**SENTIMENT ANALYSER :**

**Tools and Techniques used :**

Jupyter Notebook,pandas,numpy,nltk,NLP techniques,Pipeline,Different algorithms like Naïve Bayes,Logistic Regression,XG Boost and Random Forest classifier

**STEPS :**

* Importing the libraries
* Importing the training and test datasets
* Merging both the dataset using concat function of pandas and saving as new dataframe “data”
* Renaming the column names as ‘id’,’text’ and ‘emotion’
* Counting the frequencies for each emotion using pd.value\_counts()
* Plotting the emotion output as a bar graph for percentage visualization
* From the bar plot for emotion we are find that ‘Surprise’ has very low count while ‘Joy’ has the most
* Starting with the data preprocessing part
* Checking the missing values/Nan values.There were no missing values
* Now to bring uniformity with the text we make everything lowercase, remove irrelevant and stopwords as well
* Converting to lowercase using str.lower()
* Replacing irrelevant characters other than alpha numeric with space using str.replace()
* Removing stopwords
* Performing lemmatization so as to gain insights as this gets all the words to their root form
* Assigning target variable ‘emotion’
* Label encoding the target variable ‘emotion’
* Splitting the dataset into train and test set in 70:30 ratio
* Extracting count vectorizer parameters which transforms our ‘text’ into an array having the count of appearances of each word alongwith their frequencies .
* Defining a function for printing accuracy
* Using pipeline for models – Naïve Bayes,Logistic Regression,XG Boost and RandomForestClassifier we train and predict using these four different machine learning algorithms
* Accuracy for Naive Bayes 86.70, Logistic Regression 89.27,XG Boost 80.34 and RandomForestClassifier 85.42