Comparing the Powerhouses: A Battle of Three-Layered Neural Networks with Word Embeddings (FastText, Word2Vec, ...) and **fine-tuned BERT** Model for Text Classification

Sayyed Ahmad Hosseini, Sithursan Sivasubramaniam

School of Engineering

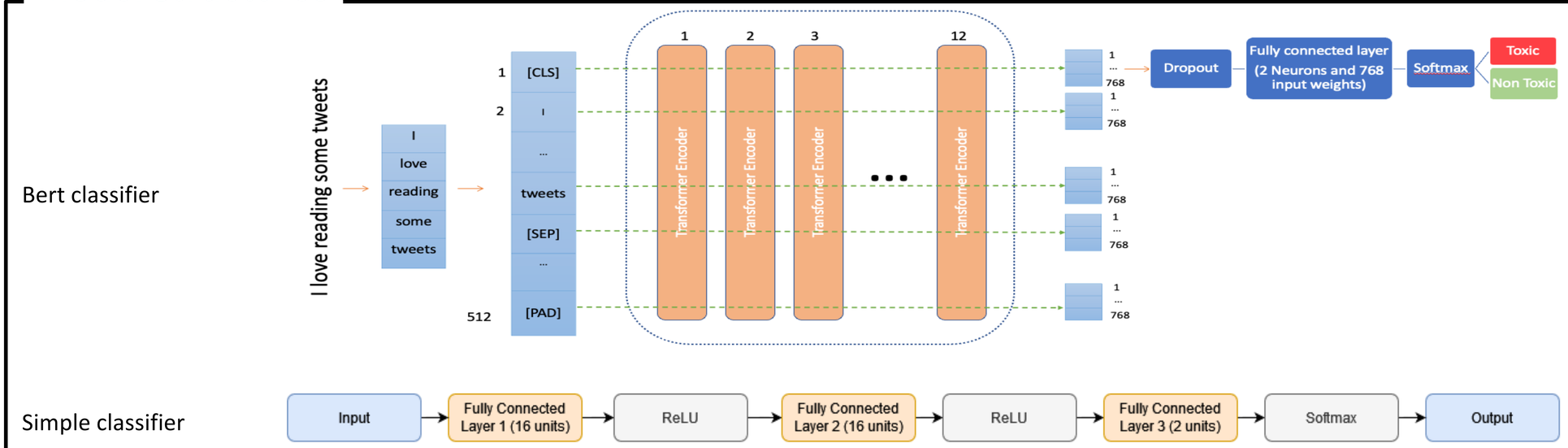
Introduction

In this project, we aim to investigate the efficacy of various embedding techniques in the classification of hate speech. We will be comparing different embedding methods to assess their impact on the accuracy and efficiency of hate speech classification models. Our goal is to offer valuable insights into the performance of these embedding methods and provide a framework for future research in this area.

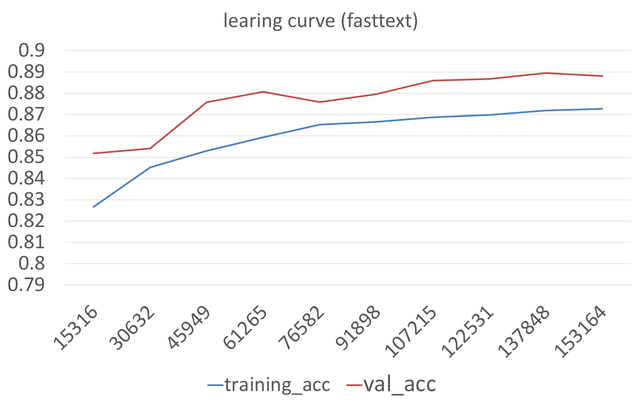
Research Questions

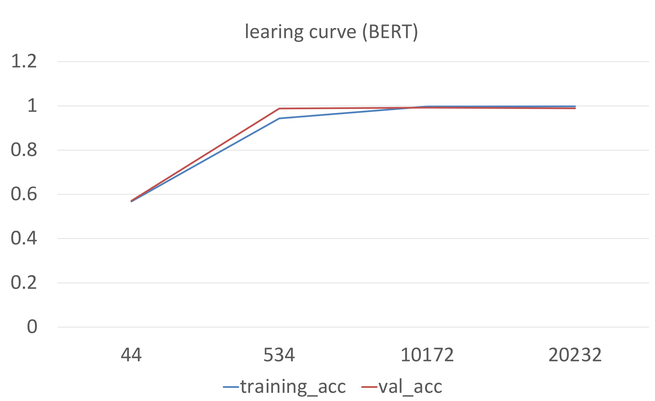
1. Which embedding method(s) perform best for hate speech classification, and why?
2. How do different embedding methods affect the accuracy , precision, recall and F1-score of hate speech classification models?
3. How does it compare to the performance of Bert embeddings with its own preprocessing methods?

Model Structures



Results





Embeddings	Accuracy	Precision	Recall	F1
tfidf	0.9003	0.9023	0.9003	0.9003
boW	0.8959	0.8977	0.8959	0.896
spacy	0.8971	0.8983	0.8971	0.8971
word2vec_avg	0.8982	0.8996	0.8982	0.8982
word2vec_sum	0.9016	0.903	0.9016	0.9016
word2vec_concat	0.8973	0.8994	0.8973	0.8972
fasttext	0.9025	0.9042	0.9025	0.9025
Bert*	0.980	-	-	-

* was trained with 20232 data points and tested with 2681 Datapoints

Conclusions

- The differences in the evaluation metrics between the different embeddings appear to be relatively small. The **choice of embedding method does not have a large impact** on the overall performance of the model.
- The Bert model's training accuracy and validation accuracy increase significantly as the number of training points increases, indicating that the model is learning better with more data**. And as test accuracy shows fine-tuned Bert model outperforms other models by about 0.8 percent.**
- Further analysis, such as individual example prediction examination and t-SNE visualization of the embeddings, can provide a deeper understanding of the differences between the embeddings and their impact on the model's performance.