Results Output

Data distribution:

a.FormSpring Data distribution:

```
df_form.groupby(['category', 'Label', 'data_type']).count()
```

category	Label	data_type	Text	hate_words	Text_clean	hate_words_clean
hate_speech	1	test	341	341	341	336
		train	2763	2763	2763	2700
		val	307	307	307	299
non_hate_speech	0	test	1362	1362	1362	1351
		train	11028	11028	11028	10941
		val	1226	1226	1226	1215

a.YouTube Data distribution:

```
df_utube.groupby(['category', 'Label', 'data_type']).count()
```

category	Label	data_type	Text	hate_words	Text_clean	hate_words_clean
hate_speech	1	test	32	32	32	32
		train	256	256	256	256
		val	28	28	28	28
non_hate_speech	0	test	314	314	313	314
		train	2546	2546	2543	2546
		val	284	284	283	284

c.YouTube Data distribution:

```
df_tw.groupby(['category', 'Label', 'data_type']).count()
```

category	Label	data_type	Text	Text_clean
hate_speech	1.0	test	535	535
		train	4331	4329
		val	481	481
non_hate_speech	0.0	test	1148	1126
		train	9295	9158
		val	1034	1021

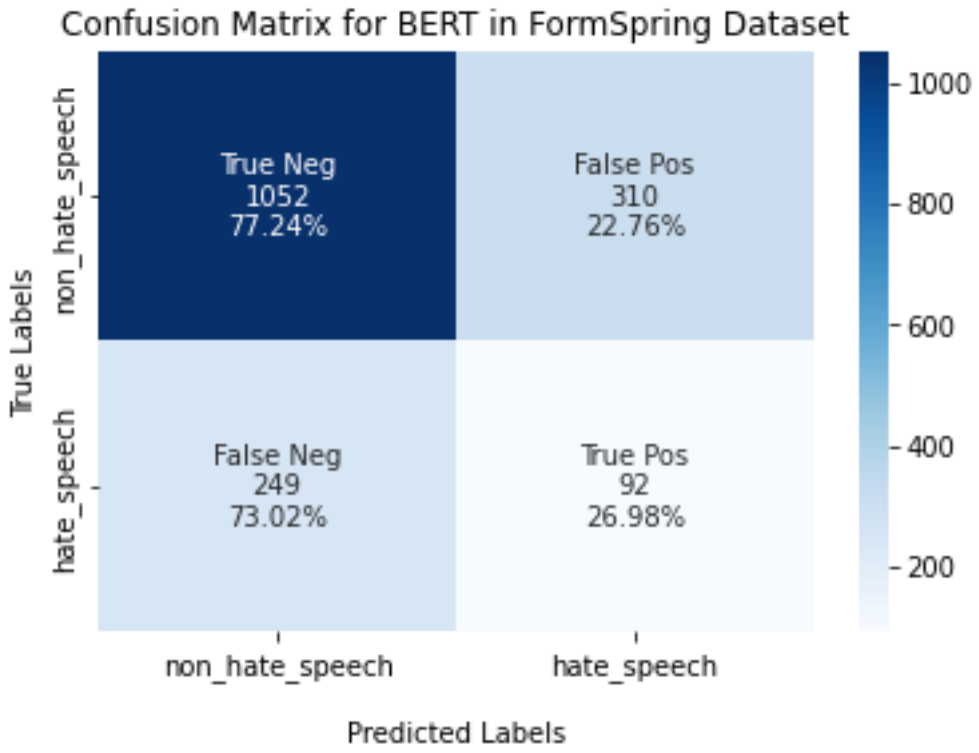
Models:

1. Model 1 (BERT-Baseline):

a.FormSpring Dataset

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, results, average='macro')
print('Precision for BERT in FormSpring Dataset is:',round(result[0],3))
print('Recall for BERT in FormSpring Dataset is:',round(result[1],3))
print('macro F1 score for BERT in FormSpring Dataset is:',round(result[2],3))
```

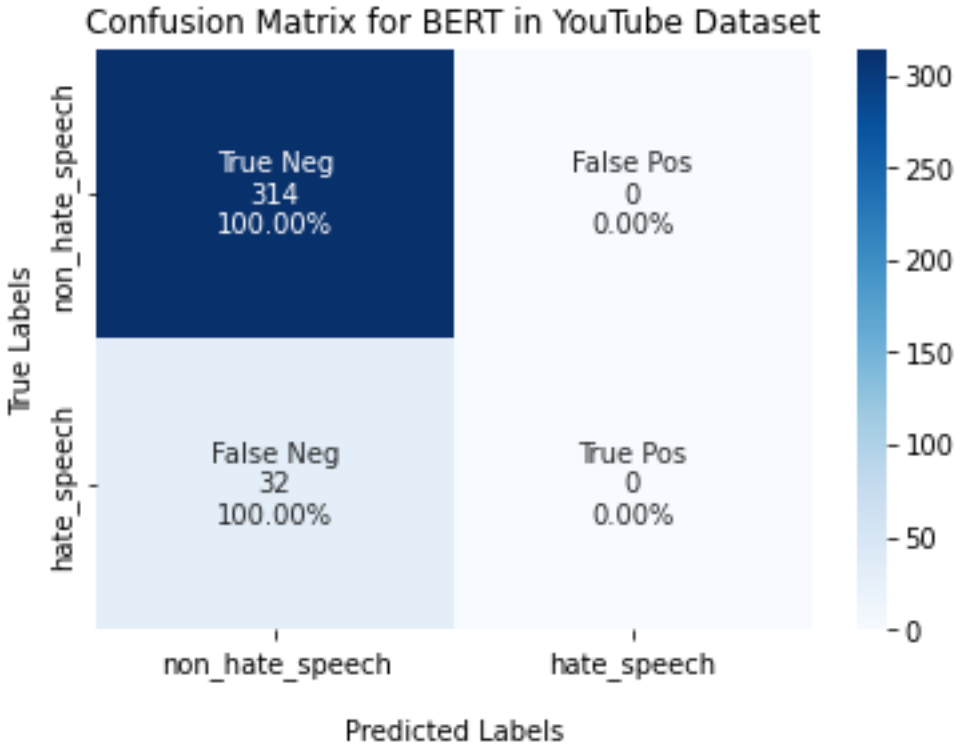
```
Precision for BERT in FormSpring Dataset is: 0.519
Recall for BERT in FormSpring Dataset is: 0.521
macro F1 score for BERT in FormSpring Dataset is: 0.519
```



b.YouTube dataset:

```
result=precision_recall_fscore_support(y_test, results, average='macro')
print('Precision for BERT in YouTube Dataset is:',round(result[0],3))
print('Recall for BERT in YouTube Dataset is:',round(result[1],3))
print('macro F1 score for BERT in YouTube Dataset is:',round(result[2],3))
```

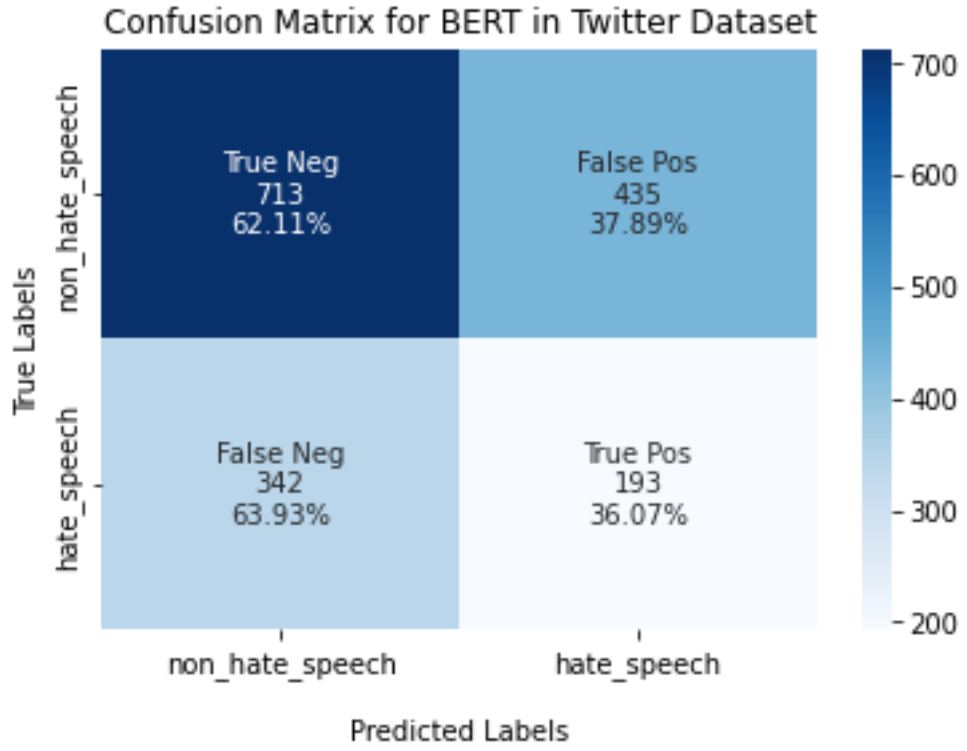
```
Precision for BERT in YouTube Dataset is: 0.454
Recall for BERT in YouTube Dataset is: 0.5
macro F1 score for BERT in YouTube Dataset is: 0.476
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:131
  _warn_prf(average, modifier, msg_start, len(result))
```



c.Twitter Dataset:

```
result=precision_recall_fscore_support(y_test, results, average='macro')
print('Precision for BERT in Twitter Dataset is:',round(result[0],3))
print('Recall for BERT in Twitter Dataset is:',round(result[1],3))
print('macro F1 score for BERT in Twitter Dataset is:',round(result[2],3))
```

```
Precision for BERT in Twitter Dataset is: 0.492
Recall for BERT in Twitter Dataset is: 0.491
macro F1 score for BERT in Twitter Dataset is: 0.49
```

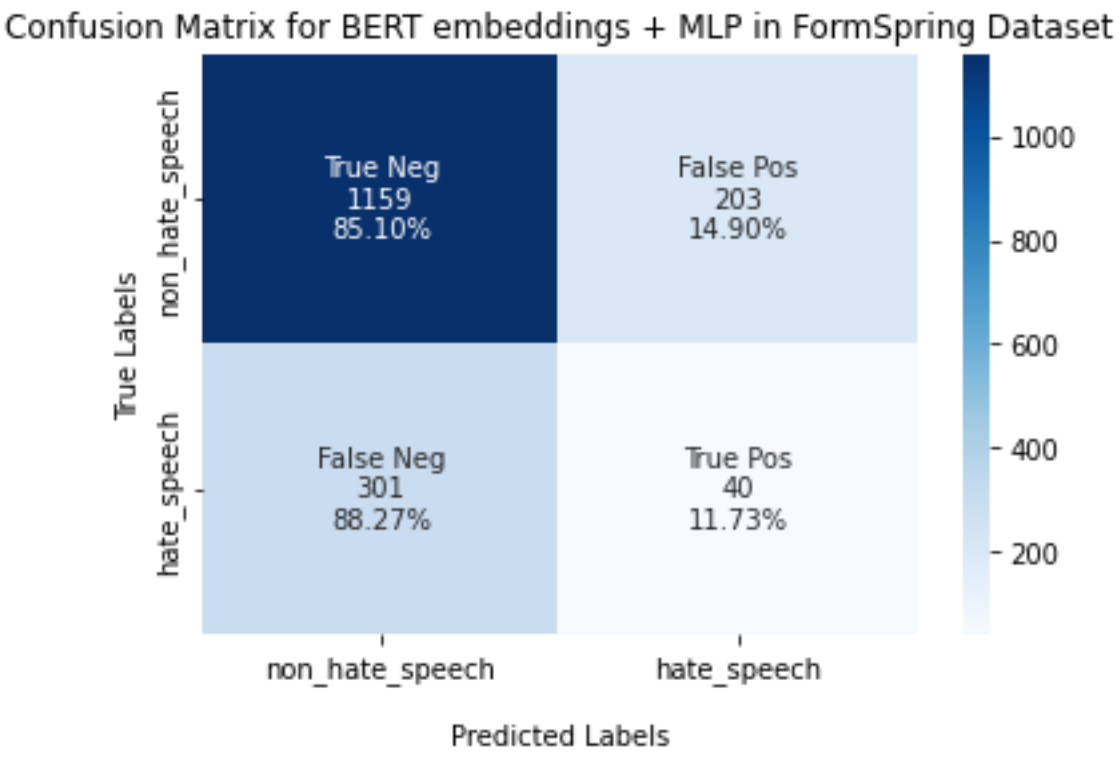


2. Model 2 (BERT Embeddings + MLP):

a.FormSpring Dataset

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, y_pred_mod, average='macro')
print('Precision for BERT embeddings + MLP in FormSpring Dataset is:',round(result[0],3))
print('Recall for BERT embeddings + MLP in FormSpring Dataset is:',round(result[1],3))
print('macro F1 score for BERT embeddings + MLP in FormSpring Dataset is:',round(result[2],3))
```

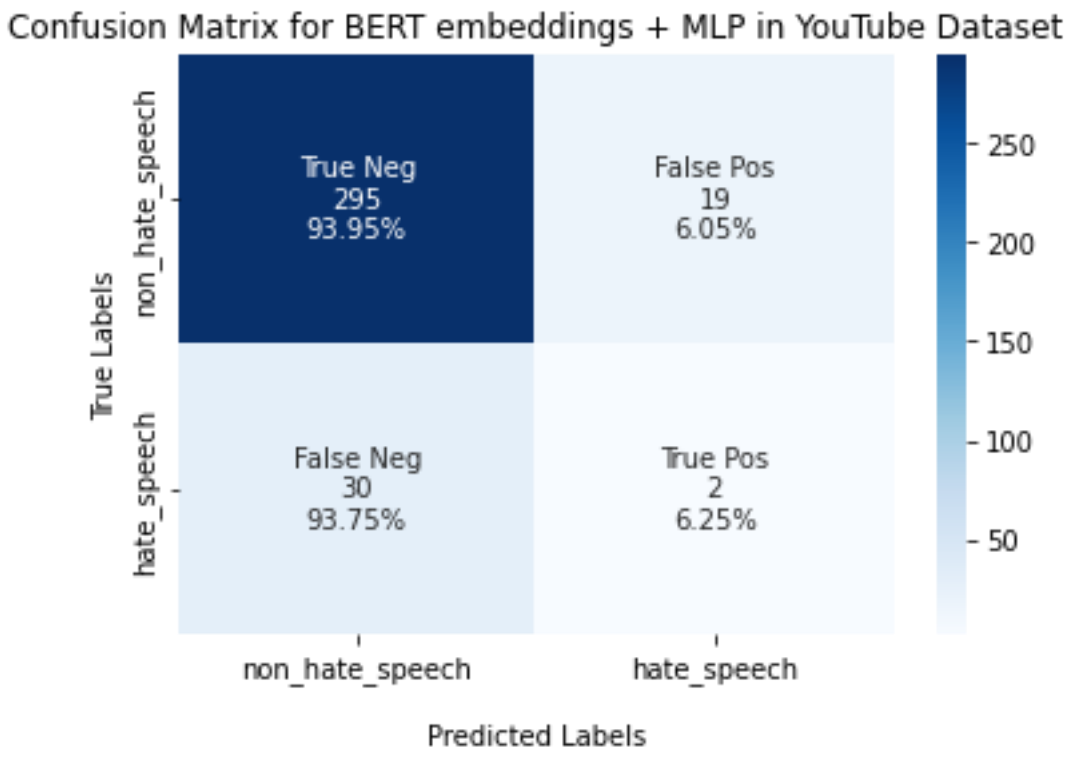
```
Precision for BERT embeddings + MLP in FormSpring Dataset is: 0.479
Recall for BERT embeddings + MLP in FormSpring Dataset is: 0.484
macro F1 score for BERT embeddings + MLP in FormSpring Dataset is: 0.479
```



b.YouTube dataset:

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, y_pred_mod, average='macro')
print('Precision for BERT embeddings + MLP in Youtube Dataset is:',round(result[0],3))
print('Recall for BERT embeddings + MLP in Youtube Dataset is:',round(result[1],3))
print('macro F1 score for BERT embeddings + MLP in Youtube Dataset is:',round(result[2],3))
```

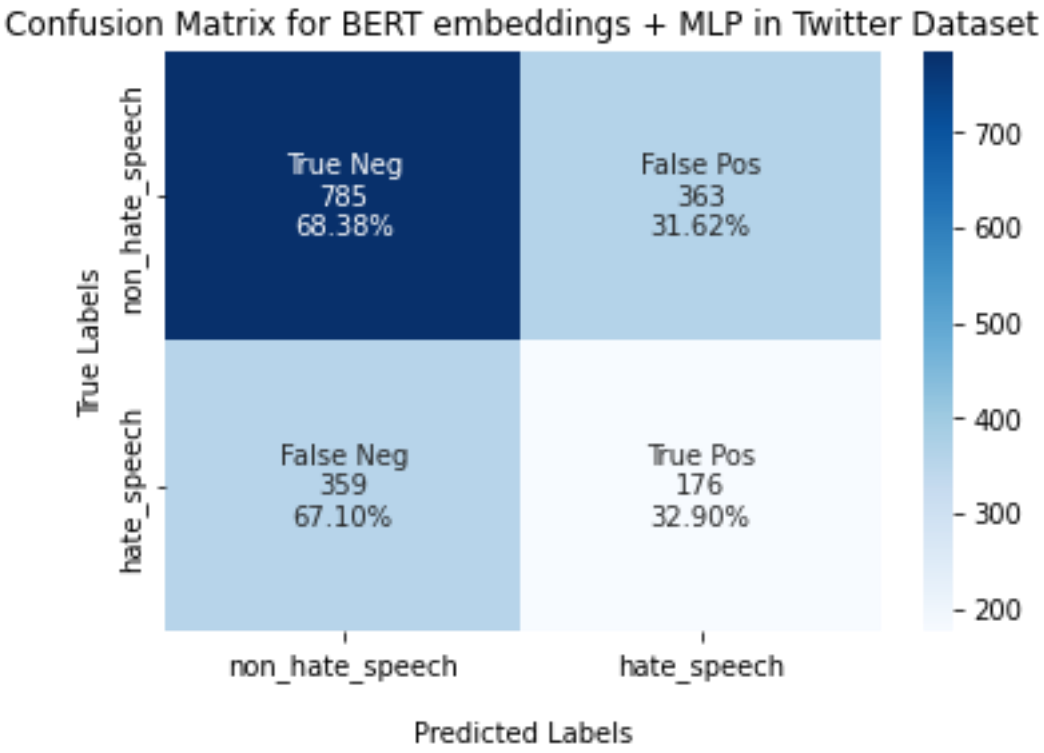
```
Precision for BERT embeddings + MLP in Youtube Dataset is: 0.501
Recall for BERT embeddings + MLP in Youtube Dataset is: 0.501
macro F1 score for BERT embeddings + MLP in Youtube Dataset is: 0.499
```



c. Twitter Dataset:

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, y_pred_mod, average='macro')
print('Precision for BERT embeddings + MLP in Twitter Dataset is:',round(result[0],3))
print('Recall for BERT embeddings + MLP in Twitter Dataset is:',round(result[1],3))
print('macro F1 score for BERT embeddings + MLP in Twitter Dataset is:',round(result[2],3))
```

```
Precision for BERT embeddings + MLP in Twitter Dataset is: 0.506
Recall for BERT embeddings + MLP in Twitter Dataset is: 0.506
macro F1 score for BERT embeddings + MLP in Twitter Dataset is: 0.506
```

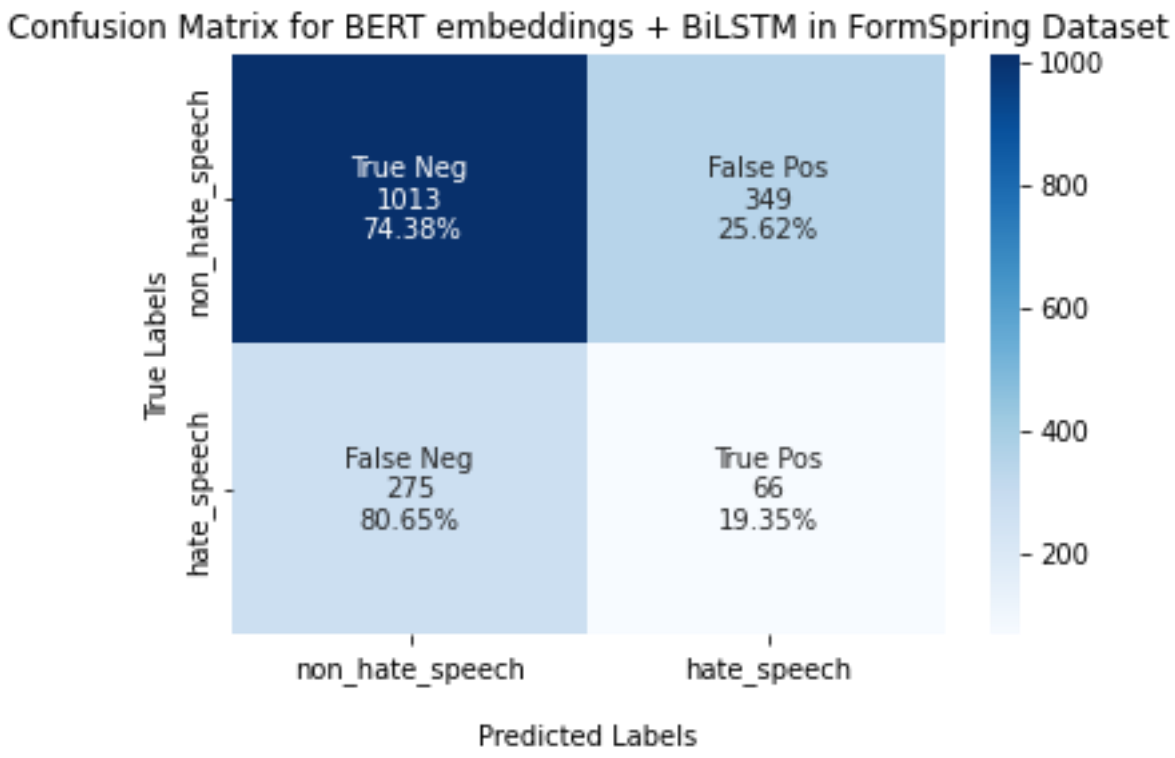


3. Model 3 (BERT Embeddings + BiLSTM):

a.FormSpring dataset:

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, y_pred_mod, average='macro')
print('Precision for BERT embeddings + BiLSTM in FormSpring Dataset is:',round(result[0],3))
print('Recall for BERT embeddings + BiLSTM in FormSpring Dataset is:',round(result[1],3))
print('macro F1 score for BERT embeddings + BiLSTM in FormSpring Dataset is:',round(result[2],3))
```

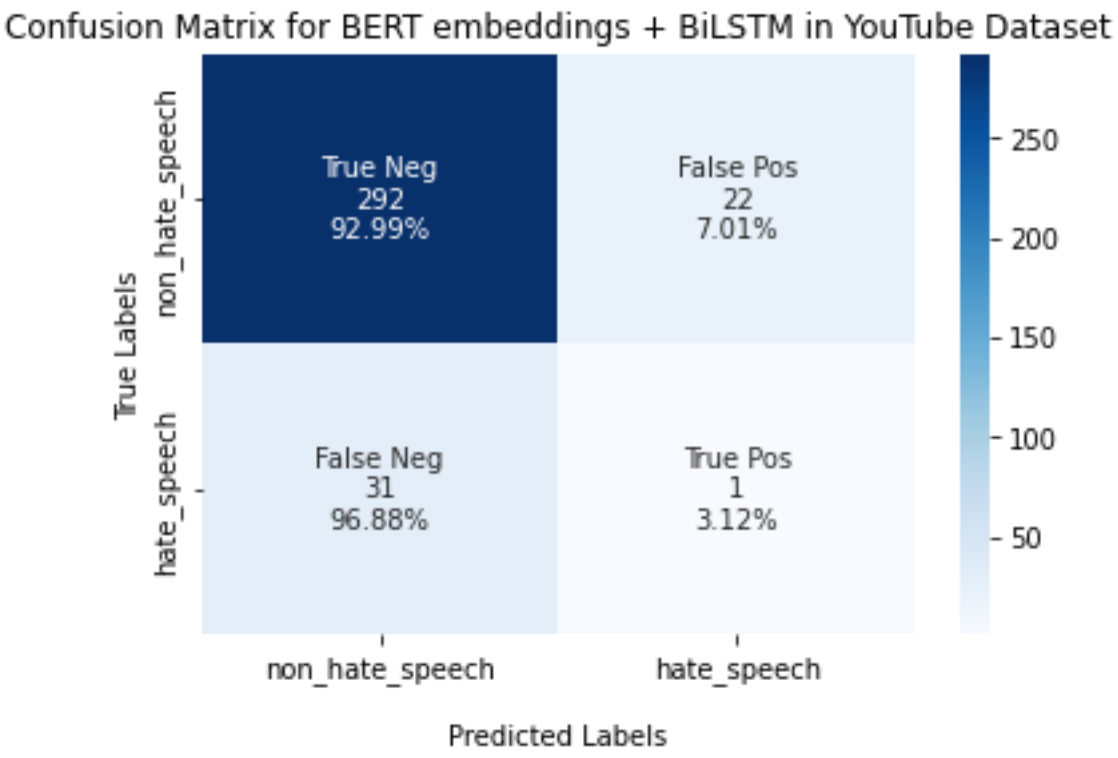
```
Precision for BERT embeddings + BiLSTM in FormSpring Dataset is: 0.473
Recall for BERT embeddings + BiLSTM in FormSpring Dataset is: 0.469
macro F1 score for BERT embeddings + BiLSTM in FormSpring Dataset is: 0.47
```



b. YouTube dataset:

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, y_pred_mod, average='macro')
print('Precision for BERT embeddings + BiLSTM in Youtube Dataset is:',round(result[0],3))
print('Recall for BERT embeddings + BiLSTM in Youtube Dataset is:',round(result[1],3))
print('macro F1 score for BERT embeddings + BiLSTM in Youtube Dataset is:',round(result[2],3))
```

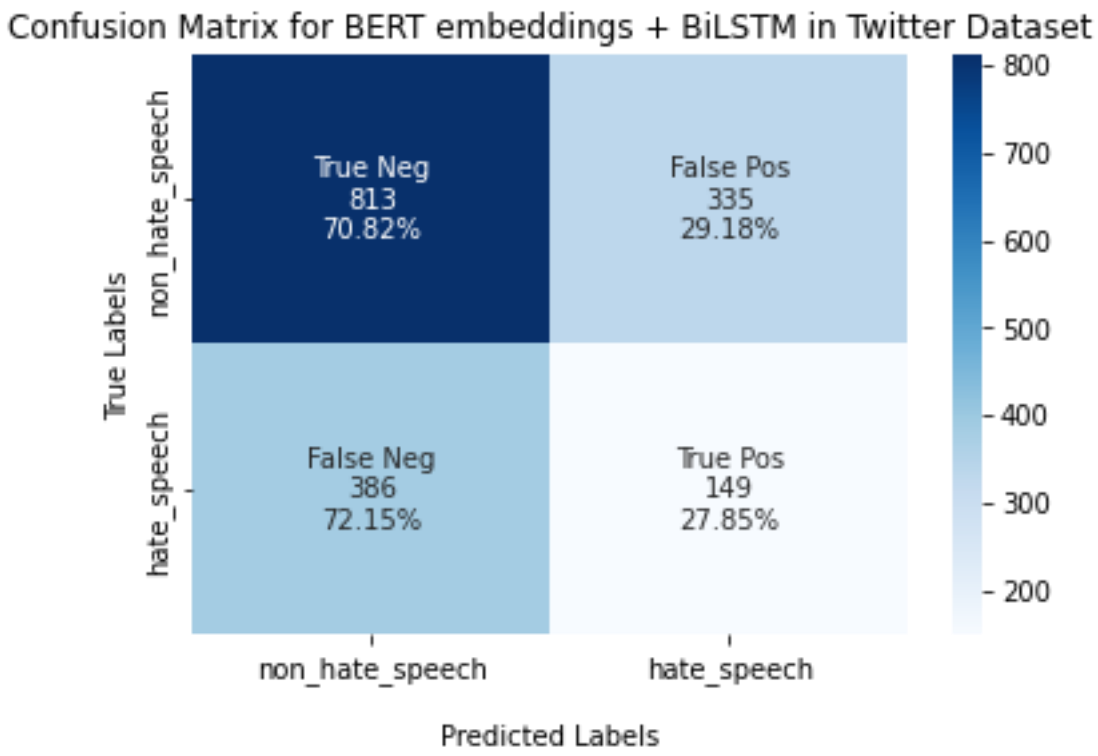
```
Precision for BERT embeddings + BiLSTM in Youtube Dataset is: 0.474
Recall for BERT embeddings + BiLSTM in Youtube Dataset is: 0.481
macro F1 score for BERT embeddings + BiLSTM in Youtube Dataset is: 0.477
```



c. Twitter dataset:

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, y_pred_mod, average='macro')
print('Precision for BERT embeddings + BiLSTM in Twitter Dataset is:',round(result[0],3))
print('Recall for BERT embeddings + BiLSTM in Twitter Dataset is:',round(result[1],3))
print('macro F1 score for BERT embeddings + BiLSTM in Twitter Dataset is:',round(result[2],3))
```

```
Precision for BERT embeddings + BiLSTM in Twitter Dataset is: 0.493
Recall for BERT embeddings + BiLSTM in Twitter Dataset is: 0.493
macro F1 score for BERT embeddings + BiLSTM in Twitter Dataset is: 0.493
```

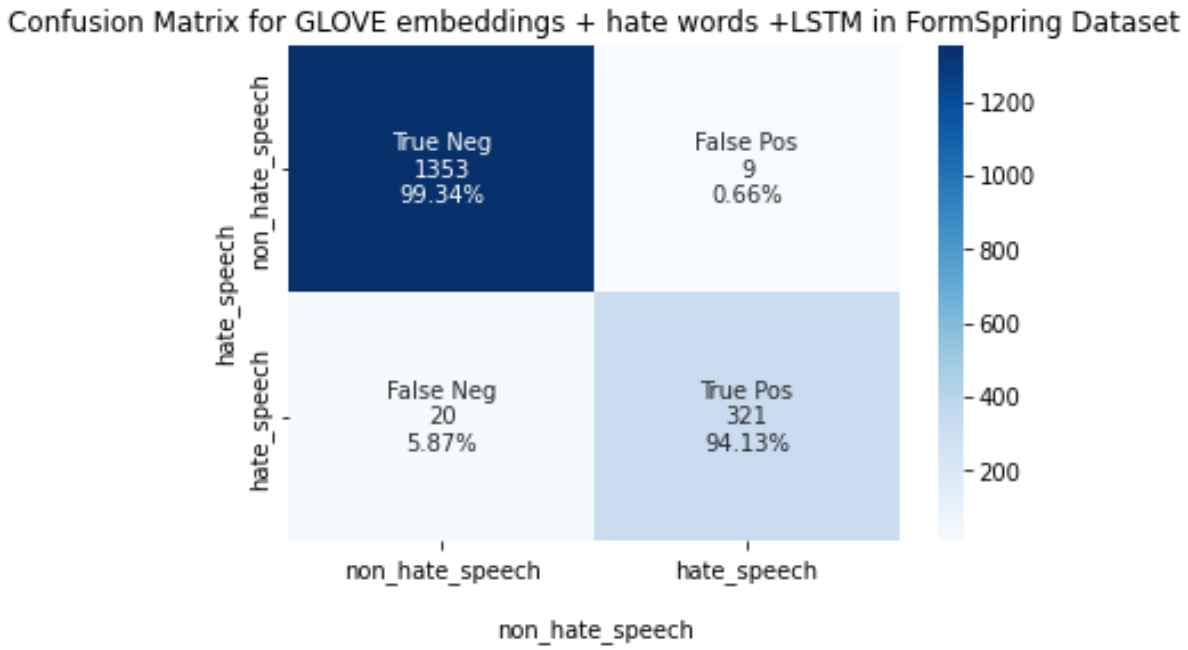


4. Model 4 (GLOVE embeddings + hate words + LSTM):

a. FormSpring dataset:

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, y_pred_mod, average='macro')
print('Precision for GLOVE embeddings + hate words + LSTM in FormSpring Dataset is:',round(result[0],3))
print('Recall for GLOVE embeddings + hate words +LSTM in FormSpring Dataset is:',round(result[1],3))
print('macro F1 score for GLOVE embeddings + hate words +LSTM in FormSpring Dataset is:',round(result[2],3))
```

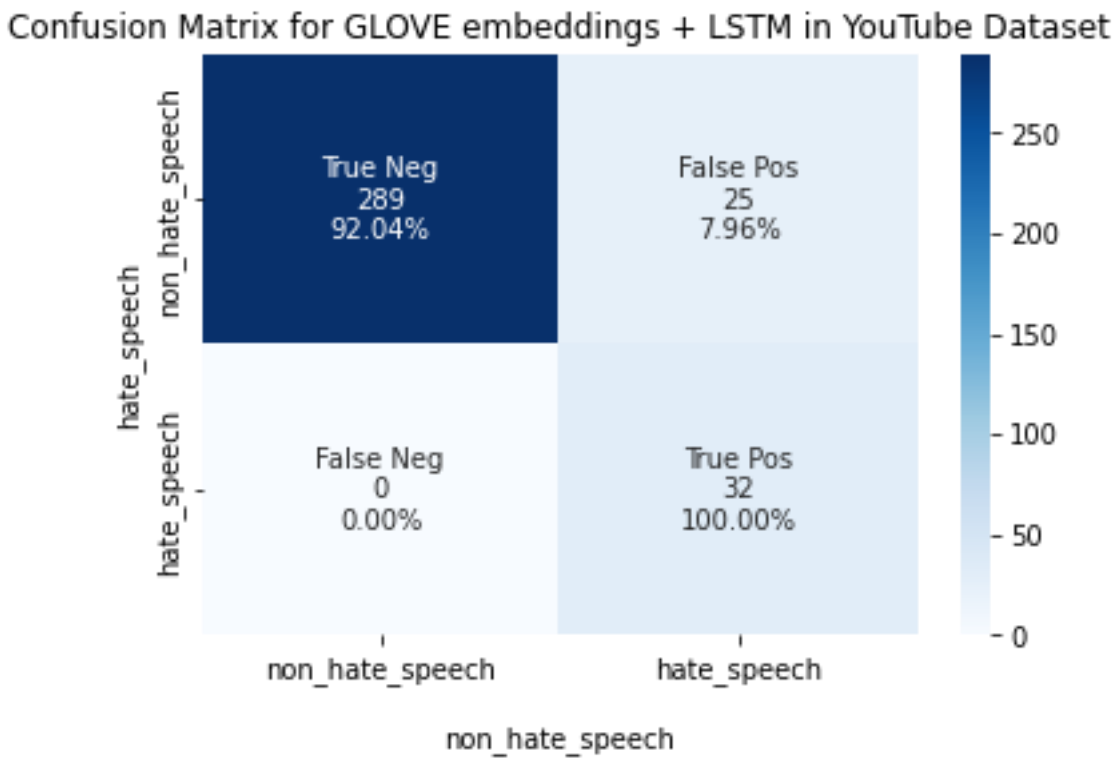
```
Precision for GLOVE embeddings + hate words + LSTM in FormSpring Dataset is: 0.979
Recall for GLOVE embeddings + hate words +LSTM in FormSpring Dataset is: 0.967
macro F1 score for GLOVE embeddings + hate words +LSTM in FormSpring Dataset is: 0.973
```



b. YouTube dataset:

```
from sklearn.metrics import precision_recall_fscore_support
result=precision_recall_fscore_support(y_test, y_pred_mod, average='macro')
print('Precision for GLOVE embeddings + hate words + LSTM in YouTube Dataset is:',round(result[0],3))
print('Recall for GLOVE embeddings + hate words +LSTM in YouTube Dataset is:',round(result[1],3))
print('macro F1 score for GLOVE embeddings + hate words +LSTM in YouTube Dataset is:',round(result[2],3))
```

```
Precision for GLOVE embeddings + hate words + LSTM in YouTube Dataset is: 0.781
Recall for GLOVE embeddings + hate words +LSTM in YouTube Dataset is: 0.96
macro F1 score for GLOVE embeddings + hate words +LSTM in YouTube Dataset is: 0.839
```


Confusion Matrix for GLOVE embeddings + LSTM in YouTube Dataset
True Neg
289
92.04%
False Pos
25
7.96%
False Neg
0
0.00%
True Pos
32
100.00%
hate_speech
non_hate_speech
non_hate_speech
hate_speech
non_hate_speech
250
200
150
100
50
0