International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT23903117

Twitter Sentiment Analysis

Gunjan R. Patil[1], Rishiraj Shrivastava[2], Advin Manhar[3]

[1,2]Student, Department of Computer Science Engineering, Amity University, Raipur, Chhattisgarh, India

[3]Assistant Professor, Department of Computer Science Engineering, Amity University, Raipur, Chhattisgarh, India

ARTICLEINFO

Article History:
Accepted: 03 June 2023
Published: 19 June 2023

Publication Issue
Volume 9, Issue 3
May-June-2023

Page Number
520-528

ABSTRACT

Social media has entered further attention currently. Public and private opinions about a wide variety of subjects are expressed and spread continually via multitudinous social media. Twitter is one of the social media that's gaining trend. Twitter offers associations a fast and effective way to dissect guests ' perspectives toward the critical success in the request place. Developing a program for sentiment analysis is an approach to be used to computationally measure guests' comprehension. This paper reports on the design of the sentiment analysis, rooting a vast quantum of tweets. Prototyping is used in this development. Results classify guests' perspectives via tweets into positive and negative, which is represented in a Graph. Still, the program has been planned to develop on a web operation system, but due to the limitation of Django which can be worked on a Linux Garcon or Beacon, further this approach needs to be done.

Keywords - Twitter, Sentiment, Analysis, Opinion Mining, Social Media, Natural Language Processing

I. INTRODUCTION

According to [1], millions of people are using social network spots to express their feelings, and opinions and expose their diurnal lives. Still, people write anything similar to social conditioning or any comment on products. The online communities give an interactive forum where consumers inform and impact others. Moreover, social media provides an occasion for businesses that gives a platform to connect with their guests similar to social media to announce or speak directly to guests for connecting with client's perspectives of products and services. In discrepancy, consumers have all the power when it comes to what consumers want to see and how consumers respond. With this, the company's success & failure is intimately participated and end up with word of mouth. Still, the social network can change the gestate and decision timber of consumers, for illustration,[2] mentions that 87% of internet users are told in their purchase and decision by client's review. So, if the association can catch up briskly on what their clients think, it would be more salutary to organize to reply on time and come up with a good strategy to contend with their challengers.

A. Problem Statement

Despite the vacuity of software to extract data regarding a person's sentiment on a specific product or service, associations, and other data workers still face issues regarding the data extraction.

• Sentiment Analysis of Web-based operations concentrate on Single Tweet Only.

With the rapid growth of the World Wide Web, people are using social media similar to Twitter which generates big volumes of opinion textbooks in the form of tweets which is available for sentiment analysis[3]. This translates to a huge volume of information from a human standpoint which makes it delicate to prize rulings, read them, dissect tweet by tweet, epitomize them, and organize them into an accessible format in a timely manner[3].

• Difficulty in Sentiment Analysis with Improper English

Informal language refers to the use of colloquialisms and shoptalk in communication, employing the conventions of spoken language [4] similar to' would not ' and ' would. Not all systems are suitable for descry sentiment through the use of informal language and this could hanker the analysis and decision-making process. Emoticons are a pictorial representation of mortal facial expressions [5], which in the absence of body language and prosody serve to draw a receiver's attention to the tenor or temper of a sender's nominal verbal communication, perfecting and changing its interpretation [6]. For illustration, ☺ indicates a happy state of mind. Systems presently in place don't have sufficient data to allow them to draw passion out of the emoticons. As humans frequently turn to emoticons to duly express what they can not put into words[6]. Not being suitable to dissect this puts the association at a loss. Short-form is extensively used indeed with short communication service(SMS). The operation of short- form will be used more constantly on Twitter so as to help to minimize the characters used. This is because Twitter has put a limit on its characters to 1 4 0[7]. For example, ' Tab ' refers to 'to be announced'.

B. Objective

The objects of the study are first, to study the sentiment analysis in micro-blogging which view to dissect feedback from a client of an association's product; and second to develop a program for guests ' reviews on a product that allows an association or existence to sentiment and analyzes a vast quantum of tweets into a useful format.

II. METHODS AND MATERIAL

This design has been divided into 2 phases. First, a literature study is conducted, followed by system development. The literature study involves conducting studies on colorful sentiment analysis ways and systems that are presently in use. In phase 2, operation conditions and functionalities are defined previous to its development. Also, the armature and interface design of the program and how it'll interact are linked. In developing the Twitter Sentiment Analysis operation, several tools are employed, similar to Python Shell 3.8.3 and tablet.

III.LITERATURE SURVEY

A. Opinion Mining

Opinion mining refers to the broad area of natural language processing, textbook mining, and computational linguistics, which involves the computational study of sentiments, opinions, and feelings expressed in text [8]. Although, view or station grounded on emotion rather than reason is frequently colloquially appertained to as a sentiment

[8]. Hence, advancing to an original for opinion mining or sentiment analysis. [9] stated that opinion mining has numerous operation disciplines including account, law, exploration, entertainment, education, technology, politics, and marketing. In earlier days numerous social media have given web users avenues for opening up to express and partake in their studies and opinions[10].

B. Twitter

Twitter is a popular real-time micro-blogging service that allows users to partake in short information known as tweets which are limited to 140 characters[2,3],[11]. Users write tweets to express their opinion about colorful motifs relating to their diurnal lives. Twitter is an ideal platform for the birth of general public opinion on specific issues[9,10]. A collection of tweets is used as the primary corpus for sentiment analysis, which refers to the use of opinion mining or natural language processing[1]. Twitter, with 500 million users and million dispatches per day, has snappily came a precious asset for associations to invigilate their character and brands by rooting and assaying the sentiment of the tweets by the public about their products, services requests, and indeed about challenges [12].[2] stressed that, from the social media generated opinions with the mammoth growth of the world wide web, super volumes of opinion textbooks in the form of tweets, reviews, blogs or any discussion groups and forums are available for analysis, therefore making the world wide web the fastest, most comprising and fluently accessible medium for sentiment analysis.

C. Micro-blogging with E-commerce

A micro-blogging platform similar to Twitter is likewise to a conventional blogging platform just single posts are shorter[13]. Twitter has a small number of words that are designed for the quick transmission of information or exchange of opinion[7]. Still, small businesses or large associations are inaugurated to the eventuality of micro-blogging as an e-commerce marketing tool[3]. Though, the micro-blogging platform has been developed in a few years for promoting foreign trade websites by using a foreign micro-blogging platform as Twitter marketing[3]. The moment of sharing, interactive, community-acquainted features are e-commerce, launched a new bright spot which it can be shown that micro-blogging platform has enabled companies to do brand image, product important deals channel, ameliorate product deals, talk to the consumer for good commerce and other business conditioning involved[2,3][14].[13] said, in fact, the companies manufacturing similar products have started to poll these micro-blogs to get a sense of general sentiment for a product. numerous times these companies study stoner responses and reply to users on micro-blogs [14].

D. Social Media

[15]defined social media as a group of the Internet-grounded operations that produce on the ideological and technological foundations of Web 2.0 which allowed to make and exchange of user-generated content. In a discussion of Internet World Start, [16] linked that a trend of internet users is adding and continuing to spend further time with social media by the total time spent on mobile bias and social media in the U.S.across PC increased by 37% to 121 billion twinkles in 2012, compared to 88 billion twinkles in 2011. On the other hand, businesses use social networking spots to find and communicate with guests, businesses can be demonstrated damage to productivity caused by social networking[17]. As social media can be posted so fluently to the public, it can harm private information to spread out in the social world[11].

On the other hand,[18] said that the benefits of sharing in social media have gone further simply social

sharing to make association's character and bring in career openings and financial income. In addition,[15],[35] mentioned that social media is also being used for announcements by companies for elevations, professionals for searching, and retaining, social literacy online, and electronic commerce.Electronic commerce or e-commerce refers to the purchase and trade of goods or services online via social media, similar to Twitter which is accessible due to its 24-hour vacuity, ease of client service, and global reach[19]. Among the reasons of why business tends to use further social media is for getting sapience into consumer behavioral tendencies, request intelligence and present an occasion to learn about client review and comprehension.

E.Twitter Sentiment Analysis

The sentiment can be set up in the commentary or tweet to give useful pointers for numerous different purposes[20]. Also,[12] and[36] stated that sentiment can be distributed into two groups, which are negative and positive words. Sentiment analysis is a natural language processing way to quantify an unexpressed opinion or sentiment within a selection of tweets[8]. Sentiment analysis refers to the general system to prize opposition and subjectivity from semantic exposure which refers to the strength of words and opposition textbook or expressions[19]. There are two main approaches for rooting sentiment automatically which are the Lexicon-based Approach approach and Machine Learning-based approach.

1) Lexicon-based Approach

The lexicon-based approach makes use of a predefined list of words where each word is associated with a specific sentiment[21]. The wordbook styles vary according to the environment in which they were created and involve calculating exposure for a document from the semantic exposure of textbooks or expressions in the documents[19]. Also,[24] also states that a lexicon sentiment is to descry word-carrying opinion in the corpus and also to prognosticate opinion expressed in the text.[20] has shown the wordbook styles which have a basic paradigm which are:

i. Preprocess each tweet, and post by removing punctuation.
ii. Initialize a total opposition score(s) equal 0-> s = 0
iii. Check if the commemorative is present in a dictionary, then if the commemorative is positive, s will be positive, if the commemorative is negative, s will be negative(-)
iv. Look at the total opposition score of tweet post
if s> threshold, tweet post as positive
if s< threshold, tweet post as negative

Still,[21] stressed one advantage of a learning-based system, is that it can acclimatize and produce trained models for specific purposes and surroundings. In discrepancy, a vacuity of labeled data and hence the low connection of the system of new data which is because labeling data might be expensive or indeed constrictive for some tasks[21].

2) Machine-Learning-based Approach

Machine-Learning-based styles frequently calculate on supervised bracket approaches where sentiment discovery is framed as a binary that is positive and negative[24]. This approach requires labeled data to train classifiers[21]. With this approach, it becomes apparent that aspects of the original environment of a word need to be taken into account similar to negative(e.g. Not beautiful) and intensification(e.g. very beautiful)[19]. still,[20] showed a basic paradigm to produce a feature vector to:
i. Create a part of speech tagger for each tweet post
ii. Accumulate all the adjectives for entire tweet posts
iii. Make a popular word set composed of the top N adjectives
iv. Traverse all of the tweets in the experimental set to produce the following:

• Number of positive words
• Number of negative words
• Presence, absence, or frequency of each word

[19] showed some examples of switch negation, negation simply to reverse the polarity of the lexicon: changing beautiful (+3) into not beautiful (-3). More examples: She is not terrific (6-5=1) but not terrible (-6+5=-1) either.

The negation of a strongly positive or negative value in this instance represents a mixed perspective, which is accurately conveyed in the shifted value. The limitation of a machine-learning-based approach, as opposed to a lexical-based one, has been cited by [21] as being more appropriate for Twitter. Furthermore, according to [20], machine learning techniques can produce a predetermined number of the most often occurring popular words, with each word's frequency on Twitter represented by an integer value.

F.Ways of Sentiment Analysis

The semantic generalities of realities uprooted from tweets can be used to measure the overall correlation of a group of realities with a given sentiment opposition[12]. Polarity refers to the most elementary form, which is if a text or judgment is positive or negative[25]. Still, sentiment analysis has ways of assigning polarity similar as:

1) Natural Language Processing(NLP):NLP techniques are grounded on machine learning and especially statistical literacy which uses a general learning algorithm combined with a large sample, a corpus, of data to learn the rules[26]. Sentiment analysis has been handled as a Natural Language Processing denoted NLP, in numerous situations of granularity. Starting from being a document position bracket task[27], it has been handled at the judgment position[28] and more lately at the expression position[13]. NLP is a field in computer science that involves making computers decide meaning from human language and input as a way of interacting with the real world.

2) Case-Based Reasoning(CBR): Case-Based Reasoning(CBR) is one of the ways available to apply sentiment analysis. CBR is known by recalling the once successfully answered problems and using the same results to break the current nearly affiliated problems[29].[25] linked some of the advantages of using CBR that CBR doesn't bear an unequivocal sphere model and so elicitation becomes a task of gathering care histories and the CBR system can learn by acquiring new knowledge as cases. This and the operation of database ways make the conservation of large columns of information easier[25].

3) Artificial Neural Network(ANN): [13]mentioned that Artificial Neural Network(ANN) or known as neural network is a fine technique that interconnects a group of artificial neurons. It'll reuse information using the connections approach to computation. ANN is used in discovering the relationship between input and output or to find patterns in data[25].

4) Support Vector Machine(SVM): Support Vector Machine is to descry the sentiments of tweets[23].[10] together with[37] stated SVM is suitable to prize and dissect to gain up to 70-81.3% of delicacy on the test set. [29]collected training data from three different Twitter sentiment discovery websites which substantially use pre-built sentiment dictionaries to mark each tweet as positive or negative. Using SVM trained from these noisy labeled data, they attained 81.3% in sentiment bracket delicacy.

G. Application Programming Interface(API)

Alchemy API performs better than the others in terms of the quality and the volume of the uprooted entities [14]. As time passed the Python Twitter Application Programming Interface(API) is created by collected tweets [30]. Python can automatically

calculate the frequency of dispatches being re-tweeted every 100 seconds, sorted the top 200 dispatches grounded on their- twittering frequency, and stored them in the designated database [12]. As the Python Twitter API only included Tweets collected for the most recent six days, the data demanded to be stored in a different database [14].

H.Python

Python was set up by Guido Van Rossum in the Netherlands, in 1989 which has been public since 1991[31]. Python is a programming language that is available and solves a computer problem which is furnishing a simple way to write out a result[31].[32] mentioned that Python can be called a scripting language. Also[32], supported that actually, Python is a just description of a language because it can be written and run on numerous platforms. Moreover,[34] mentioned that Python is a language that's great for writing a prototype because Python is lower time-consuming and working prototype handed, discrepancy with other programming languages. Numerous experimenters have been saying that Python is effective, especially for a complex design, as[33] has mentioned that Python is suitable to start up social networks or media storming systems which utmost always are web-based which is driving big data. Besides Python creates a generator that allows an iterative process of effects, one item at a time, and allows the program to snare source data one item at a time to pass through the full processing chain.

IV. RESULTS AND DISCUSSION

A. Twitter Retrieved

To associate with Twitter API, the inventor needs to agree to the terms and conditions of the development Twitter platform which has been handed to get authorization to access data. The output from this will be saved in the JSON file. The reason is, JSON(JavaScript Object Notation) is a feather-light data-interchange format which is easy to write and read. Also, stated that JSON is simple for machines to induce and parse. JSON is a text format that's completely language-independent but uses a tradition that's known to programmers of the C- family of languages, including Python and numerous others. Still, the output size depends on the time for reacquiring tweets from Twitter. Nonetheless, the affair will be distributed into 2 forms, which are decoded and unencoded. According to security issues for penetrating data, some of the affairs will be shown in an ID form similar to string ID. The tweets will be allocated the value of each word, which are classified into positive and negative words, according to the lexicon dictionary. The result will be shown in. txt, .csv, and HTML. All paragraphs must be indented. All paragraphs must be justified, i.e. both left justified and right-justified.

B. Sentiment Analysis

Tweets from the JSON file will be given the value of each word by checking with the lexicon dictionary. As a limitation of words in the lexicon dictionary it isn't suitable to assign a value to every single word from tweets. Still, as a scientific language of Python, which is suitable to dissect a sense of each tweet into positive or negative for getting a result.

C. Information Presented

The result will be shown in a graph which is representing a chance of positive, negative, and neutral sentiment hash markers. For neutral marker, the label is representing the hash markers that were assigned zero value.

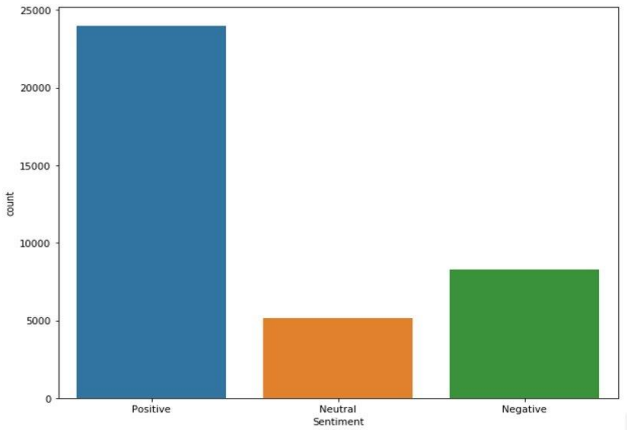

Fig. 1) Graph

As shown in Fig., the graph is representing each of the count of positive, negative and neutral sentiments in different color.

V. CONCLUSION

Twitter sentiment analysis is developed to analyze customers' viewpoints towards the critical to success in the marketplace. The program is using a machine-learning based approach which is more authentic for analyzing a sentiment; together with which NLP techniques will be used. As a result, tweets will be categorized according to the sentiments into positive and negative, which is represented in a graph and HTML page. Even though, the program has been planned to be developed as a web application, due to the limitation of Django which can only work on Linux servers. Thus, it cannot be realized. Therefore, further enhancement of this feature is recommended in future studies.

VI. REFERENCES

[1]. Rambocas, and J. Gama, 'Marketing Research:The Role of Sentiment Analysis'. The 5th SNA-KDD Workshop'11. U nivers ityof Porto, 2013

[2]. A. K. Jose, N. Bhatia, and S. Krishna, 'TwitterSentimentAnalysis'. NationalInstituteof TechnologyCalicut, 2010.

[3]. P.Lai,'ExtractingStrongSentimentTrendfromTwitter'. Stanford University, 2012.

[4]. Y. Zhou, and Y. Fan, ' A Sociolinguistic Study of American Slang,' Theory and Practice in Language Studies, 3(12), 2209–2213, 2013. doi:10.4304/tpls.3.12.2209-2213

[5]. M. Comesaña, A. P.Soares, M.Perea, A.P. Piñeiro, I. Fraga, and A. Pinheiro, ' Author ' s personal copy Computers in Human Behavior ERP correlates of masked affective priming with emoticons,' Computers in Human Behavior, 29, 588–595, 2013.

[6]. A.H.Huang, D.C. Yen, & X. Zhang, 'Exploring the effects of emoticons,' Information & Management, 45(7), 466–473, 2008.

[7]. D. Boyd, S. Golder, & G. Lotan, 'Tweet, tweet, retweet: Conversational aspects of retweeting on twitter,' System Sciences (HICSS), 2010. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5428313

[8]. T. Carpenter, and T. Way, 'Tracking Sentiment Analysis through Twitter,'. ACM computer survey. Villanova:VillanovaUniversity, 2010.

[9]. D. Osimo, and F. Mureddu, 'Research Challenge on Opinion Mining and Sentiment Analysis,' Proceeding of the 12th conference of Fruct association, 2010, United Kingdom.

[10]. A. Pak,and P. Paroubek, 'Twitter as a Corpus for Sentiment Analysis and Opinion Mining,' Special Issue of International Journal of Computer Application, France:Universite de Paris-Sud, 2010.

[11]. S.Lohmann, M. Burch, H. Schmauder and D. Weiskopf, 'Visual Analysis of Microblog Content Using Time-Varying Co-occurrence Highlighting in Tag Clouds,' Annual conference of VISVISUS. Germany: University of Stuttgart, 2012.

[12]. H. Saif, Y.He, and H. Alani, 'SemanticSentimentAnalysisof Twitter,' Proceeding of the Workshop on Information

Extraction and Entity Analytics on Social Media Data. United Kingdom: Knowledge Media Institute, 2011.

[13]. A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R.Passonneau, 'Sentiment Analysis of Twitter Data,' Annual International Conferences. New York:Columbia University, 2012.

[14]. J. Zhang, Y. Qu, J. Cody and Y. Wu, ' A case study of Microblogging in the Enterprise: Use, Value, and Related Issues,' Proceeding of the workshop on Web 2.0., 2010.G. Kalia, 'A Research Paper on Social Madia: An Innovative Educational Too', Vol.1, pp. 43-50, Chitkara University, 2013.

[15]. Internet World Start, 'Usage and Population Statistic', Retrieved 10 15, 2013 from: http://www.internetworldstats.com/stats.htm

[16]. A.M. Kaplan, and M, Haenlein, 'Users of the world, unite! The challenges and opportunities of Social Media,' France: Paris, 2010.

[17]. Q. Tang, B. Gu, and A.B. Whinston, 'Content Contribution in Social Media: The case of YouTube', 2nd conference of social media. Hawaii: Maui, 2012.

[18]. M.Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, ' Lexicon- Based Methods for Sentiment Analysis,' Association for Computational Linguistics, 2011.

[19]. M. Annett, and G. Kondrak, 'A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs,' Conference on web search and web data mining (WSDM). University of Alberia: Department of Computing Science, 2009.

[20]. P. Goncalves, F. Benevenuto, M. Araujo and M. Cha, 'Comparing and Combining Sentiment Analysis Methods', 2013.

[21]. E. Kouloumpis, T. Wilson, and J. Moore, 'Twitter Sentiment Analysis:The Good the Bad and theOMG!', (Vol.5). International AAAI, 2011.

[22]. S. Sharma, 'Application of Support Vector Machines for Damage detection in Structure,' Journal of Machine Learning Research, 2008.

[23]. A.Sharma, and S. Dey, 'Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis,' Association for the advancement of Artificial Intelligence, 2012.

[24]. J. Spencer and G. Uchyigit, 'Sentiment or: Sentiment Analysis of Twitter Data,' Second Joint Conference on Lexicon and Computational Semantics. Brighton:University of Brighton, 2008.

[25]. A. Blom and S. Thorsen, 'Automatic Twitter replies with Python,'International conference 'Dialog 2012'.

[26]. B. Pang, and L. Lee, 'Opinion mining and sentiment analysis,' 2nd workshop on making sense of Microposts. Ithaca: Cornell University. Vol.2(1), 2008.

[27]. M. Hu, and B. Liu, 'Mining and summarizing customer reviews,' 2004.

[28]. P. Nakov, Z. Kozareva, A. Ritter, S. Rosenthal, V. Stoyanov, T. Wilson, Sem Eval-2013 Task2:Sentiment AnalysisinTwitter (Vol.2,pp. 312-320 ,2013.

[29]. J. Wu, J., Wang, & L. Liu, 'Kernel-Based Method for Automated Walking Patterns Recognition Using Klnematics Data'. 5th Workshop on Natural Language Processing. China: Xi'an Jiaotong University. 2006.

[30]. T. D. Smedt, and W. Daelemans, 'Pattern for Python,' Proceeding of COLING. Belgium: University of Antwerp, 2012.

[31]. A. Sweigart, 'Invent your own computer games with Python. 2nd edition, 2012. Retrieved from http://inventwithpython.com/

[32]. C. Seberino, 'Python. Faster and easier software development,' Annual Conference. California: San Diego, 2012.

[33]. A.Lukaszewski, 'MySQL for Python. Integrate the flexibility of Python and the power of

MySQL to boost the productivity of your applications,' UK: Birningham. Packt Publishing Ltd, 2010.

[34]. V. Nareyko, 'Why python is perfect for startups,' Retrieved 01 10, 2014 from: http://opensource.com/business/13/12/why-python-perfect-startups

[35]. A. Hawkins, 'There is more to becoming a thought leader than giving yourself the title'. Retrieved 10 18, 2013. from: http://www.thesocialmediashow.co.uk/author/admin/

[36]. R. Prabowo, and M. Thelwall, 'Sentiment Analysis:A Combined Approach,' International World Wide Web Conference Committee (IW3C2), 2009. UnitedKingdom:Universityof Wolverhamption.

[37]. H. Saif, Y. He and H. Alani, 'Alleviating Data Scarcity for Twitter Sentiment Analysis'. Association for Computational Linguistics, 2012

Cite this article as :

Gunjan R. Patil, Rishiraj Shrivastava, Advin Manhar, "Twitter Sentiment Analysis", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp., May-June-2023.
doi : https://doi.org/10.32628/CSEIT23903117