

SENTIMENT ANALYSIS USING TWITTER DATA

Kirti Jain¹, Abhishek Singh², Arushi Yadav³

¹Asst. Professor, Dept. Of Computer Science, Inderprastha Engineering College

^{2,3}Student, Dept. Of Computer Science, Inderprastha Engineering College

Dr. A. P. J. Abdul Kalam Technical University

Abstract - Sentiment analysis as the name suggest it is the analysis of the sentiments/feelings/expression related to any topic, it is also known as opinion mining. Here the motive is to find the general sentiment associated to given document. We try to classify the subjective information gathered from some microblogging site according to its polarity such as positive, neutral and negative using machine learning and natural language processing. In this project, we chose Twitter as the microblogging source for getting peoples sentiments and try to classify the tweets into positive, neutral and negative sentiment.

Key Words: Sentiment Analysis, Polarity, Machine Learning, Natural Language Processing, Twitter, Microblogging.

1. INTRODUCTION

Sentiment analysis studies people judgment or thought towards certain entity. Twitter is a resourceful place to find people sentiment. Here peoples post their thought/experience/ feelings through tweets which has a character limit of 140. Twitter has a provision for developers to collect data from twitter by releasing their APIs. In this project we are using one of the twitter API i.e. streaming API which helps to extract the content in the real time.

Here we perform the linguistic analysis by building the classifier using the several machine learning techniques and natural language processing by using the collected corpus from the Kaggle as the training data to train our classifier and use the streamed corpus as the testing data to test the result of our classifier to classify the different sentiments related to tweets.

In this project we focus on the tweets related to airline as the customer shares their experience on the twitter thorough their tweets and our analyzer helps the airline company to improve their services by keeping an eye on people's sentiments by overcoming their flaws.

2. LITERATURE REVIEW

Table -1: Literature Survey Table

S. NO	Paper Title	Authors	Year	Methods	Remarks
1.	Sentiment Analysis on Twitter Data	Varsha Sahayak, Vijaya Shete, Apashabi Pathan	2015	Naïve Bayes, Maximum Entropy, SVM	In the survey, we found that social media related features can be used to predict sentiment in Twitter.

2.	Sentiment Analysis on Twitter Data	Onam Bharti, Mrs. Monika Malhotra	2 0 1 6	Naïve Bayes, KNN	Accuracy (%) Naive Bayes- 79.66 KNN – 83.59
3.	Sentiment analysis of twitter data	Hamid Bagheri, Md Johirul Islam	2 0 1 7	Naïve Bayes, Text Blob	We realized that the neutral sentiments are significantly high which shows there is a need to improve Twitter sentiment analysis.
4.	Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python	Bhumika Gupta, Monika Negi, Kanika Vishwakarma, Goldi Rawat	2 0 1 7	Bayesian logistic regression ,Naïve Bayes, Maximum Entropy Classifier , Support Vector Machine Algorithm	This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used.
5.	Sentiment Analysis of Twitter Data through Big Data	Anusha.N, Divya.G, Ramya.B	2 0 1 7	<i>Naïve Bayes Classification, Training with Mahout</i>	After the training set has been prepared, data is analyzed by uploading it on HDFS and Naive Bayes classification is carried out.
6.	Machine Learning-Based Sentiment Analysis for Twitter	Ali Hasan, Sana Moin, Ahmad Kari and Shahaboddin Shamshirband	2 0 1 8	Naïve Bayes Classifier, SVM Classifier	This paper focuses on the adoption of machine-learning algorithms to determine the highest accuracy for election sentiments.

3. CORE MODULES

These are the modules which are used in our project along with their functions are given in the table below:

Table -2: Core Modules

MODULES	FUNCTION
Twitter_data_streaming	It uses tweepy module for streaming live tweets using Oauth Handler.
Create_dataset	It creates training and testing data set from the dataset downloaded from kaggle.
Create_validation_dataset	Using the tweets downloaded, validation dataset is generated.
Feature_extraction_word2vec	Finds most similar words present in file.
Term_frequency	Feature extraction using TF-IDF(term frequency-inverse document frequency)
Term_frequency_computaion	Feature extraction using bag of words approach
Naïve_bayes_Classifier	Building classifier using naïve bayes algorithm
SVM_classifier	Building classifier using SVM approach
Data_setup_neural_network	It adds weight, if positive then 1 then if negative then 0 if neutral then 2
Neural_networks	Using training data set creates a model using backward propagation and predicts the results i.e. by improving weights
Classification	Using training data set model created using SVM, Naïve Bayes and neural networks and the model is

	tested on testing set and validation set and accuracy is also computed.
Tweets_extraction	Extract tweets to respective documents documents.

4. RESULT

In this section we present the results of our three classifier i.e. Naïve Bayes, SVM, Neural Network.

The accuracy of all the three classifier is computed and shown below:

Table -3: Results

Classifier	Accuracy (%)
SVM Classifier	81.57
Naïve Bayes	71.05
Neural Networks	44.73

As the SVM has the highest accuracy that's why we use this classifier to classify our tweets and the result is as follows:

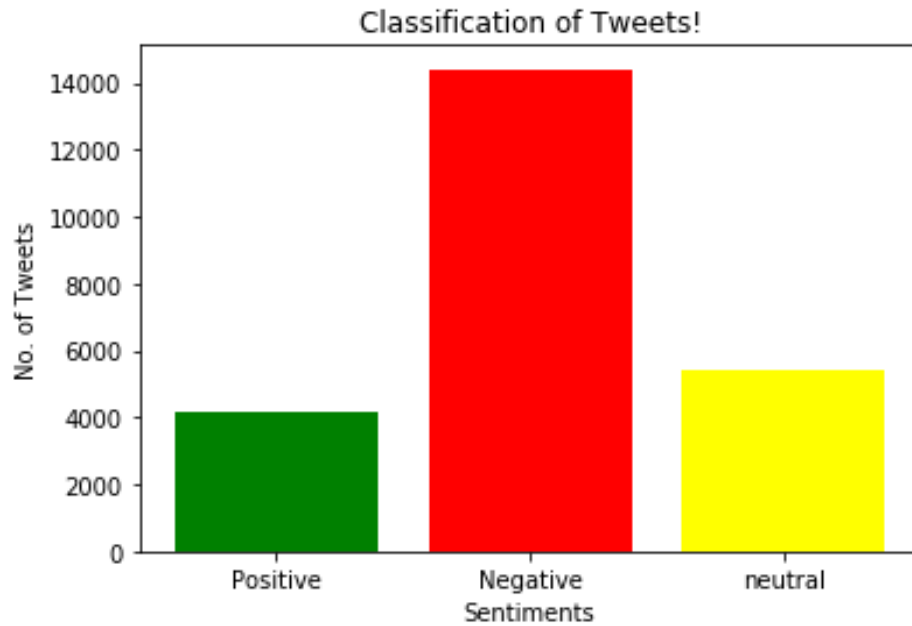


Chart -1: Sentiment Analysis Bar Graph

The total no. of tweet was 23986 out of which 4153 were positive, 14407 were negative and 5426 were neutral i.e.17.31%, 60.06% and 22.62% this much percentage hold by positive, negative and neutral tweets out of total no. of tweets respectively.

5. CONCLUSION

We presented our result on sentiment analysis using twitter data. We use proposed models of classification in the supervised machine learning i.e. Naïve Bayes, SVM, and Neural Networks. For feature extraction we used the bag-of-words, term frequency, tf-idf approach. And by combining them we classify the tweets into three different sentiment i.e. Positive, Neutral and Negative.

In future work we try to improve the accuracy of our classifier which can detect the sarcasm, irony, humor content in the tweet.

ACKNOWLEDGEMENT

We take this opportunity to thank our teachers and friends who helped us throughout the project. First and foremost we would like to thank my guide for the project (Ms. Kirti Jain, Assistant Professor, Computer Science and Engineering) for her valuable advice and time during development of project.

We would also like to thank Dr. Rekha Kashyap (HOD, Computer Science Department) for her constant support during the development of the project.

REFERENCES

[1] Varsha Sahayak, Vijaya Shete „Apashabi Pathan,” Sentiment Analysis on Twitter Data”,Department of Information Technology, Savitribai Phule Pune University, Pune, India,2015.

- [2] Onam Bharti, Mrs. Monika Malhotra, "SENTIMENT ANALYSIS ON TWITTER DATA", World College of Technology and Management, Gurgaon , India, 2016.
- [3] Bhumika Gupta, Monika Negi, Kanika Vishwakarma, Goldi Rawat, Priyanka Badhani , " Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python " , C.S.E.D ,G.B.P.E.C, Pauri, Uttarakhand, India, 2017.
- [4] Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts" , Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam, 2018.
- [5] Hamid Bagheri, Md Johirul Islam, "Sentiment analysis of twitter data" , Computer Science Department Iowa State University, United States of America, 2017.
- [6] Anusha.N, Divya.G, Ramya.B, "Sentiment Analysis of Twitter Data through Big Data", Computer Science and Engineering Sai Vidya Institute of Technology Bangalore, India, 2017.