Bi-LSTM for Sentiment Analysis on Amazon Product Reviews

BANOTH JAGADEESH - 21010102
Semester - VII

Dept. of Computer Science and Engineering
Indian Institute Of Information Technology, Manipur

Supervisor: Dr. Kishorjit Nongmeikapam
October 19, 2024

Introduction

- This project leverages a Bidirectional Long Short-Term Memory (BiLSTM) model to classify reviews as positive or negative.
- BiLSTM captures context in both forward and backward directions, improving sentiment classification accuracy.

Problem Statement

- Understanding the sentiment of customer reviews is essential for making informed business decisions.
- Classical models struggles to understand complex context, especially negations and ambiguous phrases.
- BiLSTM is employed to address this issue by learning both directions of the text sequence.

Supervisor: Dr. Kishorjit Nongmeikapam Octob

BANOTH JAGADEESH – 21010102Semester – VII (Dept. Bi-LSTM for Sentiment Analysis on Amazon Product F   3 / 12

Data Preprocessing

- Cleaning the dataset by removing special characters, stop words, and performing tokenization.
- Converting the text into numerical vectors using techniques like Word Embedding (TF-IDF or Word2Vec).
- Splitting the dataset into training and testing sets.

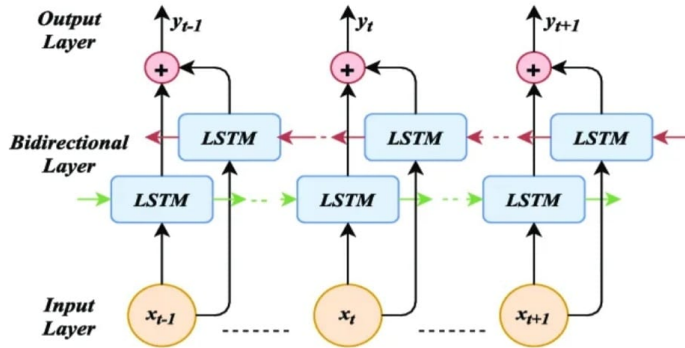Supervisor: Dr. Kishorjit Nongmeikapam Octob

BANOTH JAGADEESH – 21010102Semester – VII (Dept. Bi-LSTM for Sentiment Analysis on Amazon Product F 4 / 12

Bi-LSTM Diagram

Figure: Bi-LSTM Diagram

Supervisor: Dr. Kishorjit Nongmeikapam Octob

Implementation

Step 1: Dataset Preparation :

Step 2: Model Building :

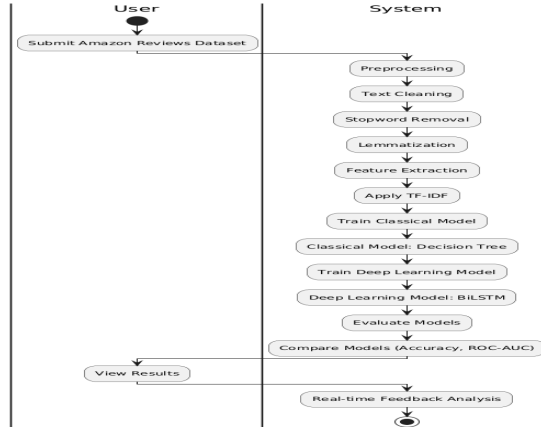Step 3: Model Training :

Step 4: Evaluation :

Activity Diagram

Figure: Activity Diagram

Supervisor: Dr. Kishorjit Nongmeikapam Octob
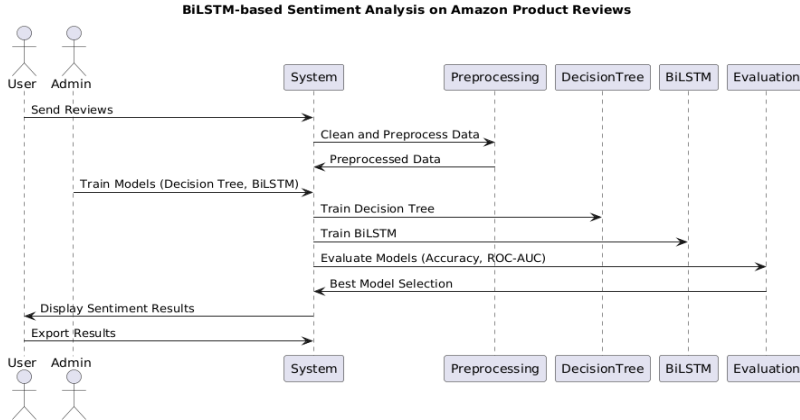
Use Case Diagram

Figure: Use Case Diagram

Model Training and Evaluation

- The BiLSTM model is trained on labeled Amazon reviews, with the following evaluation metrics:
  - **Accuracy**
  - **Precision, Recall, F1-Score**
  - **ROC-AUC Score**
- The BiLSTM outperformed classical models such as **Decision Tree**.

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Decision Tree	80.2%	79.1%	80.8%	79.9%	82.7%
BiLSTM	89.4%	88.7%	89.5%	89.1%	92.3%

Figure: Performance Comparison Model

Challenges and Solutions

- **Data Imbalance**: Handled using techniques like undersampling to balance positive and negative reviews.
- **Training Time**: Reduced using GPU acceleration for faster training of the BiLSTM model.
- **Overfitting**: Managed with early stopping and dropout layers to prevent overfitting during training.

Conclusion

- BiLSTM showed **superior performance** in sentiment analysis.
- The methodology can be applied to other **NLP tasks**.
- Future work: explore new **hyperparameters** and models.

Thank You!

BANOTH JAGADEESH – 21010102Semester – VII (Dept. Bi-LSTM for Sentiment Analysis on Amazon Product F 12 / 12

Supervisor: Dr. Kishorjit Nongmeikapam Octob