

Sentiment Analysis of Twitter Data using Python

Hetu Bhavsar¹, Richa Manglani²

^{1,2}GyanManjari Institute of Technology, GMIT, Bhavnagar-364001, India

Abstract:- Twitter is a popular social networking website where users posts and interact with messages known as “tweets”. This serves as a mean for individuals to express their thoughts or feelings about different subjects. Various different parties such as consumers and marketers have done sentiment analysis on such tweets to gather insights into products or to conduct market analysis. Furthermore, with the recent advancements in machine learning algorithms, the accuracy of our sentiment analysis predictions is able to improve. We will attempt to conduct sentiment analysis on “tweets” using various different machine learning algorithms.

Key Words: Sentiment analysis, Twitter data, Anaconda, python, positive, negative

1. INTRODUCTION

With the huge amount of increase in the web technologies, the no of people expressing their views and the opinion via web are increasing. This information is useful for everyone like businesses, governments and individuals with 500+ million tweets per day, twitter is becoming a major source of information. Input to our model is the raw data extracted from tweets. For the same, we automate the process of tweet extraction and categorizing it into two categories i.e. positive or negative. The content in twitter generated by the user is about different kinds of products, event, people and political affairs [7].

Performing sentiment analysis on tweets is considered best due to the following reasons:

1. Tweets are abstract in nature.
2. Analysis in real time can be done.
3. A vast variety of tweets for performing the analysis. [2]

2. LITERATURE SURVEY

1. Opinion Mining on Social Media Data [8]:

Po-Wei Liang et.al used Twitter API to collect data from twitter. Tweets which contain opinions were filtered out. Unigram Naive Bayes model was developed for polarity identification. They also worked for elimination of unwanted features by using the Mutual Information and Chi square feature extraction method. Finally, the approach for predicting the tweets as positive or negative did not give better accuracy by this method.

2. A Sentiment Analysis Prototype System for Social Network Data [3]:

This paper discusses a prototype system to enable automated opinion based analysis of social network data for evaluation of the public social conscience of user provided topics and events.

3. Sentiment Analysis for Social Media [1]:

Sentiment Analysis is a problem of text based analysis, but there are some challenges that make it difficult as compared to traditional text based analysis This clearly states that there is need of an attempt to work towards these problems and it has opened up several opportunities for future research for handling negations, hidden sentiments identification, slangs , polysemy. However, the growing scale of data demands automatic data analysis techniques. In this paper, a detailed survey on different techniques used in Sentiment Analysis is carried out to understand the level of work.

4. Sentiment Analysis for Using Twitter DataSet [6]:

The research of sentiment analysis of Twitter data can be performed in different aspects. This paper shows sentiment analysis types and techniques used to perform extraction of sentiment from tweets. In this survey paper, we have taken comparative study of different techniques and approaches of sentiment analysis having twitter as a data.

5. Sentiment Analysis of Tweets Using Machine Learning Approach [2]:

In this paper, SVM, ADABOOSTED DECISION TREE and DECISION TREE based hybrid sentiment classification model are presented for improving the overall accuracy of the classifier in the classification of tweets.

3. OUR PROPOSAL

The proposed model of twitter data analysis will be implemented using Anaconda python. **Anaconda** is a free and open source distribution of the **Python** and R programming languages for data science and machine learning related applications. It can be installed on Windows, Linux, and MacOS. Conda is an open source, cross-platform, package management system.

The tweets can be analysed and characterized based on the emotions used by the social users. We attempt to classify the polarity of the tweet where it is either positive or negative. If the tweet has both positive and negative elements, the more dominant sentiment should be picked as the final label. We use the dataset from Kaggle which was crawled and labelled positive/negative. The data provided comes with emoticons, usernames and hashtags which are required to be processed and converted into a standard form. It also needs to extract useful features from the text such unigrams and bigrams which is a form of representation of the "tweet".

4. CONCLUSION

The analysis of Twitter data is being done in different points of view to mine the opinion or sentiment. Our proposed approach classify the tweets as Positive and Negative tweets which further helps in sentiment analysis and uses that sentiment analysis for further decision making. For our prototype, Twitter API is utilized to gather data in real-time. The prototype back-end tests on retrieving and processing the API data indicate that it is successful in gathering large amounts of data from popular search terms in real-time. We will use various machine learning algorithms to conduct sentiment analysis using the extracted features. However, just relying on individual models did not give a high accuracy so we pick the top few models to generate a model.

REFERENCE

- [1] Harshali P. Patil, Dr. Mohammad Atique, "Sentiment Analysis for Social Media: A Survey", Department of Computer Engineering Thakur College of Engineering & Technology Mumbai, India.
- [2] Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, Sarthak Mendiratta, "Sentiment Analysis of Tweets using Machine Learning Approach", Jaypee Institute of Information Technology IIIT Sec-62 Noida, India.
- [3] Pitiphat Santidhanyaroj, Talha Ahmad Khan, "A SENTIMENT ANALYSIS PROTOTYPE SYSTEM FOR SOCIAL NETWORK DATA," Faculty of Engineering and Applied Science University of Regina Regina, Saskatchewan, Canada.
- [4] Subramaniam.G, Ranjitha.M," User Emotion Analysis using Twitter Data", IFET COLLEGE OF ENGINEERING Cuddalore, India.
- [5] M.Trupthi, Suresh Pabboju," SENTIMENT ANALYSIS ON TWITTER USING STREAMING API" 2017 IEEE 7th International Advance Computing Conference
- [6] Rasika Wagh, Payal Punde," Sentiment Analysis using Twitter Dataset", Department of Computer Science and IT, Dr. BAMU Aurangabad, India.
- [7] <https://en.wikipedia.org/wiki/Twitter>
- [8] Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social MediaData", IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3- 6, 2013, pp 91-96, ISBN: 978-1494673-6068-5
- [9] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in Proceedings of the 40th annual meeting on association for computational linguistics, pp. 417-424, Association for Computational Linguistics, 2002.