каждого товара за неделю для всей торговой сети.
3. Полученный итоговый датасет с целевым событием и фичами импортировать в Python и построить модель прогнозирования продаж с использованием методов машинного обучения на языке Python (любые алгоритмы или ансамбли, число моделей и разрезов прогнозирования неограниченно).

**КРИТЕРИЙ ОЦЕНИВАНИЯ**

В качестве метрики оценки качества модели используется MAPE. Для тестовых результатов необходимо подсчитать точность по формуле: 1 – сумм (ошибка) / сумма (факт), где ошибка это abs(прогноз – факт).

**РЕЗУЛЬТАТЫ РАБОТЫ**

В качестве результатов выполнения задания необходимо собрать архив:

1. SQL-скрипты по обработке данных, формированию предикторов и целевых событий.
2. Закомментированный программный код модели на языке Python в формате .ipynb или .py.
3. Полученные показатели точностей на тестовых данных в разрезе Календарная неделя – SKU – Тотал сеть. Файл с результатами прогнозирования в формате .xlsx. Примерный вид результирующего файла:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID\_SKU** | **Week ID** | **Forecast** | **Fact** | **ABS Error** | **Accuracy** |

Также по каждому SKU необходимо подсчитать средневзвешенный уровень точности

1. Краткое описание проделанных работ, использованных подходов и алгоритмов, а также сделанных выводов.
2. Ссылку на ваше резюме на hh.ru

**СРОК СДАЧИ ТЕСТОВОГО ЗАДАНИЯ**

Крайний срок сдачи тестового задания: 1 октября до 23:59