CSE 4022 : NATURAL LANGUAGE PROCESSSING

Team :
PRANJAL SINGH (19BCE1207)
ANISH SARKAR (19BCE1278)

Title: Social Media sentiment analysis

Abstract

NLP or natural language processing is a powerful and effective way for the interaction between computers / machines and human understandable language via interpretations based on multiple training sets , rules and grammar in linguistics. If we master the art of making the machine able to understand the language we do multiple applications can be used. We aim to exploit one such field of nlp in our project i.e. sentiment analysis via text on social media to detect the psyche of the user and getting the insights. Training of the model on a vast dataset exposes it to multiple texts which make it easier for accurate analysis.

Sentiment analysis is basically an idea to determine if the given data piece is positive, negative or neutral. Here the criteria - positive, negative or neutral depends on what the source and area of data set is. Example sentiment analysis for student can give whether his thoughts are suicidal or not, for an employee it can be whether is he satisfied with his job or not, for criminally active space it can give an insight if the person has criminal mindset which can alert the agencies .Trending topics on social sites like twitter can give the idea if general public is in favour of law or not. In this we will predict sentiments of public via their online presence.

In our project we aim to classify negative comments and flag it if it's too dangerous. Understanding Problem case and Importing libraries and dataset to Perform exploratory data analysis by plotting the word cloud and Perform data cleaning – removing punctuations followed by Perform data cleaning – removing stop words. In order to achieve accuracy we will perform tokenization (count vectorization) by creating pipeline and applying naive bayes classifiers after training of naive bayes classifiers. Final step is to access trained model performance.