

Прогнозирование спроса по новому ассортименту

Описание проекта

Проект состоит из двух задач, обе про предсказание спроса по новому ассортименту. В первой задаче у каждой команды свои данные (похожая структура, сложность данных - одинаковая), а вот категории товаров - разные; цель - побить бенчмарки. Вторая задача общая для всех и участие в ней будет носить соревновательный характер.

Выставление оценки по команде за участие в проекте (одна оценка для всех членов команды) будет по формуле

$$\min(\text{mark_task1}, \text{mark_task2}),$$

Т.е. чтобы получить 10 нужно чтобы и по первой и по второй задачам было 10.

Задача 1

В проекте предлагается предсказать интенсивность продаж новых товаров в розничной сети с помощью методов ML. Под интенсивностью продажи подразумевается количество спроса на товар в конкретном магазине для каждого дня в течение первых 28 дней продаж этого товара в магазине.

В качестве исходных данных для предсказания предоставляются

- 1) История продаж за последние 35 дней по **всем** товарам категорий Яйцо, Бананы, Сыры Твердые, Сыры плавленые.
- 2) различные характеристики товаров и магазинов:

Поле	Пример	Описание
PRODUCT_LVL_NM8	Яйцо	Категория товара
PRODUCT_LVL_NM13	Яйцо куриное	Субкатегория товара
PRODUCT_ID	534470	ID - товара
STORE_LOCATION_LVL_NM2	47 ЛЕНИНГРАДСКАЯ ОБЛ	область
STORE_LOCATION_LVL_NM7	22_1056420 ПИКАЛЕВО Г	населённый пункт
ST_LOC_DIVISION_NM	SZ	дивизион (аналитический) магазина
ST_LOC_MNGMT_CLSTR	D_SZ04	кластер (аналитический) магазина
STORE_LOCATION_ID	834663	ID - магазина
start_date	21.01.2019	дата, на которую нужно выдать прогноз
FCST	.	величина прогноза

Данные предоставлены в двух файлах: исходные данные (NPF_TRAIN) и список тех товаров/магазинов/дней, для которых нужно выдать прогноз NPF_TEST.csv.

Оценка результатов

В качестве функции штрафа за неточное предсказание используется WAPE:

$$WAPE = \frac{\sum_{(i: \text{тройка } SKU|TT|\text{день})} |A_i - F_i|}{\sum_{(i: \text{тройка } SKU|TT|\text{день})} (A_i)},$$

где A_i – истинное значение спроса для тройки $SKU|TT|\text{день}$;

F_i – прогноз спроса для тройки $SKU|TT|$ день;

Оценка результатов задачи 1

Каждая команда должна подготовить прогноз для определённой группы товаров, соответствующие командам категории и бенчмарки для них (чем меньше ошибка, тем лучше результат) приведены в таблице ниже:

Команда №	Категории	Benchmark 1 (Ошибка WAPE)	Benchmark 2 (Ошибка WAPE)	Benchmark 3 (Ошибка WAPE)
4	Сыры плавленые	100%	98%	97%
29	Сыры твердые	100%	98%	97%
6	Яйцо куриное	100%	99%	98%
16	Бананы	100%	99%	98%

Список групп:

номер группы	Студент №1 ФИО	Студент №2 ФИО	Студент №3 ФИО
4	Ширнин Александр Андреевич	Маркин Никита Александрович	Лупашин Евгений
29	Бабушанова Дарья Алексеевна	Скрыльникова Дарья Алексеевна	Юрлов Павел Павлович
6	Андреев Михаил Алексеевич	Зисман Илья Александрович	Сушкевич Лев Алексеевич
16	Стрипский Даниил Игоревич	Агаев Фархат Чингизович	Савосин Артем

Оценка выставляется по правилу:

- 1) если не работает присланный на проверку финальный код (или он вообще не прислан), тогда оценка 2 (по десятибальной)
- 2) иначе:

если WAPE прогноза меньше benchmark1 команда получает оценку 8, с ошибкой меньше benchmark2 — оценку 9 (в зависимости от решения), с ошибкой меньше benchmark2 — оценку 10.

Сабмит прогнозов отправляется в соревнования Kaggle: TBD.