

LP1 Assignment 14

Title:- Sentiment Analysis

Date of Completion:- 27/11/20

Problem Statement:- Twitter Data Analysis: Use Twitter data for sentiment analysis. The dataset is 3mb in size & has 31962 tweets. Identify the tweets which are hate tweets and which are not. ~~So~~

Learning Objectives:-

- 1) Understand text processing
- 2) Understand sentiment analysis

Learning Outcomes:- Students will be able to

- 1) To do sentiment analysis on twitter dataset using classification algorithm

Software/Hardware Requirements:- Python Libraries, OS (Linux)

Theory:-

1) Sentiment Analysis:-

It is the process of determining whether a piece of writing is positive, negative or neutral.

2) Applications:-

- a) Understand customer's attitude towards products & services.
- b) Marketers can get public opinion of their company.

- and product
- c) Gather critical feedback about problems in newly released products.

3) Steps

- Gather relevant tweets from twitter
- Preprocessing (stopword removal)
- Feature Extraction
- Feature Selection

a) Gather relevant tweets

Gather sample data from relevant sources & datasets to prepare training and testing set.

b) Preprocessing:-

- Without preprocessing there is high chance of text having noisy or inconsistent data.
- Punctuations, special characters, numbers are not required to do sentiment analysis.
- Removing punctuations, special characters, numbers
- Remove short words
- Tokenize data :- split strings into tokens (smaller words)
- Stemming :- getting the root word. (removing suffixes).

c) Feature extraction:-

- Selection of useful words is called feature extraction.
- Three different features namely unigram, bigram, n-gram
- Part of speech tags.
- Negation :- It may change the prior polarity of sentiment

d) Feature Selection:-

4 types

- i) Natural language processing
- ii) Statistical
- iii) Clustering based
- iv) Hybrid
- v) Classification

Dataset :- Twitter datasetResult:-Test case

Confusion matrix $\begin{bmatrix} 7373 & 57 \\ 230 & 331 \end{bmatrix}$

accuracy score :- 0.96

recall score :- 0.59

Precision score :- 0.85

F1 score :- 0.69

Conclusion:- Thus I have understood sentiment analysis and how to perform it on String dataset. I have also completed the assignment successfully.