Problem Statement

The data provided consisted of the text column and target column. So, I figured it is a Text Classification problem. The labels were provided and so it is a Supervised Leaning Task. I first looked at the data as a whole and understood its contents. Text Classification problem begins with cleaning the text and converting it into a vector of numbers so that the model can interpret it.

Model Interpretation:

Since the problem at hand was of Classifying the text, I tried a number of Classifier models such as Naïve Bayes, Support Vector Machine and Random Forest Classifier. The models have the following test and training accuracies.

Model	Training Accuracy	Testing Accuracy
Naïve bayes	0.576	0.534
SVM	0.807	0.679
Random Forest Classifier	0.616	0.514
Gradient Boosting	0.935	0.641

Limitations:

Text Classification Models have the following limitations:

- Data sparsity
- Limited Generalizability

Support Vector machine has been considered best for Text Classification. Also while considering Deep Neural Network, Transfer Learning and pretrained models reduce the cost of training a deep neural network and these models have already been trained on a quality aspect. Some of these models are XLNet, ERNIE and Text-to-Text Transfer Transformer (T5)