Sentiment Analysis Project Report

By Tanishq Selot, 3rd year (IIT Indore) for TrueFoundry

Model/Metric	Training Accuracy	Validation Accuracy	Testing Accuracy
LSTM + Word2Vec	0.9815	0.9647	0.9687
LSTM + GloVe	0.9481	-	0.4880
BERT	0.9966	0.9831	-

Steps:

1. **Data Reader Class & Text Pre-processing:**
   1. Cleaning tags/mentions
   2. Cleaning non-alphanumeric characters
   3. Cleaning URLs
   4. Cleaning Punctuations
   5. Cleaning repeating characters
   6. Cleaning numbers
   7. Stemming

   Resampling because of class imbalance in the favour of "Negative" class.

2. **Tokenization** – Converting sentences into list of strings

3. **Word Embedding** – Representing words in a form which can be understood by Deep Learning Models. Used Word2Vec because of its effectiveness in representing a relatively small vocabulary. Fitting the model on all text. Later, using it for the embedding layer of the DL model.

4. **Train-test split & Padding** – Making the length of all the input vectors same.

5. **Model class** – Building, compiling, summarising, and fitting the model.

6. **Model Evaluation**

7. **Inference class & testing** a random example.