

Описание MVP сервиса

Описание структуры проекта

Фронт-енд часть сервиса представляет собой Streamlit приложение, которое использует реализованный на FastApi функционал по обучению бейзлайн-модели для предсказания исхода матча. Решение обернуто в докер. Для получения прогноза используются исторические данные по матчам и их исходам с 2017 года. Сервис предсказывает победу домашней команды (1 - победа домашней команды, 0 - ничья или проигрыш).

Для получения предсказания пользователю необходимо загрузить данные в нужном формате для обучения модели. Сервис дает возможность проанализировать данные, обрабатывает их перед обучением и предлагает задать гиперпараметры для модели. Все вычисления происходят на сервере через FastAPI, фронт-енд часть используется для получения инпутов от пользователя и связывает их с API.

Описание функционала API

Бэкенд содержится в двух файлах – `api_route.py`, где находятся все ручки, и `main.py`, который запускается для запуска сервиса.

Бэкенд развернут на FastAPI, при запуске загружается информация о бейзлайновой модели – объект класса стандартизации числовых фичей, объект класса кодирования категориальных признаков и веса обученной модели. Помимо этого присутствует два словаря – `models` и `models_load`; первый содержит в себе все когда-либо обученные модели и бейзлайновую модель с ключом `model_id` и значением в виде словаря гиперпараметров; словарь `models_load` имеет ту же структуру, только в нем лежат загруженные модели, доступные для получения предсказания.

Бэкенд содержит в себе следующие ручки:

- 1) `show_example` – данная ручка не принимает на вход параметров и возвращает набор случайных наблюдений из обучающего датасета в формате csv, используемый для обучения бейзлайновой модели
- 2) `fit` – данная ручка получает на вход json с ключами `model_id` (значение строкового формата) и `hyperparameters` (значение формата Dict) и запускает обучение модели и сохраняет объект класса стандартизации числовых фичей, объект класса кодирования категориальных признаков и веса обученной модели; результатом работы ручки является возвращенное сообщение "Model `model_id` is trained and saved"
- 3) `predict` – данная ручка получает на вход json с ключом `model_id` (значение строкового формата) и набор данных для предсказания целевой переменной; модель может быть использована для предсказания только в том случае, если до этого была загружена с помощью ручки `set_model`; результатом работы ручки является возвращенный изначальный файл формата csv с дополнительной бинарной колонкой `preds`

- 4) `show_models` – данная ручка не принимает на вход параметров и возвращает словарь активных моделей в формате `model_id: hyperparameters: {}`
- 5) `set_model` – данная ручка принимает на вход json с ключом `model_id` (значение строкового формата) и возвращает сообщение “Model `model_id` is loaded”. С помощью этой ручки модель из словаря `models` добавляется в словарь `models_load`, после чего с помощью нее можно получить предсказание.

Описание функционала Streamlit-приложения

Streamlit-приложение - это одностраничный веб-сервис, который предлагает пользователю получить предсказания по необходимым матчам.

Видео по работе сервиса: https://disk.yandex.ru/i/HyfB-tL_t63luA

Структура приложения:

1. Загрузка данных - приложение проверяет, чтобы содержание данных было подходящим для обучения. В случае, если данные отличаются от нужного формата - приложение выводит предупреждение.
2. Если данные загружены, можно увидеть сделать краткий разведывательный анализ данных (для MVP сервиса он сокращен, в финальной версии хотим предоставить больше возможностей). Возможность просмотра представлена через чекбокс.
3. Часть обучения модели: пользователь должен задать ID для модели, чтобы записать ее в системе. Опционально можно задать гиперпараметры для модели - сервис выводит полученные гиперпараметры для сверки.
4. Если все гиперпараметры введены верно, то ниже выводится сообщения о статусе обучения - если все успешно, то выводится “Модель обучена и сохранена”.
5. Для получения предсказания необходимо загрузить статистики команд в матче.
6. На загруженных данных сервис строит прогноз для каждого матча. Пользователю выводятся все загруженные статистики и предсказания исходов.