

Задание 3.

Введение.

Необходимо разработать no-code платформу для анализа данных из различных источников с возможностью визуализации результатов. Такие решения существуют на рынке и пользуются популярностью (**Microsoft Azure, RapidMiner**), т.к. позволяют анализировать данные людям без навыков программирования (бизнес-аналитики, учёные из других областей) в данном случае реализуется прототип с базовой функциональностью.

Основная часть.

Схема решения, описание процессов, сценария использования, описание информационных потоков и предлагаемых технологий в решении (техническое решение) – до 5 баллов.

Онлайн-сервис позволяет пользователю осуществлять базовый анализ и визуализацию данных, взаимодействуя с графическим интерфейсом.

Исходный код представлен в репозитории

<https://github.com/yaggg/yandex-profi-se-semifinal-2022.git>

Для реализации использовался web-фреймворк **Flask**, библиотека для работы с таблицами **Pandas** и библиотека с алгоритмами машинного обучения **Scikit-Learn** и библиотека **Swagger** для документации **API**.

Основные способы взаимодействия пользователя с системой:

- Авторизация в системе
- Возможность загрузки данных из внешних источников (csv, xls, реляционная или noSQL база данных)
- Создание конвейеров для обработки данных с использованием графического интерфейса
- Предобработка данных (обработка отсутствующих значений, выбросов, нормализация данных)
- Применение алгоритмов машинного обучения, включая предсказание временных рядов (**time series prediction**), кластеризация (**clustering**), детектирование аномалий (**anomaly detection**)
- Визуализация результатов анализа данных

Процессы взаимодействия пользователя с такой системой принято описывать с использованием **UML sequence diagram**.

Описание алгоритма (алгоритмов) в виде блок-схемы – до 5 баллов.

Схема алгоритма авторизации:

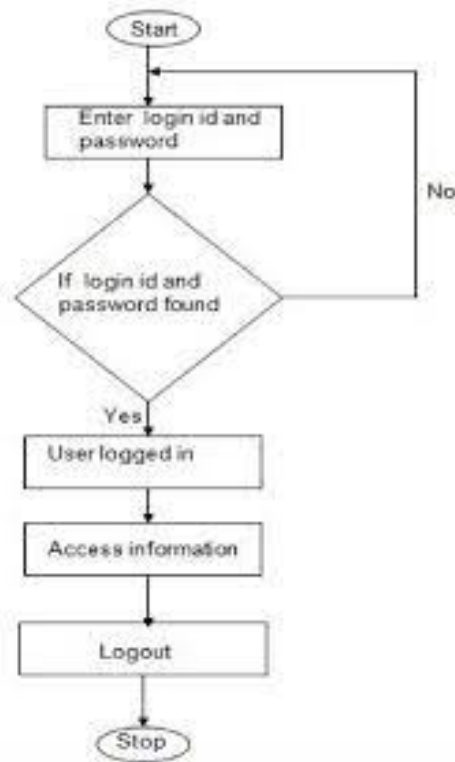
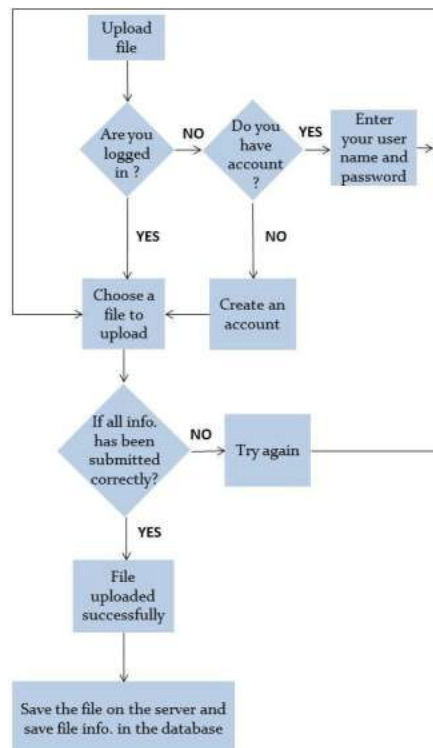


Схема загрузки данных:



Пользовательская документация – до 5 баллов.

Документация реализована на основе библиотеки **Swagger**:

auth Authorization

POST /auth/

Parameters

Try it out

No parameters

Responses

Response content type application/json

Code	Description
200	Success

psql PostgreSQL operation

POST /psql/{user_id}/read

Parameters

Try it out

Name	Description
user_id <small>required</small> string (path)	<input type="text" value="user_id"/>

Responses

Response content type application/json

Code	Description
200	Success

file File operation

POST /file/{user_id}/read

Parameters

Try it out

Name	Description
user_id <small>required</small> string (path)	<input type="text" value="user_id"/>

Responses

Response content type application/json

Code	Description
200	Success

Оценка точности и скорости работы алгоритма (алгоритмов) – до 5 баллов.

Заключение

Платформы no-code для анализа данных, хотя и популярны среди специалистов без навыков программирования, имеют ряд ограничений, связанных с тем, что их функциональность ограничена пользовательским интерфейсом, и при необходимости реализовать продвинутые алгоритмы или специфичные настройки алгоритмов приходится прибегать к классическому программированию.