**Part B : News Article Classification**

*Explanation video link:- https://drive.google.com/file/d/1XbveLi9dfhJZRntxXn611deVbPbCo9ju/view?usp=sharing*

1. After loading dataset I handle all the missing values by replacing ‘unknown’ in category and empty string in headline and short\_description columns
2. After that I created preprocess\_text() function in which I convert text to lower case, remove non alphabetic character and remove extra white spaces that text contain and then I apply preprocessing by using preprocess\_text method.
3. After that I combine the headline and short description to extract the feature and then I print the no. of rows in each category using value\_counts() function and visualize the distribution of category using countplot.
4. After doing this I find out that that there is equal distribution meant all category has equal count.
5. Then I split data into training and testing phase and after that I extract the feature by using tf-idf
6. After that I created a dictonary models having Logistic Regression, Naive Bayes and SVM and I train each model and print metric like accuracy, F1socre,precision etc using a for loop which will build and train each model which is used in models dictonary.
7. From analyzing each algorithm I found out that Logistic Regression has highest Precision and F1 Score and SVM has highest Recall and Accuracy and this project is of classification so I consider Logistic Regression will be the best for this situation.
8. After that I test logistic regression model for new unseen data and it gives accurate result.

*Explanation video link:- https://drive.google.com/file/d/1XbveLi9dfhJZRntxXn611deVbPbCo9ju/view?usp=sharing*