**Models Operations:**

* Understanding the data and getting an insight into the data to get comfortable with the dataset.
* Converted dataset available in JSON format into standard panda’s data frame.
* Performed basic text analytics.
* Perform *sentence* embeddings that provide vectorized semantic sentence representations. This is done using InferScent.
* Calculate cosine similarity and Euclidean distance with each sentence in the context and predict the output with minimum distance.
* Using these distance vector and cosine similarity, made feature vector.
* Apply Multinomial Logistic regression, Random Forest Classifier, and XGBClassifier on that feature vector.

**Results:**

Accuracy of Prediction Using Euclidean Distance 0.4004383611685065  
Accuracy of Prediction Using Cosine Similarity 0.4905877920980833

Multinomial Logistic regression Train Accuracy: 0.4276022086466165  
Multinomial Logistic regression Test Accuracy: 0.4317434210526316

Random Forest Train Accuracy: 0.7173695958646616  
Random Forest Test Accuracy : 0.5330122180451128  
  
XGBoost Train Accuracy : 0.6573073308270677  
XGBoost Test Accuracy : 0.5615601503759399

**Way Forward**

We will be using BERT (Bidirectional Encoder Representations from Transformers) pre-trained model, fine-tune it as per our data and add a custom model to it to increase its benchmark accuracy on our dataset.

The accuracy we achieved in our model above is low as many advanced techniques are already applied with greater accuracy, but using these models we have learnt and explore our dataset at a lower level which will help in fine-tuning and enhancing accuracy in later steps