**Observations Criteria:**

Note:

1. For pre-processing of OLID train and test data, a built-in generic tokenizer class from hugging face ` AutoTokenizer` has been used to prepare the inputs for the pre-trained Bert base uncased model
2. Converted target labels from string to integer **(OFF – 0, NOT – 1),** so that they can be in the acceptable format to be able to pass as input to the pre-trained BERT base model
3. To overcome the time consumption constraint, 50% of the OLID training dataset is used, with 100% test dataset for calculating the accuracy

* **BERT base vs MLP vs Logistic Regression vs Naïve Bayes:**

|  |  |
| --- | --- |
| **MODEL** | **Accuracy** |
| BERT base uncased (fine-tuned)  F1 score – **90%** | **95%** |
| MLP | 71.74% |
| Logistic Regression | 72% |
| Naïve Bayes | 90% |

* **SUMMARY OF ANALYSIS OF RESULTS:**

1. Among BERT, MLP, LR, and Naïve Bayes, the fine-tuned version of the BERT base uncased pre-trained model outperformed with 95% of accuracy using only 50% of the OLID dataset which is expected from the model because of its complex learning architecture
2. BERT model with an F1 score of 90% clearly indicates that the prediction of OFF or NOT classification also met the expectation
3. Along with the use of Transformer encoder architecture to process each token of input text in the full context of all tokens before and after and fine-tuned version of the BERT base increased the performance for the text classification task with an accuracy of around
4. Although Naïve Bayes is also in the competition with an accuracy of 90%, because the NB model deals in a better way with probability calculations
5. Other models MLP and LR had the same accuracy with noise in the dataset because of the pre-processing mostly
6. On the other hand, BERT base uncased model pre-processing is handled by built-in hugging face classes AutoTokenizer, which took care of returning the correct tokenizer class instance based on the BERT model for the train and test dataset
7. Only constraint with the BER base model is the time consumption, with a large dataset it was taking more than 11 hours to train the model.
8. Based on model retraining frequency one can choose the model
9. BERT base uncased model experimented with a small amount of dataset of 2k, 1k, and 100 samples only
10. The model performed well over 2k samples of OLID training dataset with an accuracy of 83.5%, 1k samples the accuracy was 75%, and 65% for 100 samples
11. This demonstrated that the BERT base model performed comparably well as compared to MLP and LR models with significantly fewer data
12. Despite being trained on a full dataset, MLP and LR models' accuracy was only about 72%, which was not even close to or more efficient than the BERT base model with less dataset