**Twitter Sentiment Analysis**

Submitted by : Vibha Pandey

Email: vibhapandey2001@gmail.com

**Problem Statement:**

Build model to predict the twitter sentiments. Understand the dataset and cleanup.

**Approach to the problem:**

I used SimpleRNN model to understand the working.

In this model I performed my training and got accuracy: 0.7381 and in testing I got accuracy: 0.5682.

This was a classification problem and my platform setup were:

1. Platform used : Python 3.10 with Keras.
2. Model : SimpleRNN
3. Dataset used : Twitter Tweets Sentiment Dataset from Kaggle.

**Pre-processing Steps:**

The following pre-Processing steps were carried out on the data:

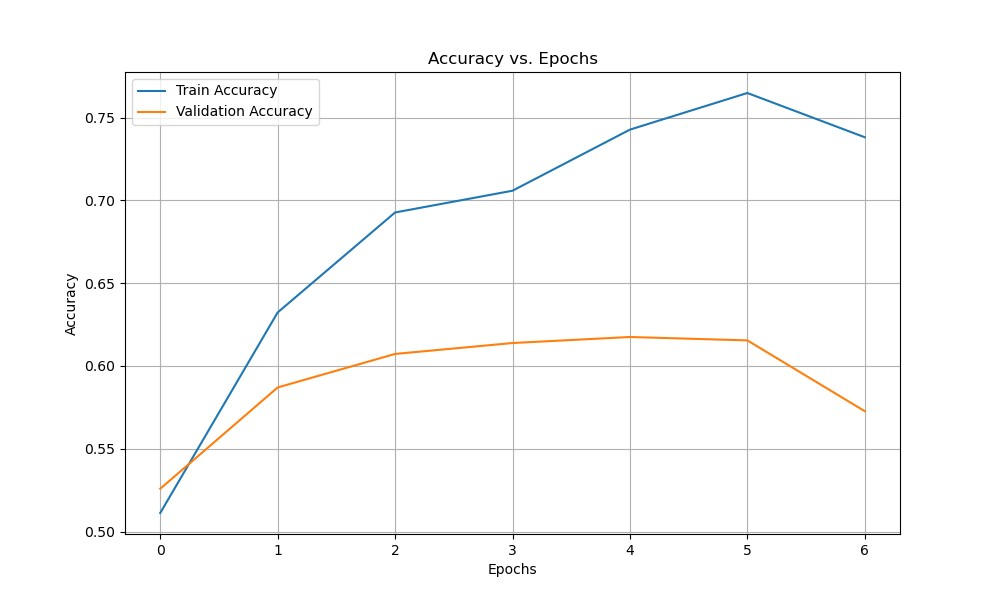
1. Tokenization : For separating sentences into individual words
2. Padding : Making the input sizes equal
3. Text-cleanup: using Regular expression (RE) to remove unnecessary texts from dataset.

**Hyperparameter Optimization:**

I conducted various hyperparameter optimization to determine the best epoch value and the best batch size. Each test was run 12 independent times.

1. Batch size: Batch\_sizes were varied from 10 to 64. A value of 32 to 64 turned out to be the best one.
2. RNN units: RNN units were varied from 25 to 125. I found out it was inversely proportional to embedding units.
3. Embedding Units: with low RNN units the embedding units were set to 200.

The following is the loss graph and accuracy graphs. The training stopped at 7 epochs.



A graph with lines and numbers

Description automatically generated

**Future work:**

To apply different models like LSTM, GRU etc and see how the model works.