

# Model Report for Fine-Tuned BERT for Three-Way Sentiment Analysis of Twitter Data

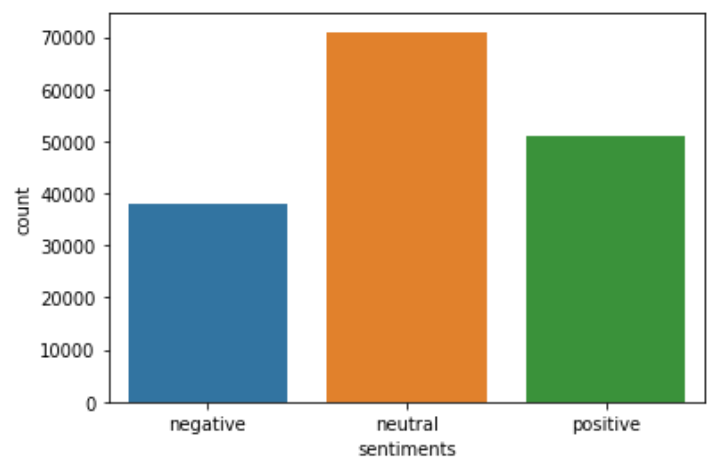
Shyamsunder Nukala

## Report:

The results of the model's metrics show that it is currently making bad predictions for the sentiment of tweets. Based on the information from the classification report we see that the model is always predicting that the tweet is neutral. The reason this model had poor performance was most likely due to the constrained training hyperparameters; number of epochs and batch size, that were changed to account for time and system limitations. The BERT paper suggests a batch size of 8, 16, 32, 64, or 128 but for 4 epochs on fine tuning. A CPU based personal system was originally used but when that proved infeasible; a tesla GPU google colab instance was used to run the code and to finish running in a reasonable time a batch size of 16 at 1 epoch was used. These are not ideal, so the model only seemed to learn that neutral tweets were more likely than other labels as we can see in the confusion matrix and classification report for recall. The first course of action would be to rerun the training this time with the recommended hyperparameters as a high batch size and low epoch number most likely contributed to model inaccuracy. Next adjusting the dataset to be more evenly distributed. As we see in the countplot there is the noticeably higher number of neutral samples in the dataset. We could fix this by either over or under sampling the data to get a more even distribution which would benefit most classification models. To prepare the model for production no changes other than increasing epochs and possibly lowering batch size would be necessary since tweets don't need to be classified in real time quickly and other available instances of BERT for sentiment classification demonstrate that they are able to run quickly with around 90% accuracy. This model would be able to classify tweets in a reasonable time but again the performance would have to be improved.

Performance Metrics and Data Insights

CountPlot:



Training Insights:

Epoch 1/1  
-----

Train loss 1.0724247884816593 accuracy 0.4383680555555556

Val loss 1.0685507994294166 accuracy 0.438625

Test accuracy 0.4408138632640745

	precision	recall	f1-score	support
negative	0.00	0.00	0.00	16571
neutral	0.44	1.00	0.61	30678
positive	0.00	0.00	0.00	22345
accuracy			0.44	69594
macro avg	0.15	0.33	0.20	69594
weighted avg	0.19	0.44	0.27	69594

Confusion Matrix:

