

Методические материалы к хакатону «Tender Hack – Москва»

Цель и задачи

Разработать предиктивную систему для анализа ключевых метрик котировочных сессий.

Кейс будет в момент официального начала хакатона. Датасет будет представлять собой выгрузку в формате CSV (данные по котировочным сессиям).

Портал, где будет размещен сервис — <https://zakupki.mos.ru/>

В рамках хакатона участникам предстоит реализовать задачи:

- Изучить принципы работы котировочных сессий и их ключевые метрики;
- Проанализировать исходные данные, для понимания потенциальных путей решения задачи, отбора ключевых признаков и установления возможных зависимостей;
- Произвести предобработку данных для дальнейшей работы с ними;
- Разработать архитектуру системы для решения поставленной задачи;
- Выбрать метрики качества и оценки эффективности предложенной системы;
- Реализовать выбранную модель;
- Провести тестирование системы, оптимизировать ее работу в рамках улучшения метрик;
- Продемонстрировать работоспособность проекта;
- Определить возможности по масштабированию решения и следующим его доработкам;
- Презентовать проект;

Описание полей датасета

Номер КС — номер котировочной сессии;

ИНН —захешированный ИНН заказчика;

Статус — статус котировочной сессии на момент составления датасета;

Наименование КС — наименование котировочной сессии;

ОКПД 2 — кодировка товаров, работ или услуг – объектов котировочной сессии по общему классификатору продукции по видам экономической деятельности;

КПГЗ — кодировка товаров, работ или услуг – объектов котировочной сессии по классификатору предметов государственного заказа;

Регион — регион поставки товара, предоставления услуги или выполнения работы;

НМЦК — начальная максимальная цена;

Итоговая цена — итоговая цена на момент закрытия котировочной сессии;

Дата — Дата проведения котировочной сессии;

Участники — число участников, принявших участие в торгах за данную котировочную сессию;

Ставки – общее число ставок, сделанных участниками в данной котировочной сессии.

Этапы защиты

Метрики: основными метриками качества являются MAE процентного падения НМЦК и MAE конкуренции на котировочной сессии. В случае её незакрытия считается, что уровень падения составляет 100%.

Этап 1: Отправка результатов предсказания системы на контрольной выборке в соответствующую онлайн форму и проектов на облачный репозиторий для проведения дальнейшего код-ревью. Защита проектов (презентация до 5 минут). Реальный кейс параллельно демонстрирует второй участник команды. Максимальное количество баллов по критериям — 100.

Этап 2: Защита топ-5 решений (презентация до 10 минут, затем 10 минут на вопросы жюри). Максимальное количество баллов по критериям — 100. Баллы суммируются с баллами первого этапа.

Критерии оценки:

1. Работоспособность прототипа — 70 баллов.
2. Оригинальность идеи — 10 баллов.
3. Масштабируемость — 10 баллов.
4. Применение передовых Computer Science подходов — 10 баллов.

Полезные материалы:

- <https://zakupki.mos.ru/cms/Media/docs/%D0%98%D0%B-D%D1%81%D1%82%D1%80%D1%83%D0%BA%D1%86%D0%B8%D1%8F%20%D0%BF%D0%BE%20%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%B5%20%D1%81%20%D0%9F%D0%BE%D1%80%D1%82%D0%B0%D0%BB%D0%BE%D0%BC%20%D0%B4%D0%BB%D1%8F%20%D0%B7%D0%B0%D0%BA%D0%B0%D0%B7%D1%87%D0%B8%D0%BA%D0%B0.pdf> — руководство для заказчиков по работе с порталом;
- <https://colab.research.google.com> — среда разработки от Google;
- <https://scikit-learn.org/> — Python-библиотека для машинного обучения;
- https://radimrehurek.com/gensim_3.8.3/index.html — Python-библиотека для работы с текстом и создания текстовых моделей;
- <https://fasttext.cc/> — Python-библиотека для создания текстовых моделей;
- <https://docs.python.org/3.8/> — документация Python;
- <https://www.anaconda.com/> — дистрибутив языков Python и R, включающий библиотеки для машинного обучения.