# **Steps of Project**

## *Pre-processing:*

* Removing punctuations
* Convert to Lower Case
* Tokenization
* Removing Stop Words
* Lemmatization
* Removing empty strings
* Class Imbalance
* ***Bow:***
* N-Grams
* Hashing TF
* TF-IDF Model
* K-Means
* LDA
* **Train-test split(70-30 split)**
* ***Model Building: (with Cross validation 3 folds, Grid Search Hyperparameter optimization)***
* Naïve Bayes
* Logistic Regression
* Random Forest
* GBT
* SVM
* ***Evaluation Metrics***
* Accuracy
* Area under ROC
* Area under Precision-Recall Curve
* Precision
* Recall