**Nimit Gupta**

**Sentiment Analysis**

Steps Involved in sentiment analysis:

* Installed all the libraries used for dataset processing and visualizing the data.

### *Loading the data*

* Read the dataset using pandas and run a few commands to check any anomalies in the given dataset.
* After this I moved onto the understanding of the NLP pipeline. I used Natural Language Toolkit (Nltk) which is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language.

## *Preprocessed the data*

* Next, performed steps involved in NLP which goes like this:

1. Tokenization
2. Stopword removal
3. part-of-speech tag
4. Entity extraction

### *Splitting*

* Split the dataset using train\_test\_split in the 20:80 ratio and use the data shuffle function so that order of examples doesn't impact the model's training.

### *Visualizing the data*

* Used seaborn and matplotlib to get the visual of how the data and the sentiments are related to each other.

### *Training the model*

* I used the pre-trained model Bidirectional Representation for Transformers (BERT) for sentiment analysis which is by Huggingface which is a quite efficient model.
* Then we created a model pipeline and trained our data against the model and used various classifiers for better reports and results.

### *Result and analysis*

* In the end we got good and satisfactory results and used various parameters like recall, f1 score and precision to compare the results between different classification models.
* Also we used heat maps, which represent the coefficients, to visualize the strength of correlation among variables.
* And at last, plotted the confusion matrix to represent the prediction summary in matrix form. It shows how many predictions are correct and incorrect per class.

My approach and solution towards this problem is quite easy to understand and implement. Tried not to make the problem more complicated, got the problem analyzed, studied about it, learned new things and lessons and developed some significant knowledge.

If given more time, I was ready to study more deeply about it and used some methods to increase the accuracy of the model like Tuning the model and using classifiers like neural networks and some other things could be done. Also used the concepts of hyperparameters to improve the efficiency of it.