Page :
Date : / /

Assignment : C4

- TITLE : Twitter Data Analysis.
- Problem Statement : Twitter Data Analysis : use Twitter Data for sentiment analysis. The dataset is 3 MB in the size and has 8523 ~~rows~~ & 12 variables. ~~Predict the Sales of a Store~~. 31,962 Tweets. Identify the hate tweets & which are not.
- Objective : i) To do analysis of twitter data.
  ii) Sentiment analysis.
- Outcome : i) To learn the data analytics concepts.
  ii) To learn the sentiment analysis.
- Software and Hardware Requirements : OS: windows / ubuntu distribution 4GB RAM. 500 GB HDD, Python frameworks, Python libraries, Twitter Dataset.
- Theory :

Sentiment analysis is the automated process of analyzing text data & sorting it into sentiments positive, negative or neutral. Using sentiment analysis tools to analyze opinions in Twitter data can help companies understand how people are talking about their brand. Twitter boasts 330 Million monthly active users, which allows businesses to reach a broad audiences and connect with customers without intermediaries. On

Page:
Date: / /

the downside. there is so much information that it's hard for brands to quickly detect negative social mentions that could harm their business. This is why social listening, which involves monitoring conversations on social media platforms has become a key strategy in social media marketing.

The most common type of sentiment analysis is called polarity detection and involves classifying a statement as positive, negative or neutral. A sentiment analysis model would automatically tag a tweet like "love the new security feature" as positive. Try this phrase out for yourself in the sentiment analyzer below. Sentiment analysis uses Natural Language Processing to make sense of human language and machine learning to automatically deliver accurate results.

How do Do twitter Sentiment Analysis?
Performing sentiment analysis on twitter data involves several steps:

1) Get twitter Data: It's important that your data is representative of what you're trying to find out because you'll use it to:
   - Train your sentiment analysis model.
   - Test how your model performs on Twitter Data.

You should consider the types of tweets you want to analyse:

- Current tweets : useful to track keywords or hastags in real-time.
- Historical tweets : useful to compare sentiments over different periods.

2) Prepare your Data : Once you've captured the tweets you need for your sentiment analysis, you'll need to prepare your data. As we mentioned earlier social media data is unstructured. That means it's raw noisy and needs to be cleared before we can start working on our sentiments analysis model. This is important step because good quality data will lead to more accurate results. Preprocessing a Twitter datasets involves a series of tasks like removing all types of irrelevant information like emojis, special characters and extra blank space.

3) Create a Twitter Sentiment Analysis Model : If you want to get predictions with a higher level of accuracy, tailored to your criteria and domain then the best way is to create your own customized sentiments Analysis model by training it with your own twitter data.

4) Visualize your Data :
Data visualization tool help explain sentiment Analysis results in a simple and effective way.

Page:
Date: / /

- Test Cases:

	Description	Output	Result
i]	Identify hate tweets count	~~29720~~ 2242	Pass
ii]	Identify non hate tweets count	29720	Pass
iii]	Tokenized tweets	[when, father, and, ... [thanks, for, ...] , etc	Pass
iv]	Dataset Shape	(31962, 3)	Pass

- Conclusion: Thus we successfully done twitter sentiment Analysis and identified hate tweets and non hate tweets.