

Approach Toward Problem Solving

Final\_submission.ipynb is final file

ANIMESH KUMAR SINGH | 04/01/2019

**Approach Towards Solving Problem**

1)Reading the Raw Data.

2)Checking for Any NA value Present in the Data

3)Now Removing Device and Browser Used as It will Not Play any Significant Rule in Prediction, they can come handy in Clustering.

4)Then Using Label Encoder Converting the Response in Numerical Categorial Value as it is much easier to handle.

5)Applying Text Pre-processing Techniques Like Removal of Punctuations, Stopwords and Digits We can also Remove Special Character

6)In EDA, I calculated the Length of each observation Due to Computational Problem Couldn’t able to See the Boxplot but the First Observation was an Outlier

7)In this, I used Stemming instead of Lemmatization because Stemming is much faster than Lemmatization.

8)Now in Feature Engineering, we used both Count and TF-IDF and reduce the features from 960000 to 45000 without compromising the accuracy much.

10)I Got 88% Accuracy with Logistic Regression

11)Few Last Notes, Due to Computer Restriction not able to Perform SVM and XGboost Hopefully they Would’ve Increased the Accuracy of the Model.

12)The Final Result is in Form of Good and Bad.

***Final\_Submission.ipynb & Final\_Submission.csv Is Final File****.*