Alexandria University- Faculty of Engineering

Computer and Systems Engineering Department

Assignment 4 AI

Movies Reviews Classification Using BERT

Name: Marwan Mohamed Saad Abougabal
ID: 18011736
Name: Ahmed Gamal Mahmoud Hefny
ID: 18010083
Name: Mohamed mofreh abdelmonem elgazzar
ID: 18011626
Name: Youssef Hany Fathy Shamsia
ID: 18015025

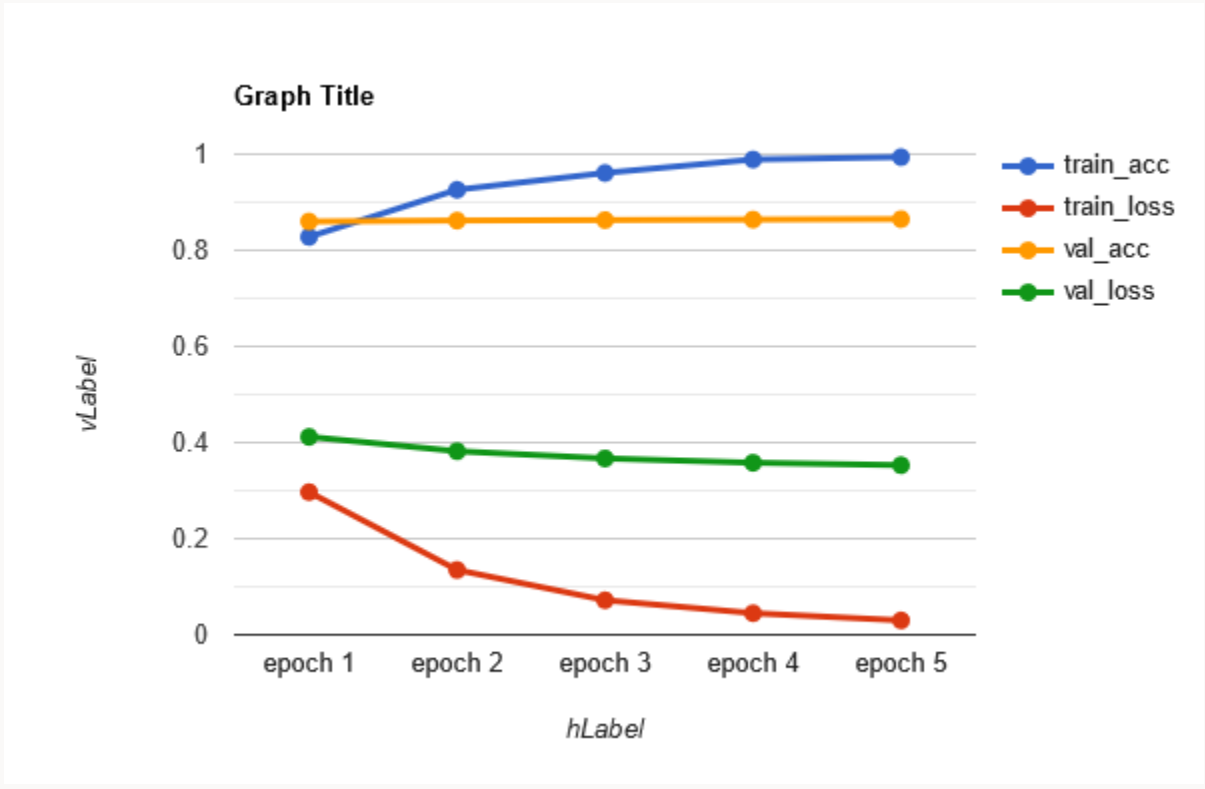
Colab link:

https://colab.research.google.com/drive/1DWvCZG7wZW7UwPpxO9vgtAhaxnRHwW6d?usp=sharing

Change of training and Validation Accuracies

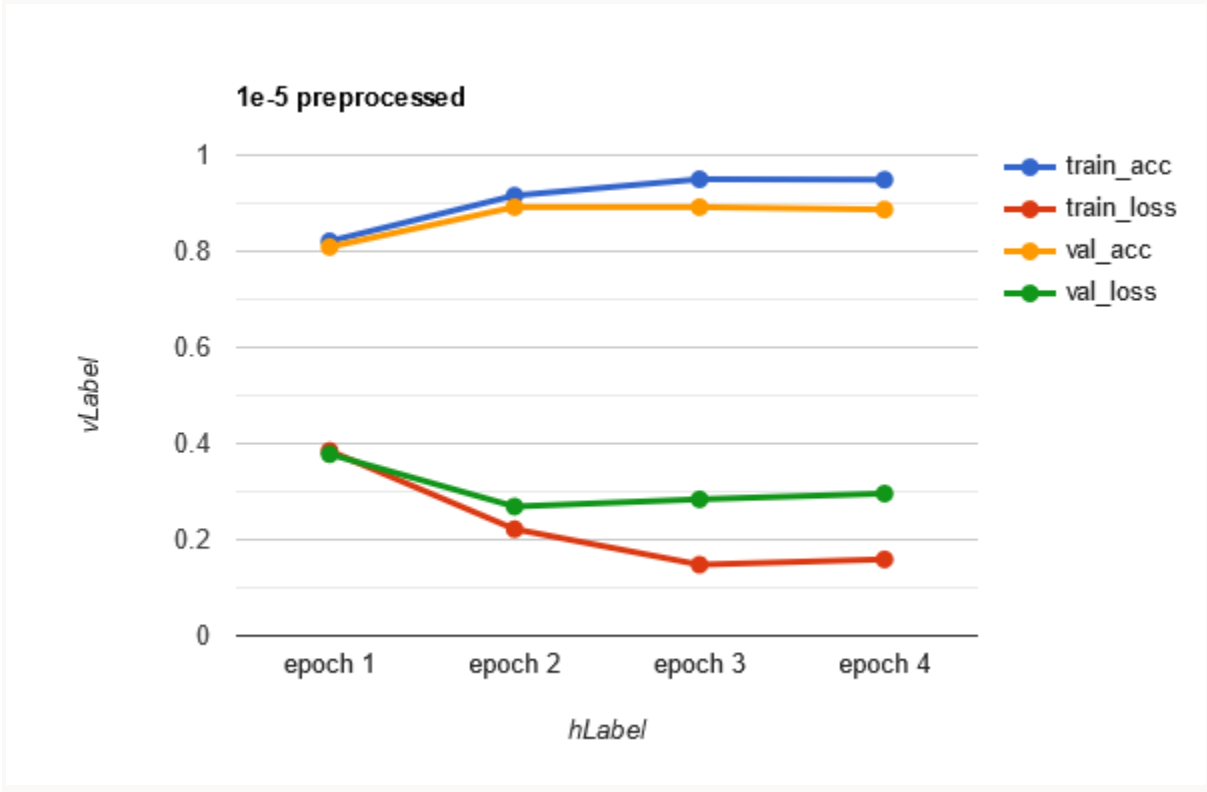
1e-6 preprocessed

```
100%|| 35000/35000 [2:25:18<00:00,  4.01it/s]
Epochs: 1 | Train Loss:  0.045 | Train Accuracy:  0.989 | Val Loss:  0.358 | Val Accuracy:  0.864
1e-06
100%|| 35000/35000 [2:25:35<00:00,  4.01it/s]
Epochs: 2 | Train Loss:  0.030 | Train Accuracy:  0.994 | Val Loss:  0.353 | Val Accuracy:  0.865
1e-06
```



1e-5 preprocessed

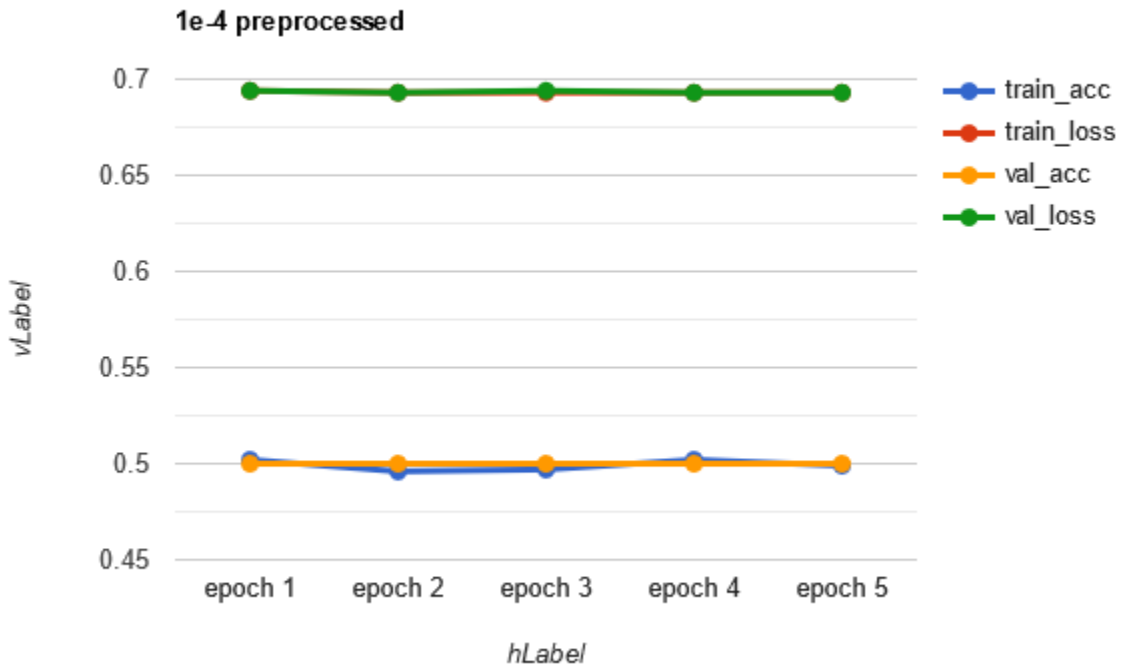
```
Some weights of the model checkpoint at bert-base-cased were not used when initializing BertModel: ['cls.seq_relationship.weight', 'cls.seq_relationship.bias', 'cls.predictions.decoder.weight', 'cls.predictions.transform.LayerNorm.weight', 'cls.predictions.transform.LayerNorm.bias', 'cls.predictions.bias', 'cls.predictions.transform.dense.bias', 'cls.predictions.transform.dense.weight']
- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).
100%|██████████| 35000/35000 [50:39<00:00, 11.52it/s]
Epochs: 1 | Train Loss:  0.385                | Train Accuracy:  0.821                | Val Loss:  0.378                | Val Accuracy:  0.809
1e-05
100%|██████████| 35000/35000 [50:23<00:00, 11.58it/s]
Epochs: 2 | Train Loss:  0.222                | Train Accuracy:  0.916                | Val Loss:  0.269                | Val Accuracy:  0.892
1e-05
100%|██████████| 35000/35000 [50:24<00:00, 11.57it/s]
Epochs: 3 | Train Loss:  0.148                | Train Accuracy:  0.950                | Val Loss:  0.284                | Val Accuracy:  0.892
1e-05
100%|██████████| 35000/35000 [50:32<00:00, 11.54it/s]
Epochs: 4 | Train Loss:  0.159                | Train Accuracy:  0.949                | Val Loss:  0.296                | Val Accuracy:  0.887
1e-05
 55%|█████     | 19187/35000 [27:39<22:36, 11.66it/s]
```



1e-4 preprocessed

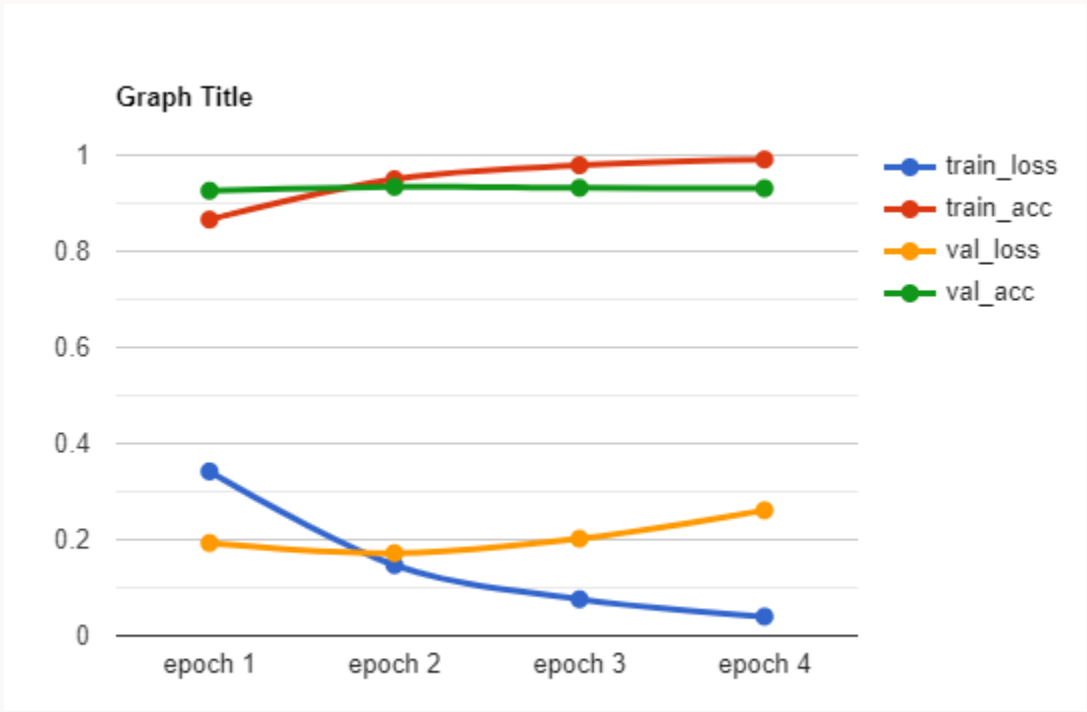
```
Downloading: 100% 416M/416M [00:18<00:00, 24.9MB/s]

Some weights of the model checkpoint at bert-base-cased were not used when initializing BertModel: ['cls.predictions.decoder.weight', 'cls.seq_relationship.bias', 'cls.predictions.transform.LayerNorm.weight', 'cls.predictions.transform.dense.weight', 'cls.predictions.bias', 'cls.predictions.transform.LayerNorm.bias', 'cls.predictions.transform.dense.bias', 'cls.seq_relationship.weight']
- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).
100%|██████████| 35000/35000 [50:09<00:00, 11.63it/s]
Epochs: 1 | Train Loss:  0.694 | Train Accuracy:  0.502 | Val Loss:  0.694 | Val Accuracy:  0.500
0.0001
100%|██████████| 35000/35000 [50:17<00:00, 11.60it/s]
Epochs: 2 | Train Loss:  0.693 | Train Accuracy:  0.496 | Val Loss:  0.693 | Val Accuracy:  0.500
0.0001
100%|██████████| 35000/35000 [50:16<00:00, 11.60it/s]
Epochs: 3 | Train Loss:  0.693 | Train Accuracy:  0.497 | Val Loss:  0.694 | Val Accuracy:  0.500
0.0001
100%|██████████| 35000/35000 [50:15<00:00, 11.61it/s]
Epochs: 4 | Train Loss:  0.693 | Train Accuracy:  0.502 | Val Loss:  0.693 | Val Accuracy:  0.500
0.0001
100%|██████████| 35000/35000 [50:13<00:00, 11.62it/s]
Epochs: 5 | Train Loss:  0.693 | Train Accuracy:  0.499 | Val Loss:  0.693 | Val Accuracy:  0.500
1e-05
```

+ Code + Markdown



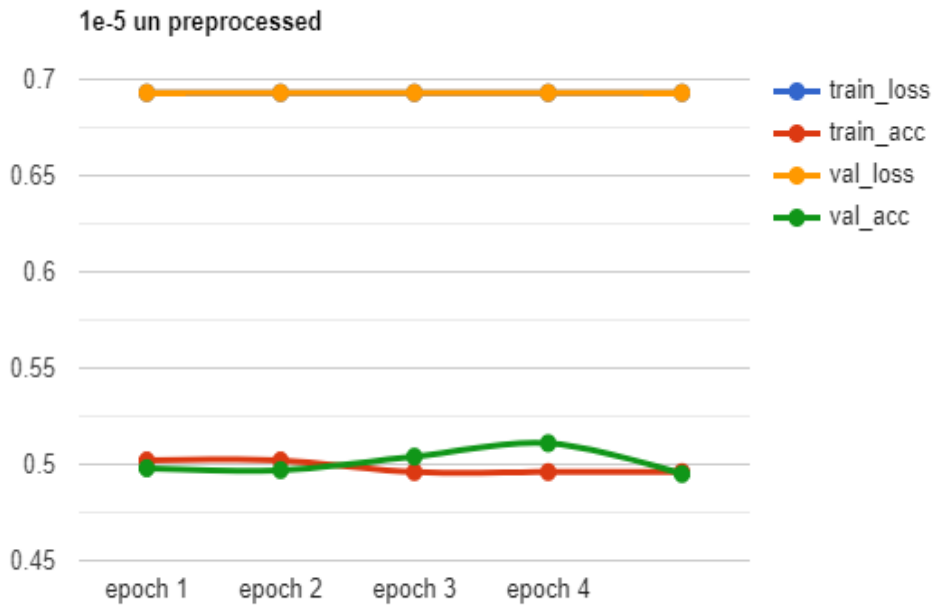
1e-6 unprocessed

```
Some weights of the model checkpoint at bert-base-cased were not used when initializing BertModel: ['cls.predictions.transform.LayerNorm.weight', 'cls.predictions.tran
- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or with another architecture (e.g. initializing a BertForSe
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exactly identical (initializing a BertForSequenceClassifi
100%|██████████| 35000/35000 [1:17:51<00:00,  7.49it/s]
Epochs: 1 | Train Loss:  0.342               | Train Accuracy:  0.866               | Val Loss:  0.193               | Val Accuracy:  0.926
1e-06
100%|██████████| 35000/35000 [1:18:06<00:00,  7.47it/s]
Epochs: 2 | Train Loss:  0.147               | Train Accuracy:  0.950               | Val Loss:  0.172               | Val Accuracy:  0.934
1e-06
100%|██████████| 35000/35000 [1:18:04<00:00,  7.47it/s]
Epochs: 3 | Train Loss:  0.076               | Train Accuracy:  0.979               | Val Loss:  0.202               | Val Accuracy:  0.932
1e-06
100%|██████████| 35000/35000 [1:16:53<00:00,  7.59it/s]
Epochs: 4 | Train Loss:  0.040               | Train Accuracy:  0.991               | Val Loss:  0.261               | Val Accuracy:  0.931
1e-06
 89%|████████▉ | 31313/35000 [1:08:37<08:05,  7.60it/s]
```

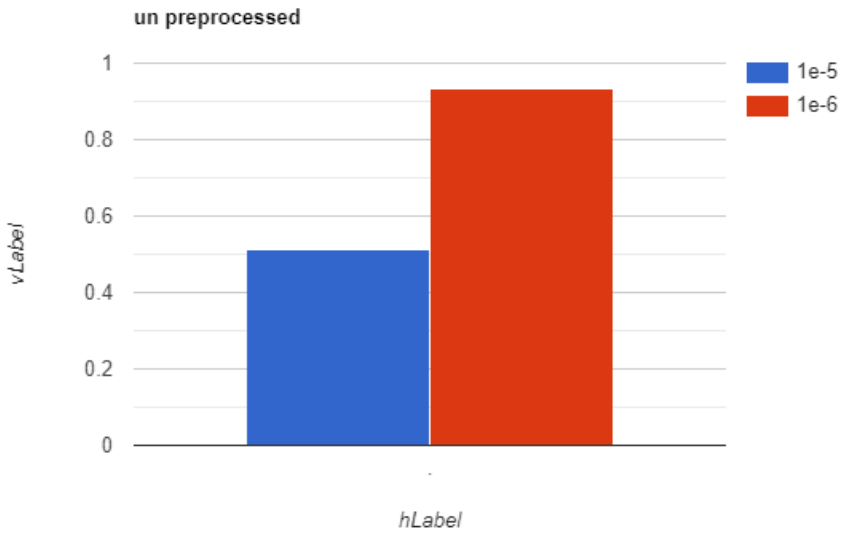


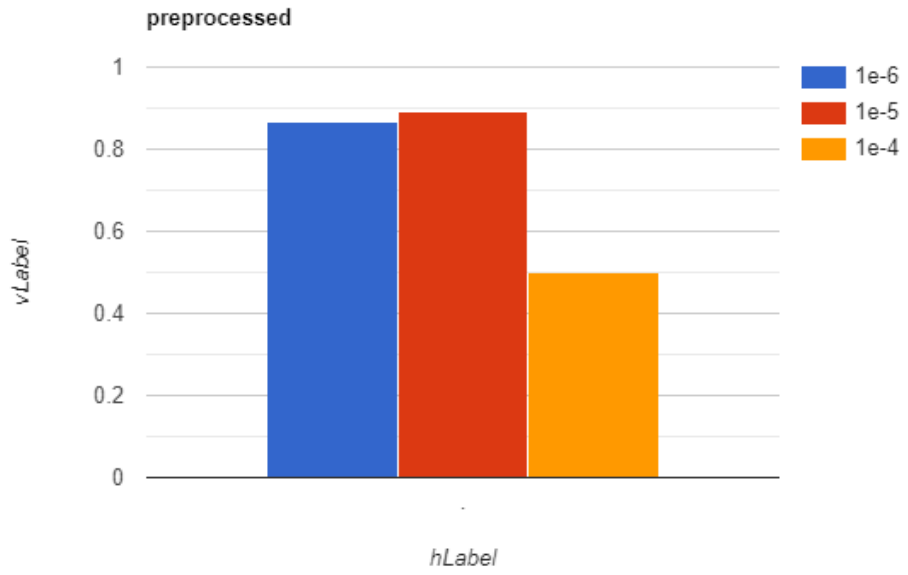
1e-5 unprocessed

```
100%|██████████| 35000/35000 [50:09<00:00, 11.63it/s]
Epochs: 1 | Train Loss:  0.693                 | Train Accuracy:  0.502                 | Val Loss:
0.693                 | Val Accuracy:  0.498
1e-05
100%|██████████| 35000/35000 [50:08<00:00, 11.63it/s]
Epochs: 2 | Train Loss:  0.693                 | Train Accuracy:  0.502                 | Val Loss:
0.693                 | Val Accuracy:  0.497
1e-05
100%|██████████| 35000/35000 [49:56<00:00, 11.68it/s]
Epochs: 3 | Train Loss:  0.693                 | Train Accuracy:  0.496                 | Val Loss:
0.693                 | Val Accuracy:  0.504
1e-05
100%|██████████| 35000/35000 [50:01<00:00, 11.66it/s]
Epochs: 4 | Train Loss:  0.693                 | Train Accuracy:  0.496                 | Val Loss:
0.693                 | Val Accuracy:  0.511
1.0000000000000002e-06

100%|██████████| 35000/35000 [50:33<00:00, 11.54it/s]
Epochs: 1 | Train Loss:  0.693                 | Train Accuracy:  0.496                 | Val Loss:
0.693                 | Val Accuracy:  0.495
1e-06
```



Best Validation Accuracies for Different Learning Rates



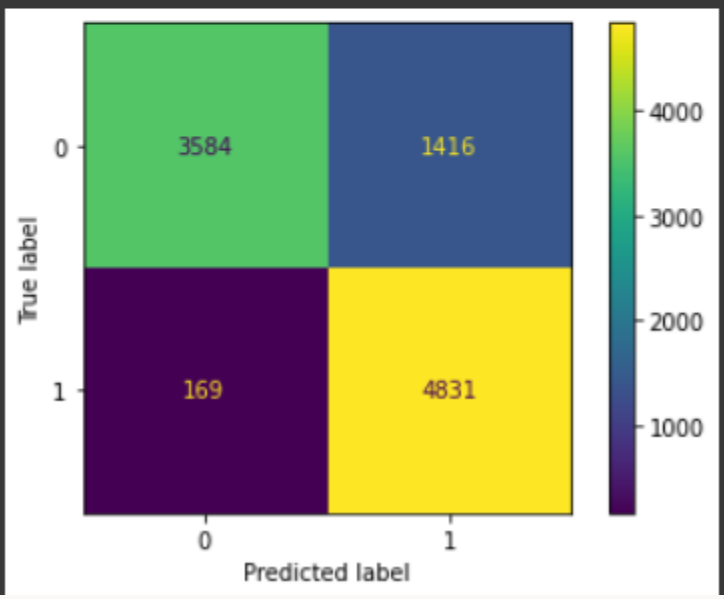


Scoring

==========

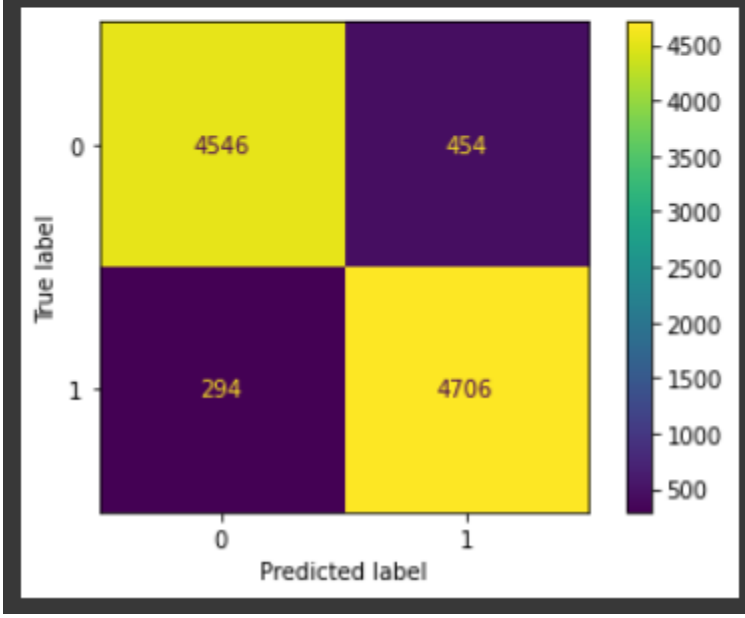
Preprocessed 1e-6

```
Confusion Matrix:
[[3584 1416]
 [ 169 4831]]
Accuracy:
84.15 %
              precision    recall  f1-score   support

           0       0.95      0.72      0.82      5000
           1       0.77      0.97      0.86      5000

    accuracy                           0.84     10000
   macro avg       0.86      0.84      0.84     10000
weighted avg       0.86      0.84      0.84     10000

==============================
```



Un Preprocessed 1e-6

```
Confusion Matrix:
[[4546  454]
 [ 294 4706]]
Accuracy:
92.52 %
              precision    recall  f1-score   support

           0       0.94      0.91      0.92      5000
           1       0.91      0.94      0.93      5000

    accuracy                           0.93     10000
   macro avg       0.93      0.93      0.93     10000
weighted avg       0.93      0.93      0.93     10000

==============================
```



Comments

The previous results show that the BERT models with the un-preprocessed dataset had better accuracy on average when compared to the BERT models with the dataset that was pre-processed using NLTK library. For models with the un-preprocessed dataset, the model with the learning rate of 1e-6 had the best performance. These results seem to indicate that the BERT model performs better given full sentiments in natural language than when the sentiments are pre-processed.