

Buisness problem Framing:- The use of social media this days enable people to express their views and opinions openly online, thus at the same time this has resulted in conflicts and hate between people and this results the environment unsafe for the users. The researchers have also come to the conclusion that the main reason is hate on such social media platforms and there is no such models for online hate detection.

Conceptual Background of the Domain Problem:- Internet comments are the main reason for violence and hate, Thus machine learning can be used for this purpose. The problem we are here to solve is of tagging the online internet comments that are aggressive against the other users. This as a result insults of the parties like celebrities.

Review of Literature:- There has been a remarkable increase in the cases of cyberbullying and trolls on various social platforms. Many celebrities and the other users are facing problems from people which lead them into depression, they are mentally unstable and also they may lead to attempt suicide.

Motivation for the problem Undertaken:- The comments which are considered as Malignant here it shows the human tendency and behaviour. Here Data Scientist plays a very vital role where he/she filters such wrong words used on the social media platform and not showing such content and blocking such users who do such type of activities on regular bases.

Analytical Problem Facing:-

Here we deal with the main columns which have some importance of the data and the other multiple types of behaviour. I basically tried and focussed on such content which has great value and importance.

Countvector is the NLP term which I am going to apply in columns. This converts the important words.

Data source and Formats:- I have got the data set from my SME which is a csv file. Here I have dropped out the unnecessary data which is not important to me during working on this project.

Data Pre-processing:- Here we have main focus on removal of special characters punctuations from text.

Data Inputs-Logic-output Relationships:- Firstly the data have been converted into neumerical format in the vector form. It feed to the model in the form of series which help to analyze the score through the use of the performamce metrics.

Hardware and Software Requirememts And Tools Used:- Laptop with 8Gbram,jupyterNB,Pandas for Processing.Sciket learn library