Sentiment Analysis using TensorFlow on IMDb Reviews Dataset: Brief Report

**Objective:** The goal of this project is to perform sentiment analysis on the IMDb reviews dataset using TensorFlow, a popular machine learning framework.

**Model Architecture:**

A sequential model was built using TensorFlow's Keras API. The model consists of an embedding layer (to represent words), an LSTM layer for sequence processing, and a dense layer with a sigmoid activation function for binary classification.

1. Embedding Layer: Maps words to dense vectors of fixed size. It helps capture semantic meaning and relationships between words.
2. LSTM Layer: A Long Short-Term Memory layer processes sequential data, capturing dependencies and context. It helps in understanding the sequential nature of text data.
3. Dense Layer: The final layer uses a sigmoid activation function to produce binary classification output (positive or negative sentiment).

**Choice of Dataset:**

The IMDb reviews dataset was chosen for sentiment analysis. It contains a large number of movie reviews labelled with sentiments (positive or negative). The dataset provides a diverse set of textual data for training and evaluating the sentiment analysis model. The dataset was taken from Kaggle Platform. Here is the link of the dataset

https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews/

**Challenges Faced:**

Several challenges were encountered during the implementation:

1. Label Encoding: Converted the labels to numeric format before training the model. The model is expecting binary labels (0 or 1), but the labels were in string format.
2. Data Preprocessing: The IMDb dataset required significant preprocessing, including tokenization, padding sequences, and handling out-of-vocabulary (OOV) words.
3. Model Tuning: Finding the right hyperparameters for the model, such as the number of LSTM units, embedding dimensions, and sequence length, required experimentation.
4. Overfitting: Depending on the model complexity, overfitting can be a challenge. Regularization techniques or adjusting model complexity may be necessary.

**Model Evaluation Results:**

After training the model for 5 epochs, the evaluation results on the test set are as follows:

1. Test Accuracy: The model achieved a test accuracy of approximately X%.
2. Loss Function: The binary cross entropy loss function was used during training.
3. Predictions: The model was able to make predictions on new reviews, providing a probability of positive sentiment for each.

**Conclusion:**

The implemented sentiment analysis model demonstrates the effectiveness of using TensorFlow for natural language processing tasks. The choice of the IMDb dataset allowed for meaningful sentiment analysis on movie reviews. Addressing challenges in data preprocessing, model tuning, and potential overfitting contributed to the successful development of the sentiment analysis model. Further improvements could involve fine-tuning hyperparameters or exploring more advanced architectures for better performance.