



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com



A Survey on Twitter Sentiment Analysis by Using Machine Learning

Parimal Ganorkar, Dhiraj Kaulkar, Nayan Sonkusre, Kartik Tayade, Prof. Pratik Deshmukh

U.G. Student, Department of Computer Science and Engineering, Prof.Ram Meghe College of Engineering and

Research Badnera Amravati, India

Associate Professor, , Department of Computer Science and Engineering, Prof.Ram Meghe College of Engineering and

Research Badnera Amravati, India

ABSTRACT: Investigation of open data from internet-based expressions and opinions could yield fascinating outcomes and bits of knowledge into the universe of popular feelings about any item, administration or identity. The blast of Web 2.0 has prompted expanded action in Podcasting, Blogging, Tagging, Social Bookmarking, and Social Networking. Subsequently there has been a sudden increase of enthusiasm for individuals to mine these tremendous assets of information.

In this project we will talk about a procedure which permits use and understanding of twitter information for sentiment analysis. We perform several steps of text pre-processing, and then experiment with multiple classification mechanisms using a dataset of 50000 tweets. We compare the accuracy obtained using various machine learning classifiers for this task. Sentiment analysis classifier could be useful for many applications like market analysis of different features of a new product or public opinion for a new movie or speech by a political candidate.

KEYWORDS: Natural Language Processing, Machine Learning, Text PreProcessing.

I. INTRODUCTION

A popular social media platform is Twitter, a micro-blogging site that allows users to write textual entries of up to 140 characters, commonly referred to as tweets. As of June 2015, Twitter has over 302 million monthly active users according to their homepage, where as approximately 88 % have their tweets freely readable. Additionally, Data created by Twitter is made available through Twitter's API, and represents a real-time information stream of opinionated data. Performing Natural language processing on textual data from Twitter presents new challenges because of the informal nature of this data. Tweets often contain misspellings, and the limit of 140 characters encourages slang and abbreviations. Additionally, tweets contain special features like hash tags that may have an analytical value. Hash tags are labels used for search and categorization, and are included in the text prepended by a '#'. Finally, if a tweet is a reply or is directed to another Twitter user, mentions can be used by prepending a username with '@'. Twitter as of now has three unique variants of APIs accessible, to be specific the REST API, SEARCH API, and STREAMING API. With the REST API, engineers can assemble status information and client data. The Search API enables designers to inquiry explicit Twitter content, while the Streaming API can gather Twitter content progressively.

Analyzing the public sentiment is important for many applications such as firms trying to find out the response of their products in the market, predicting political elections and predicting socioeconomic phenomena like stock exchange. The aim of this project is to develop a functional classifier for accurate and automatic sentiment classification of an unknown tweet stream.

II. RELATED WORK

Pang B. and Lee L in 2004, expressed that analysis of sentiments tries recognizing the perspectives of a basic content span; an example is classifying a film audit as positive or negative. To decide the polarity, the creator defines a novel machine-learning strategy which applies content classification strategies to only the subjective parts of the text document. Separating these parts can be executed utilizing proficient strategies for discovering least cuts in graphs which significantly encourages consolidation of cross sentence relevant limitations [1].

Bing Liu. 2012 expressed that Opinions are vital to every single human activity and are key influencers of our practices. This isn't valid for people yet additionally valid for organizations. Opinions and its connected ideas, for example, feelings are the concern of investigation of sentiments. The initiation and fast development of the field match the online networking e.g., surveys, sites, micro blogs, Twitter, and social networks, on the grounds that without precedent for mankind's history, an immense volume of obstinate information is being recorded in computerized frames[2].

Gautam G. and Yadav P in 2014 define another method for communicating the feelings and emotions of people. In general it is a way which tremendously measures the information where clients can see the emotions of different clients which are categorized into various classes of sentiments and are progressively developing as a main factor in basic leadership. The work defined is useful to view the data as the quantity of tweets where sentiments or emotions are either good or bad, or neutral.[4]

Hassan S. described in their paper that analysis of sentiment over Twitter gives companies a very best way to analyze the public's sentiment for their brand, business, products, etc. An extensive variety of features and techniques for preparing opinion classifiers for Twitter datasets have been inquired about as of late with varying [4].

III. METHODOLOGY

Building a classifier can be used to implement supervised approaches. Examples that can be manually labeled are used to train this classifier. Support Vector Machines (SVM), Random Forest Classifier, Naive Bayes classifier, and Maximum Entropy are the most commonly used supervised algorithms. Supervised Techniques have been found to outperform unsupervised techniques in terms of performance [1]. SVMs, according to Cui et al. [2], are better for sentiment classification since they may perform better when a review contains both positive and negative words. When the amount of training data is limited, however, a Naive Bayes classifier may be preferable because SVMs require a big amount of data to develop a highquality classifier. Selecting an adequate set of features is one of the most critical jobs in sentiment categorization. Below are the most widely used features in sentiment categorization.

- Term presence and their frequency: Unigrams or n-grams, as well as their frequency or presence, are among these characteristics. In sentiment categorization, these features have been frequently and successfully used. In movie review sentiment analysis, Pang et al. [1] say that unigrams produce better results than bi-grams, however, Dave et al. [3] argue that bigrams and tri-grams produce superior product-review polarity categorization.

- Part of speech information

The part-of-speech method is used to decipher sense, which is then used to guide feature selection [4]. Part-of speech tagging is effective for detecting adjectives and adverbs in sentences that identify opinion words and nouns that define product characteristics

- Negations: Negation is a key aspect to consider because it has the ability to reverse a sentiment [4]. "The movie is great" and "The movie is not fantastic," for example. The negation word "not" in the second statement makes it negative.

- Opinion words and phrases: Words and phrases expressing favorable or negative opinions include "like," "lovely," "hate," "I'd propose," and "I'd suggest..." Statistical and lexicon-based approaches are the most common methods for determining the semantic orientation (positive or negative) or polarity of opinion words.

IV CONCLUSION

The use of sentiment analysis to harvest massive amounts of data has become a major academic topic. This document outlines some of the most often utilized sentiment analysis applications and obstacles. Now, commercial groups and academia are collaborating to develop the finest sentiment analysis technology. Although various algorithms have been utilized in sentiment analysis and have produced positive results, no algorithm can tackle all of the problems. Random Forest Algorithm have higher accuracy than other algorithms, according to the



majority of academics The classification of sentiment is found to be domain-dependent. Different types of classification algorithms should be integrated to overcome their unique flaws, profit from each other's strengths, and improve classification accuracy. Such apps are in high demand in the market since every company wants to know how customers feel about their products and services, as well as those of their competitors. New applications for sentiment analysis can be created. Although sentiment analysis techniques and algorithms have progressed, there are still many obstacles to overcome in this discipline. More research can be done in the future to address these issues.

REFERENCES

- [1] Pang B, Lillian Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts"
- [2] Bing Liu, University of Illinois at Chicago, "Sentiment Analysis and Opinion Mining"
- [3] Joshi, Aditya & R Balamurali & Bhattacharya, Pushpak & Mohanty, Rajat (2011) c-feel It: A sentiment Analyser for Micro Blogs
- [4] Gautam G and Yadav P 2014 "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis"



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details