

Тестовое задание Биокад (Зима 24)

Тестовое задание DS

Вам необходимо разработать вопросно-ответную систему на основе документа.

Аннотация к данным

Датасет в представлен в формате jsonl

Id - индекс наблюдения

passage - документ

text - текст документа

questions - вопросы к тексту

answers - ответы

label - класс ответа (0 - правильный, 1 - неправильный)

Задачи:

1. Провести эксплораторный анализ данных
2. Сформулировать гипотезы для решения задачи
3. Выбрать и обосновать метрики оценки качества
4. Разработать алгоритм, который будет отвечать на вопросы на основе предоставленных документов.
 1. Вход: Вопрос или Вопрос + Документ
 2. Выход: Ответ на вопрос
5. Провести эксперименты и оценить качество работы алгоритма
6. Добавить комментарии к коду, поясняющие его работу и особенности реализации.

Оформление результатов

1. Создать архив с файлами .ipynb, в которых представлены результаты вашего исследования.
2. Добавить в архив файл с requirements.txt
3. Опционально: добавить Dockerfile с указанием требований к системе
4. Архив с файлами должен содержать любые дополнительные материалы, необходимые для воспроизведения вашего решения на стороннем ПК