Programming Assignment 02 – LSTM Models for Tweet Classification

Name : D.R.K.Dhanushka
Reg. Number : S/19/809

---

1. Introduction

This report presents my work on classifying tweets as either **personal health mentions** or **non-personal health mentions** using deep learning models. Given the sequential nature of textual data, I employed **Long Short-Term Memory (LSTM)** and **Bidirectional LSTM (Bi-LSTM)** architectures. The objective was to evaluate and compare their effectiveness in learning contextual relationships in tweets for binary classification.

2. Dataset Description

The dataset contains tweets that have been labeled as:

- **1**: Personal health mentions
- **0**: Non-personal health mentions

Text preprocessing was performed to clean and tokenize the data. Each tweet was then converted into sequences and padded to a fixed length.

3. Model Architecture

LSTM Model

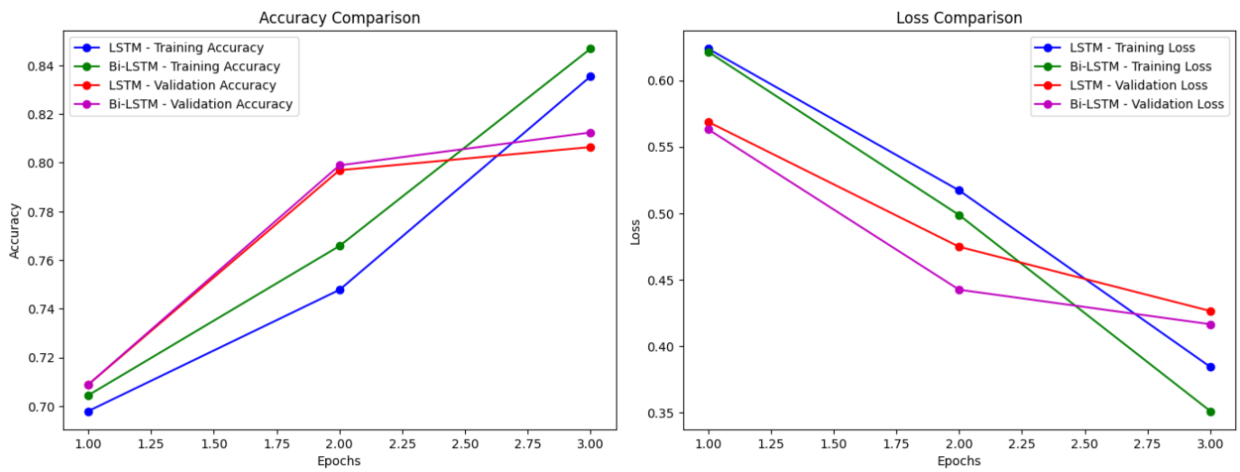
- **Embedding Layer**: Converts words into dense vector representations.
- **LSTM Layer**: Learns dependencies in tweet sequences.
- **Dense Layers**: For classification (sigmoid activation in the final layer).
- **Loss Function**: Binary Crossentropy.
- **Optimizer**: Adam.

Bi-LSTM Model

- Same structure as the LSTM model, except the LSTM layer is replaced by a **Bidirectional wrapper**, allowing the model to capture both past and future context in the sequence.

4. Results

Model	Final Training Accuracy	Final Validation Accuracy	Test Accuracy	Final Training Loss	Final Validation Loss	Best Epoch
LSTM	0.835460	0.806403	81.146803	0.384505	0.426537	3
Bi-LSTM	0.846847	0.812406	82.137496	0.351269	0.416574	3



Both models showed consistent convergence, but Bi-LSTM converged slightly faster and achieved lower validation loss and higher validation accuracy.

5. Discussion

- Both models performed well in terms of training and validation accuracy, with the Bi-LSTM slightly outperforming the standard LSTM across all key metrics. While the difference in test accuracy was modest (approximately 1%), the **Bi-LSTM model** consistently showed better generalization ability, as evidenced by slightly lower validation and test losses.

- The LSTM model learned patterns in the data effectively, achieving an accuracy of over 83% on the training set. However, it seemed to plateau a bit earlier, with signs of slight overfitting after the best epoch (epoch 3), where validation loss began to increase.

- The **Bi-LSTM model**, in contrast, maintained a more stable learning curve and showed improved validation loss and accuracy even toward later epochs, suggesting it was more efficient at learning contextual patterns without overfitting too quickly.

Through this assignment, I gained hands-on experience in implementing and comparing sequence models for text classification tasks. The **Bi-LSTM model** proved to be more effective due to its ability to capture richer contextual dependencies.