

Проект предназначен для предсказания решения Центрального банка по изменению ключевой ставки на основе анализа текста пресс-релизов.

1. Структура проекта:

- **Jupyter Notebook checkpoint_4_27_12_2024**. В данном ноутбуке выполнено получение, исследование и препроцессинг данных; обучение классификаторов и выбор лучшей модели. Здесь определяется лучшая модель для предсказания изменения ключевой ставки (1 — повышение, 0 — без изменений, -1 — понижение). Обученная модель сохранена в файл `best_model.joblib`.

- **API-сервис (main.py)**. Сервис, реализованный с помощью FastAPI, предоставляет доступ к лучшей из обученных моделей. Он принимает на вход один или несколько текстов пресс-релизов и возвращает предсказания изменения ставки. Помимо этого, сервис предоставляет информацию об используемой модели.

- **Streamlit-приложение (app.py)**. Сервис, который позволяет пользователям взаимодействовать с моделью через простой и понятный интерфейс.

2. Функционал API и Streamlit-приложения:

API (main.py):

Ручка `/one_predict` (метод POST):

- **Входные данные:** JSON объект с полем `text` (строка, содержащая текст пресс-релиза).
- **Выходные данные:** JSON объект с полем `target` (целое число: 1, 0, или -1, представляющее предсказанное изменение ставки).
- **Функционал:** принимает на вход текст одного пресс-релиза и возвращает предсказание изменения ставки от лучшей модели.

Ручка `/multiple_predict` (метод POST):

- **Входные данные:** JSON массив, где каждый элемент — это JSON объект с полем `text` (строка, содержащая текст пресс-релиза).
- **Выходные данные:** CSV-файл с двумя колонками: `text` и `target` (предсказанное изменение ставки: 1, 0, или -1).
- **Функционал:** принимает на вход список пресс-релизов и возвращает предсказания для каждого из них в виде CSV-файла.

Ручка `/model_info` (метод GET):

- **Входные данные** отсутствуют.
- **Выходные данные:** JSON объект с информацией о модели: название модели и её версия.
- **Функционал:** возвращает информацию о лучшей модели.

Streamlit приложение (app.py):

Прогноз по одному пресс-релизу: Пользователь вводит текст пресс-релиза в текстовое поле и получает прогноз изменения ставки.

Обработка CSV файла: Пользователь загружает CSV-файл с текстами пресс-релизов (колонка text). Доступны две операции:

1. Предсказать ставку: Приложение отправляет данные в API, получает прогнозы и предоставляет возможность скачать результаты в формате CSV.
2. Исследовать: Приложение предоставляет базовую статистику по загруженному CSV файлу: количество строк и столбцов, первые 5 строк, статистику по длине текстов (средняя, минимальная, максимальная), гистограмму и boxplot распределения длин текстов.

Информация о модели: По запросу отображает название и версию используемой модели, полученные из API.

3. Инструкция по использованию:

Инструкция по запуску приложения описана в [README.md](#).

Для прогноза изменения ставки по одному пресс-релизу необходимо скопировать текст пресс-релиза с решением по поводу ставки с сайта ЦБ (например, https://cbr.ru/press/pr/?file=20122024_133000key.htm). Вставить в окно ввода и нажать “Сделать прогноз”. Приложение выведет прогноз в человекочитаемом виде.

В приложение можно загрузить csv-файл с набором текстов пресс-релизов ЦБ для получения прогноза по каждому пресс-релизу. Для обучения модели тексты были собраны с сайта ЦБ. Часть таких текстов выложена в файл [data.csv](#). Для быстрой проверки работы приложения необходимо загрузить в него этот файл (или вручную собрать файл в аналогичном формате).

Для получения прогноза необходимо выбрать опцию “Предсказать ставку”, после этого будет доступен для скачивания csv-файл с результатами прогноза по каждому пресс-релизу.

Для исследования датасета необходимо выбрать опцию “Исследовать”. Приложение покажет базовую статистику по csv-файлу и распределение текстов по длине.

Нажатие по кнопке “Получить информацию о модели” выведет название модели, используемой для предсказания