Sentiment Analysis of Tweets Based on the Polarity of their Bag of Words Representation

By Melanie Qu
Summer of 2020

In a nutshell…

1. Web Scrape tweets data from the website using API.
2. Tokenize, filter by length, lowercase the words, and extract the major features from 20 tweets based on frequency distribution.
3. Rank each tweet by the number of positive/negative terms matched with a dictionary. This will serve as training set.
4. Use machine learning algorithms, such as Naive Bayes Classifier, Support Vector Classifier, and Decision Tree Classifier to see which is most accurate in predicting polarity.

End users

- Market Research
- Stock Returns for Investors and Hedge Fund Managers
- Political Campaigns

Minimum Requirements for 8 Weeks

Basic Functionalities:

- ability to assign ratings to tweets
- extract relevant key features for evaluation
- compare the test results with the training results to ensure accuracy and precision.

Ideal Data

- Loughran McDonald Master Dictionary of business
- NYtimes Tweets

Existing Solution & More...

- Inspired by RavenPack
- My solution is more applicable to general news data and is based on statistical analysis.

Validation of Success of the Project

- Naive Bayes Classifier - 3rd
- **Vector Support Machine Classifier - 1st**
- Decision Tree Classifier - 2nd

Bibliography

Github Repository:
https://github.com/marlesque/Sentiment_Analyzer/tree/master/LICENSE.md

Master Dictionary:

https://sraf.nd.edu/textual-analysis/resources/

Thank you!