

# Report

## 1) **Model Interpretation:**

As this is a classifier-based data so each sentence is very much responsible for classifying it.

Sentence has to be filtered out by using stemming and removing the stop words in order to get the important words which are affecting the target variable.

And based on the data I thought that multinomial naive bayes will be best but it is giving a test accuracy of 63 percent.

## 2) **Train and Test Accuracy:**

After filtering out the stopwords and implementing stemming the model is giving a train accuracy of 0.71 and test Accuracy of 0.63.

## 3) **Limitations of the model:**

- Contextual words and phrases and homonyms.
- Synonyms.
- Irony and sarcasm.
- Ambiguity.
- Errors in text or speech.
- Colloquialisms and slang.
- Domain-specific language.
- Low-resource languages.

## 4) **Adding my own points as well:**

I think that dataset has been build built in a pretty good way because categorizing these sentences is not easier as it looks. As multinomial naïve bayes is best for classification type data so I think it will be good fit for the model and the model will be generalized with low bias and low variance. Label encoded the target categories so that it will be convenient for the model to predict the output. Applying count vectorizer to covert the words into vectors. I have also implemented the multinomial naïve bias with hyperparameter but it increases the test accuracy by one percent which 64.