[NN'23] Arabic Sentiment Analysis (NLP)

Name	Department	ID
1. Heba Hossam Mohamed Ali	CS	20201700958
2. Bassant AbdElraouf Elsayed	CS	20201700189
3. Wessam Samir Mohamed Elnawawy	CS	20201700973
4. Sondos Tarek Abdel Mageed Ali	CS	20201700370
5. Osama Ali AbdElsamea Elsharqawy	CS	20201700106
6. Mohamed Ali Mahmoud Mohamed	CS	20201700719

1. Introduction:

Presenting an analysis and comparison of text classification models for sentiment analysis on a dataset containing textual reviews.

Three distinct models were implemented and evaluated:

1. Transformer Model
2. LSTM Model
3. RNN Model

Then provide insights into the preprocessing steps, model architectures, and their corresponding accuracies.

2. Data Description:

The dataset consists of textual reviews with associated sentiment labels. The task involves predicting the sentiment (positive, negative, or neutral) based on the review text.

3. Preprocessing:

3.1 Common Preprocessing Steps:

- Text Cleaning: Removal of punctuation and special characters show in:
  - **(remove_punctuation)**
  - **(remove_numbers)**
  - **(remove_unusual_sequences)**
  - **(remove_non_arabic)**
  - **(remove_emojis)**

- Tokenization: Using the nltk library to split text into tokens.
- Stopword Removal: Elimination of common words in both Arabic and English.
- Lemmatization: Reduction of words to their base form.

3.2 Model-Specific Preprocessing:

- For the Transformer Model: tokenization and padding were applied.
- For LSTM and RNN Models: label encoding used to convert sentiment labels into numerical values.

4. Model Architectures:

4.1 Transformer Model:

The Transformer model utilized a multi-layer architecture with attention mechanisms, positional encoding, and feedforward layers. It was trained for sentiment analysis with a focus on capturing contextual information.

4.2 LSTM Model:

The LSTM model featured an embedding layer, followed by a combination of Conv1D, MaxPooling1D, and two LSTM layers. It employed dropout for regularization and utilized SoftMax activation for multi-class classification.

4.3 RNN Model:

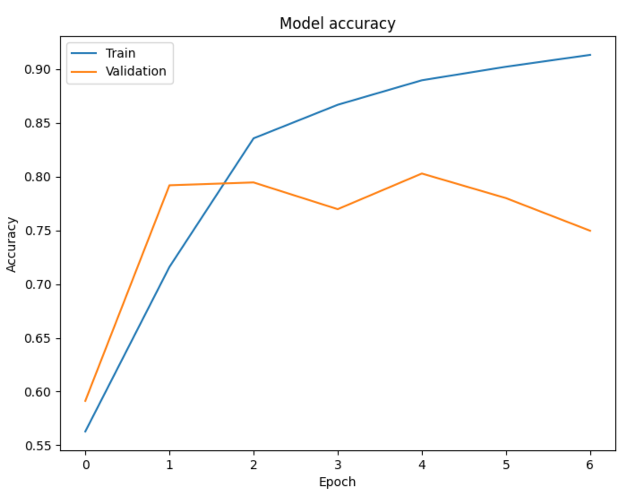
The RNN model incorporated embedding layer, followed by a SimpleRNN layer with dropout, a dense layer, and flattening. It used SoftMax activation for classification.

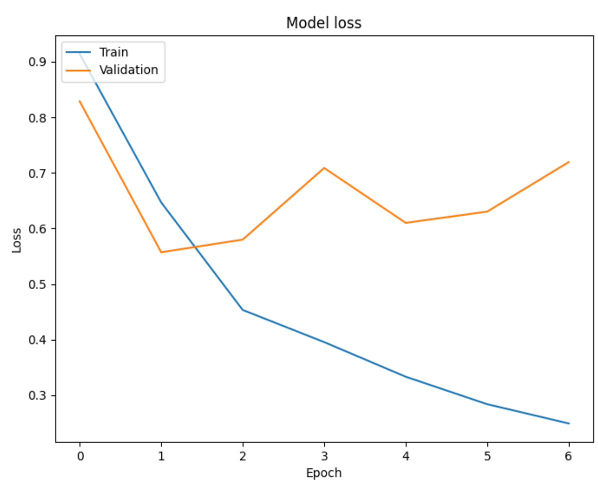
5. Model Training and Evaluation:

Each model was trained using a portion of the labeled dataset and evaluated on a validation set. Training involved minimizing categorical cross-entropy loss, and early stopping was applied to prevent overfitting.

5.1 Transformer Model:

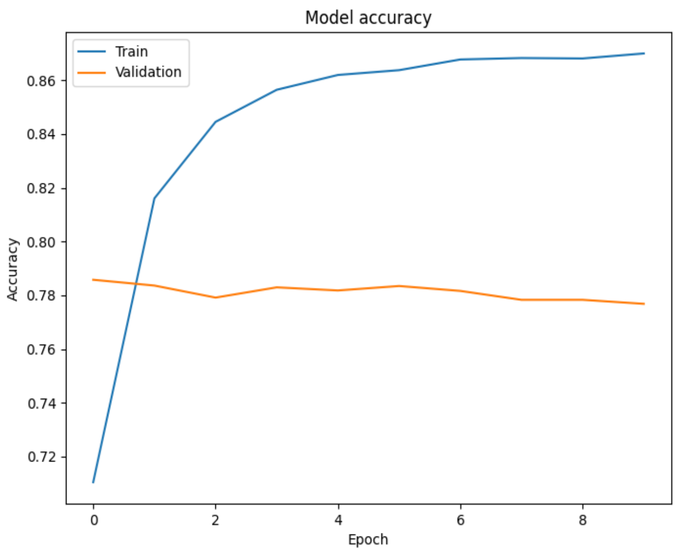
The Transformer model achieved a validation accuracy of approximately 80.26% after 30 epochs.

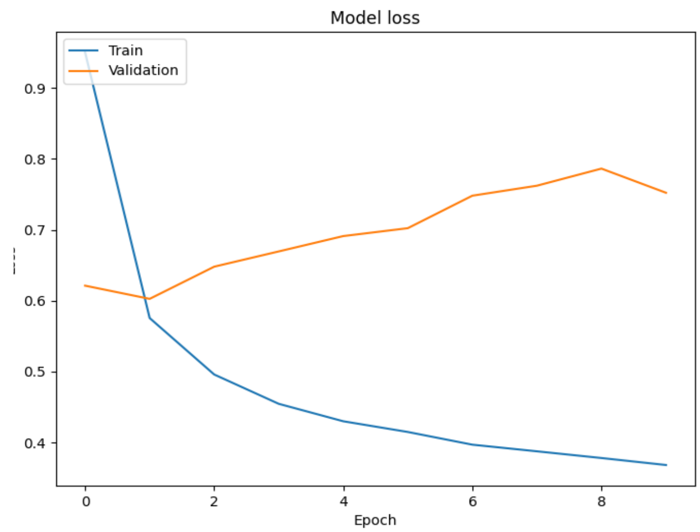




5.2LSTM Model:

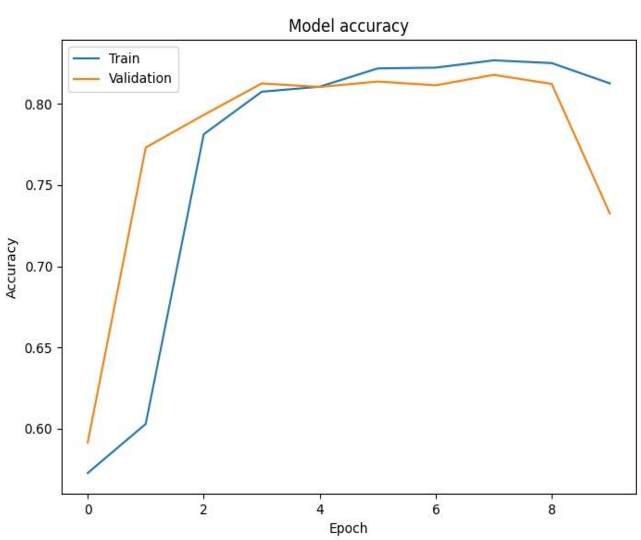
The LSTM model achieved a final validation accuracy of approximately 79.42% after 10 epochs

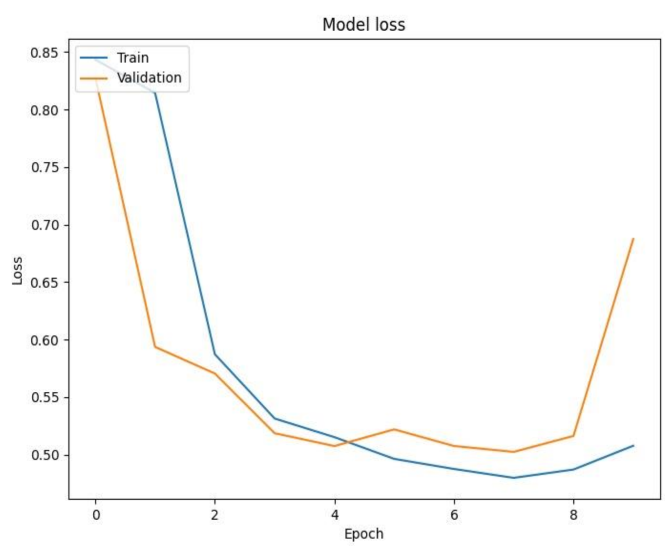




5.3 **RNN Model:**

The RNN model achieved a final validation accuracy of approximately 80.3% after 10 epochs





6. Conclusion:

In conclusion, the Transformer Model demonstrated competitive performance, while the LSTM and RNN models provided satisfactory results as well.