Classification Report

Bhosale Rushikesh Anant

September 17, 2025

1 Introduction

This report summarizes the performance of three different neural network architectures – a simple Recurrent Neural Network (RNN), a Long Short-Term Memory (LSTM) network, and a Transformer Encoder model – on a sentiment analysis task using the Corona_NLP dataset. The models were implemented from scratch and evaluated on a separate test set. The key metrics presented are accuracy and F1-score for each sentiment class.

2 Model Performance

2.1 RNN Classification Report

The RNN model achieved a macro F1-score of 0.37, indicating low performance in correctly classifying the sentiment of the tweets.

```
                    precision    recall  f1-score   support

           Neutral       0.37      0.40      0.39       841
          Positive       0.37      0.37      0.37       943
Extremely Negative       0.39      0.31      0.35       592
          Negative       0.39      0.39      0.39       768
Extremely Positive       0.37      0.39      0.38       654

          accuracy                           0.37      3798
         macro avg       0.38      0.37      0.37      3798
      weighted avg       0.37      0.37      0.37      3798
```

2.2 LSTM Classification Report

The LSTM model showed a significant performance improvement with a macro F1-score of 0.74, demonstrating its ability to handle long-range dependencies in the text data.

```
                    precision    recall  f1-score   support

           Neutral       0.77      0.72      0.74       841
          Positive       0.75      0.80      0.77       943
Extremely Negative       0.73      0.72      0.73       592
          Negative       0.75      0.73      0.74       768
Extremely Positive       0.72      0.75      0.73       654

          accuracy                           0.74      3798
         macro avg       0.74      0.74      0.74      3798
      weighted avg       0.74      0.74      0.74      3798
```

2.3 Transformer Classification Report

The Transformer model achieved the best results with a macro F1-score of 0.75. This indicates that its attention-based architecture is highly effective for this text classification task.

```
                    precision    recall  f1-score   support

           Neutral       0.76      0.76      0.76       841
          Positive       0.76      0.78      0.77       943
Extremely Negative       0.75      0.71      0.73       592
          Negative       0.74      0.73      0.74       768
Extremely Positive       0.76      0.79      0.78       654

          accuracy                           0.75      3798
         macro avg       0.75      0.75      0.75      3798
      weighted avg       0.75      0.75      0.75      3798
```

3 Conclusion

Based on the evaluation metrics, the Transformer model demonstrated the highest performance in sentiment classification, followed closely by the LSTM. The simple RNN model's performance was significantly lower, highlighting the importance of more advanced architectures like LSTMs and Transformers for complex NLP tasks.