Controversy Identification Using Machine Learning

INSTAGRAM , META(FACEBOOK ) AND TWITTER ARE ONE OF THE MAIN OF THE MOST TRENDING AND MOST USED SOCIAL MEDIA PLATFORMS.

Beginning of this topic is very simple and essential as in today world Social Media is not just a part of every person life but also most people of this generation are dependant on the social media for their welfare and it is their main source of income as well So audience interaction is very crucial but some people use this in an offensive way to hurt others and in order to exploit or ruin the person who is posting it as well as it doesn't effect the person who posted that person but also the rest of the people also it influence the other people thinking too. So in order to restrict the users from what they are doing this controversy identification using machine learning is made.

It is good to put your views on something but someone should realise there is selection of words to do the same .

**AIM:**

**To identify the offensive Statements and Users as well by analysing it and predicting sensitivity of the content.**

**OBJECTIVES :**

Through this we are able to modify the sensitivity of the content .

User can be restricted with what they are doing and this is helpful not only on the feed but through this we can analyse the reporting of the users on the basis of their conversation with the other .

Multiple warnings to the same user and ultimately blocking their account can be easily tracked

We can keep track on the activity of the user easily by highlighting it .



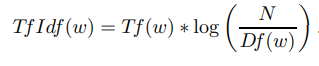
It is very difficult for any social media platform to keep on track of everyone's feed . So this model will help them to do so by this we can clearly analyse their activity as well as set the sensitivity of words by ourselves. Here we have use a Tf-Idf vectorizer or you can say that vector space model for text analysis . A Tf-Idf vectorizer is a part of natural language processing this would help us to know the relation between two statements and how two statements are related to each other.

**TF-ID VECTORIZER works as :**

By measuring the inverse frequency of a word with respect to a statement that is how many times a word is appearing within a text .

This vector space model requires a weighting scheme for assigning weights to terms in a document. There exist many such methods, the most well known of which is the term frequency inverse document frequency (tf-idf) . Let Tf (w) be the term frequency or number of times that word w occurred in the

document, and let Df (w) be the document frequency (number of documents that contain the word). Let N be the number of documents. We define Tf Idf(w) as



The vector space model, when accompanied by some distance metric, allows one to perform many useful tasks. We can use tf-idf and the vector space model to identify documents of particular interest.

As in summary this is helpful in analysing how a word is related to a document by determining the occurrence of a word within the text .



REFERENCES :

Interactive Data Visualization Book by MATTHEW O. WARD