CS772: DL4NLP
Assignment Evaluation

Team Id
Manoj Bhadu, 170010036
Piyush Onkar, 160050012

Date: 09/03/2021

Problem Definition

Sentiment Analysis using various word embedding techniques such as Word2Vec, GloVe and FastText.

Compare the performance of model obtained after various word embedding.

Error analysis of the models obtained.

System Architecture

Libraries Used for this task:

- Pandas
- Numpy
- Tensorflow
- Keras
- NLTK
- Gensim
- Sklearn
- Seaborn
- Matplotlib

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 64)                160064
_________________________________________________________________
dense_1 (Dense)              (None, 5)                 325
=================================================================
Total params: 160,389
Trainable params: 160,389
Non-trainable params: 0
```

Table 1: Layers vs embeddings (sigmoid)

No. of hidden layers	Word2vec	GloVe	fastText
0	Precision: 0.49 Recall: 0.59 F1-score: 0.48 Accuracy: 0.593	Precision:0.50 Recall: 0.60 F1-score: 0.50 Accuracy: 0.607	Precision: 0.48 Recall: 0.59 F1-score: 0.46 Accuracy: 0.59
1 (64)	Precision: 0.48 Recall: 0.6 F1-score: 0.48 Accuracy: 0.60	Precision: 0.51 Recall: 0.61 F1-score: 0.51 Accuracy: 0.61	Precision: 0.48 Recall: 0.59 F1-score: 0.47 Accuracy: 0.59

Table 2: Data imbalance addressed (sigmoid)

	word2vec	GloVe	fastText
<technique-name>			
Under sampling	0.53	0.54	0.52
Over sampling	0.532	0.55	0.51

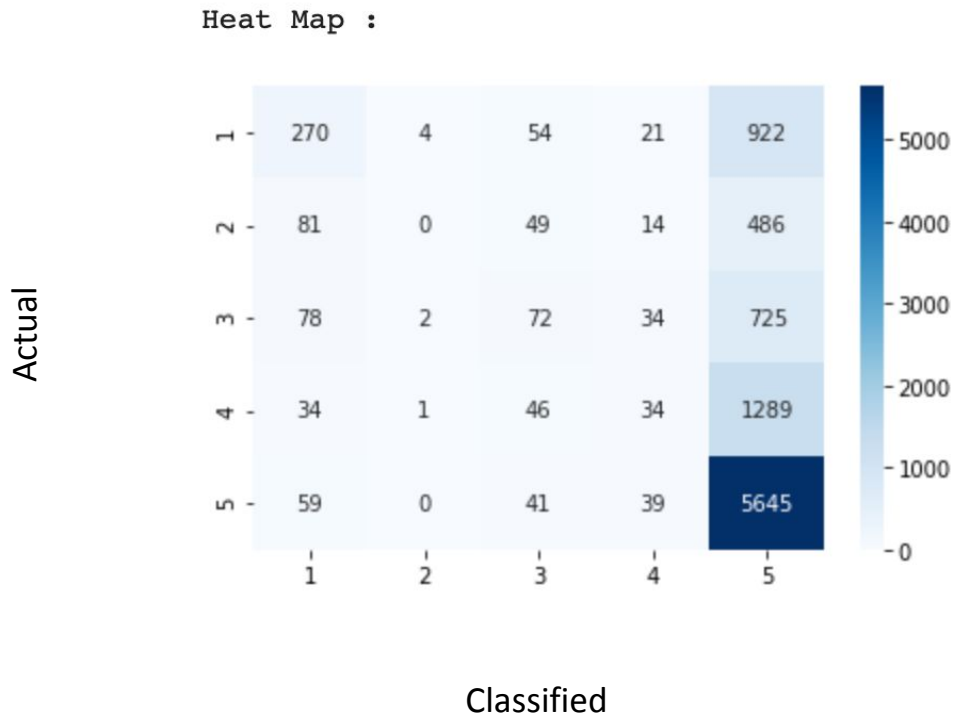
Table 3: Layers vs Embeddings (ReLU)

No. of hidden layers	word2vec	GloVe	fastText
0	Precision: 0.49 Recall: 0.59 F1-score: 0.48 Accuracy: 0.591	Precision: 0.50 Recall: 0.60 F1-score: 0.49 Accuracy: 0.603	Precision: 0.48 Recall: 0.59 F1-score: 0.46 Accuracy: 0.585
1 (64)	Precision: 0.49 Recall: 0.59 F1-score: 0.50 Accuracy: 0.592	Precision: 0.48 Recall: 0.60 F1-score: 0.50 Accuracy: 0.61	Precision: 0.47 Recall: 0.59 F1-score: 0.46 Accuracy: 0.59

Table 4: Data imbalance addressed (ReLU)

	word2vec	GloVe	fastText
<technique-name>			
Under Sampling	0.56	0.57	0.55
Over Sampling	0.572	0.58	0.53

Confusion matrix for the best model



GUI screenshot/working