**Тестовое задание. ML-разработчик.**

**Формат результата:**

Jupyter notebook с выполненным заданием

**Критерии оценки:**

1. Качество выполнения задания и соответствие решения тексту задания
2. Качество кода (повод показать, что умеете писать код) и его организация
3. Обоснование выбора алгоритма или модели (там, где требуется)
4. Текстовое сопровождение кода (комментарии и тп)

**Часть 1. Первичный анализ данных и тематическое моделирование**

1. Скачать датасет с HuggingFace<https://huggingface.co/datasets/MonoHime/ru_sentiment_dataset/tree/main>
2. Выполнить разведывательный анализ текстовых данных (EDA)
3. Выполнить тематическое моделирование текстовых данных. Перед выполнением необходимо выбрать и обосновать выбор алгоритма для тематического моделирования.
4. Визуализировать полученные на предыдущем шаге темы
5. Предложить варианты дальнейшего улучшения тематического моделирования для имеющегося набора данных (перечислить и описать варианты, без реализации)

**Часть 2. Обучение классификатора**

1. Набор данных тот же
2. Обучить классификатор для определения тональности текста
   * Разбить данные на train/val/test части
   * Выбрать и обосновать выбор параметров для обучения классификатора
   * Выбрать и обосновать выбор метрики качества и функции потерь
   * Обучить модель
   * Оценить качество классификатора и скорость работы обученной модели на CPU и GPU
3. Предложить варианты для дальнейшего улучшения качества и скорости работы модели (перечислить и описать варианты, без реализации).

***Часть 3.\* Оптимизация классификатора***

*Факультативная часть (не является обязательной для выполнения)*

Выполнить оптимизацию модели для инференса на CPU, для этого нужно:

1. Выбрать и обосновать выбор способа оптимизации модели для инференса на CPU
2. Выполнить оптимизацию
3. Подготовить сводную таблицу, где будет представлено сравнение качества и скорости работы оригинальной модели и ее оптимизированной версии.

P.S. Под обоснованием выбора того или иного алгоритма (модели) предполагается краткое описание причин выбора в сравнении с альтернативными вариантами.