

DA Assignment - 4

Title :- Twitter Data Analysis

Problem Statement :-

Use twitter data for sentiment analysis. The dataset is 3MB in size and has 31962 tweets. Identify the tweets which are hate tweets and which are not.

Objective :- To analyze twitter data using any classifier.

Outcomes :- To do sentiment analysis on twitter data using classification algorithm.

Theory :-

- sentiment analysis is the process of determining whether a piece of writing is positive, negative or neutral.
- It can be used to identify the customer's attitude towards a brand through the use of variables such as context, emotion.

Steps to perform sentiment analysis

- Gather relevant tweets
- Preprocessing
- Feature extraction
- Feature selection

1) Gather relevant tweets
sample test data made available by twitter.

2) Preprocessing
Removing twitter handles
Removing punctuations, numbers, special chars
Removing short words
Tokenization
Stemming

3) Feature extraction:-
selection of useful words from the tweet

4) Feature selection:-
correct feature selection techniques are used in sentiment analysis that has got a significant role for identifying relevant attributes and increasing accuracy

Natural language Processing
statistical
Clustering based
Hybrid

3] Classification:-

classification algorithm can be used to classify tweets.

a) Naive Bayes classifier

b) SVM.

Conclusion:-

Thus sentiment analysis on twitter data was successfully done using SVM, After performing sentiment analysis tweet was classified in to hate/not hate.