Graded Problemset: Sentiment Analysis

Your role is that of a data scientist working for a company that provides sentiment analysis services to restaurants. The company has provided you with a dataset of reviews (`restaurant-reviews.csv`) for different restaurants located in Kiel. The dataset contains the following columns:

- `name`: Name of the restaurant
- `restaurant_url`: URL of the restaurant
- `title`: Title of the review
- `text`: Text of the< review
- `rating`: Rating of the review (1-5)

Your task is to develop a sentiment analysis model that is able to predict the sentiment of a review on a scale from 1 - 5, based only on the text and title of the review.

Your boss asks you to try out 3 different approaches and evaluate their performances:

- **1. Heuristic approach**: The Python packages `vaderSentiment` and `textblob` provide heuristic approaches to sentiment analysis. Select one of these packages, research how the algorithm works and describe it concisely in your own words. Then apply it to the dataset.

- **2. Finetuned transformer model**: Several Python packages (e.g. `transformers`) provide transformer models that are already finetuned for sentiment analysis. Research at least one suitable finetuned transformer model and apply it to the dataset. Briefly explain how/on what basis you picked the candidate model.

- **3. Train a model yourself**: choose one of the following options

  - Train a suitable machine learning model based on *bag of words or TF-IDF*. You do not need to carry out a full-fledged hyperparameter optimization, but you should try out a few different options.
  - Train a suitable machine learning model based on *word embeddings*. You do not need to carry out a full-fledged hyperparameter optimization, but you should try out a few different options.
  - *Fine-tune a transformer model* yourself

Provide one table with the performances of the three approaches on a test data set. Then write a summary in which you (1) describe your main findings, (2) provide a recommendation to your boss, and (3) suggest ways how the performance could be further improved in the future.

Note: No further restrictions are given by your boss. There will be different ways of approaching the problem, e.g. in terms of preprocessing, model evaluation, or other aspects. You are free to make your choices, but you need to justify them