Seif Eldin Amr Mostafa Yassine

20P2006

PROJECT 1

IMDP MOVIES Review Analysis

1) Overview

This project aims to build and optimize an LSTM (Long Short-Term Memory) neural network for sentiment analysis on movie reviews. The primary objective is to classify reviews into positive or negative sentiments. The process involves several key steps:

2) Data Loading and Preprocessing:

- **Dataset:** The dataset, consisting of movie reviews and their sentiment labels, is loaded in chunks to manage large data sizes efficiently.
- **Label Encoding:** The sentiment labels are encoded into numerical values (0 for negative, 1 for positive) using `LabelEncoder`.
- **Tokenization:** The text reviews are tokenized using `Tokenizer` to convert words into numerical sequences. The vocabulary size is limited to 5,000 most frequent words.
- **Padding:** Sequences are padded to ensure uniform input length for the model, with a maximum length of 200 words.

3) Model Definition and Hyperparameter Tuning:

- Model Architecture: An LSTM-based neural network is constructed with the following layers:
- ‘Embedding’ layer to convert tokenized sequences into dense vectors.
- ‘LSTM’ layer to capture temporal dependencies in the sequences.
- ‘Dropout’ layer to reduce overfitting.
- ‘Dense’ output layer with a sigmoid activation function for binary classification.
- Hyperparameters: Using Optuna, various hyperparameters are optimized:
- Number of LSTM units (between 50 and 100).
- Dropout rate (between 0.2 and 0.5).
- Optimizer (either 'adam' or 'rmsprop').

4) Model Training and Evaluation:

- **Training:** The model is trained for 5 epochs with a batch size of 64. The training process is conducted with validation data to monitor performance.
- **Evaluation:** The model’s performance is evaluated based on validation accuracy. The best hyperparameters are selected to train a final model.
- **Final Training:** The final model is trained again with the best-found hyperparameters.

5) Sentiment Prediction on Custom Reviews:

- **Custom Reviews:** A set of custom text reviews is preprocessed similarly to the original dataset.
- **Prediction:** The final model predicts sentiments for these custom reviews.
- **Decoding Predictions:** Predictions are mapped back to sentiment labels ('positive' or 'negative') based on the model's output.

6) Results Export:

- **CSV Export:** The results, including custom reviews and their predicted sentiments, are saved into a CSV file for further use or analysis.

We managed to reach 90% accuracy. Sample of the output of the model on unseen custom data:

index	review	predicted_sentiment
0	The film was an absolute delight from start to finish. The plot was engaging and the characters were well-developed. The cinematography was stunning, and the soundtrack perfectly complemented the mood of the movie. The actors delivered their roles with such authenticity that it was easy to get lost in the story. This movie is a must-watch for anyone who loves a heartwarming and well-crafted story. I left the theater feeling uplifted and inspired. Highly recommended for a great cinematic experience!	positive
1	What a fantastic film! The storyline was incredibly gripping, and the character development was top-notch. The performances were stellar, and each scene was beautifully shot. The pacing was perfect, keeping me engaged throughout. The movie managed to combine humor with emotional depth in a way that felt genuine and impactful. It's the kind of film that stays with you long after the credits roll. If you're looking for a memorable and enjoyable movie night, this one is a winner!	positive
2	This movie exceeded all my expectations. The direction was impeccable, and the script was both witty and profound. The chemistry between the lead actors was palpable, adding depth to their performances. The visual effects were mesmerizing and enhanced the storytelling in a meaningful way. The film's soundtrack was also noteworthy, adding an extra layer of emotion to the scenes. Overall, it's a well-rounded film that offers a rich and satisfying viewing experience.	positive
3	I was really disappointed with this movie. The plot was convoluted and failed to hold my interest. The acting was mediocre at best, and the film's attempt at emotional depth felt shallow. The special effects, while flashy, did little to enhance the story. The film's humor fell flat, and the dramatic moments felt contrived. Overall, it was an underwhelming experience that didn't deliver on its promises.	negative
4	This movie missed the mark in several ways. The storyline was clichéd and predictable, offering nothing new or engaging. The characters lacked depth, making it difficult to connect with them. The pacing was erratic, with some scenes dragging on unnecessarily. The film's attempts at humor were forced, and the dramatic elements didn't resonate. It's a film that could have benefited from better writing and direction.	negative
5	I was thoroughly impressed by this film. The storyline was original and kept me hooked from beginning to end. The actors delivered powerful performances, and the film's pacing was spot-on. The cinematography was exceptional, with each frame carefully crafted to enhance the narrative. The movie balanced humor and drama beautifully, making it an enjoyable watch. It's a testament to the skill and creativity of the filmmakers. Definitely worth watching for anyone who appreciates high-quality cinema.	positive
6	This film was a delightful surprise. The plot was refreshing and the execution was flawless. The characters were well-written and the performances were heartfelt. The attention to detail in the set design and costumes added authenticity to the story. The film's ability to evoke genuine emotion and keep me entertained throughout is a mark of its quality. It's a wonderful piece of cinema that I would gladly recommend to friends and family.	positive
7	I found this movie to be quite disappointing. The plot was predictable and lacked originality. The character development was superficial, and the performances were underwhelming. The pacing felt sluggish, and the film seemed to drag on longer than necessary. Despite the high production values, the movie failed to engage me on an emotional level. Overall, it was a forgettable experience that didn't live up to its potential.	negative
8	This film was a letdown. The story was confusing and poorly executed, leaving me more frustrated than entertained. The characters felt one-dimensional, and the dialogue was often forced and unnatural. The pacing was uneven, and the film struggled to maintain any momentum. While the visuals were decent, they couldn't compensate for the lackluster script and uninspired performances. It's a movie that might be better suited for a background watch rather than a focused viewing.	negative
9	I was not impressed with this film at all. The narrative was disjointed, and the character arcs felt incomplete. The acting was lackluster, and the dialogues were often cringeworthy. The film tried to tackle too many themes at once, leading to a muddled and incoherent story. The visual effects were decent but couldn't salvage the overall experience. It's a forgettable film that didn't make much of an impact.	negative

PROJECT 2

WEATHER PREDICTION

1. Overview

This project aims to predict weather conditions based on historical weather data using advanced sequential models. The data consists of various weather attributes such as temperature, precipitation, wind speed, and the overall weather condition. The objective is to build and compare predictive models using Hidden Markov Models (HMM) and Gated Recurrent Units (GRU), assessing their performance in weather forecasting.

2. Data Preparation

a) **Data Loading and Preprocessing**:

- The dataset, sourced from a weather file, includes columns for weather conditions, temperature, precipitation, and wind speed.
- Weather conditions are encoded into numerical values, and additional features like season, temperature, precipitation, and wind speed are also encoded.

b) **Feature Engineering**:

- Weather data is encoded into categorical values to facilitate model training.
- Features are derived from the date and weather attributes to enrich the dataset for model training.

c) **Data Splitting**:

- The data is split into training and testing sets to evaluate the performance of the predictive models.

3. Model Development

I. **Hidden Markov Model (HMM)**:

- An HMM is used to model the sequential nature of the weather data. The model is trained on historical weather sequences to predict future weather conditions.
- Parameter tuning is performed using Optuna to optimize model accuracy.
- The model's performance is evaluated by predicting weather conditions on the test dataset and comparing predictions with actual values.

II. **Gated Recurrent Units (GRU)**:

- A GRU-based Recurrent Neural Network (RNN) is built to handle the sequential data. GRUs are chosen for their effectiveness in capturing temporal dependencies in sequential data.
- Hyperparameter tuning is performed using Keras Tuner to optimize the GRU model's performance.
- The GRU model is evaluated on the test set, and its performance is compared with the HMM model.

4. Performance Evaluation

- **Accuracy Measurement**: Both models are evaluated based on their accuracy in predicting weather conditions. The accuracy of each model is compared to assess which model provides better forecasting performance.
- **Visualization**: A bar chart is used to visually compare the accuracies of the GRU and HMM models, highlighting their performance in weather prediction.

5. Benefits of the Project

- **Improved Weather Forecasting**: By leveraging sequential models, the project aims to provide more accurate weather forecasts, which can be valuable for various applications such as agriculture, event planning, and daily decision-making.
- **Model Comparison**: The project offers insights into the performance of different sequential models (HMM vs. GRU), helping in understanding their strengths and weaknesses in the context of weather prediction.
- **Data-Driven Decision Making**: Accurate weather predictions can aid in better decision-making and planning, enhancing preparedness for weather-related events.

6. Accuracy comparison