International Journal for Research in Science Engineering and Technology

REVIEW ON SENTIMENT ANALYSIS OF TWITTER DATA USING DIFFERENT TECHNIQUES

[1] J. Uma, [2] Dr. K. Prabha,
[1] Research scholar, [2] Assistant Professor,
[1,2] Department of computer science,
[1,2] Periyar University PG Extension centre,
[1,2] Dharmapuri.

ABSTRACT- Sentiment analysis in Twitter tackles the issue of investigating the tweets regarding the opinion they express. This survey centers chiefly around sentiment analysis of twitter information which is useful to investigate the data in the tweets where opinions are profoundly unstructured, heterogeneous and are either sure or negative, or unbiased now and again. In this paper, we give an audit and relative investigations of existing techniques like machine learning, profound learning and Hybrid strategy. Utilizing different machine learning algorithms like Naive Bayes, lexicon-based, and Support Vector Machine, we give research on twitter information streams. We have additionally examined merits and bad marks of existing techniques of Sentiment Analysis on Twitter.

Keywords: [Sentiment analysis, Twitter, opinion, machine learning algorithms.]

1. INTRODUCTION

Sentiment analysis

Sentiment analysis manages distinguishing and grouping opinions or sentiments which are available in source text. Web-based media is producing a colossal measure of sentiment rich information as tweets, notices, reviews and blog posts and so forth Sentiment analysis of this client created information is exceptionally helpful in knowing the opinion of the group. Twitter sentiment analysis is laborious when contrasted with fundamental sentiment analysis because of the presence of slang words and misspellings. The greatest furthest reaches of characters that are permitted in Twitter is 140. Machine learning approach can be utilized for breaking down sentiments from the content.

Twitter

Microblogging is an organization administration with which clients can share messages, connections to external websites, images, or videos that are obvious to clients bought in to the administration. Messages that are posted on microblogs are short rather than customary blogs. As of now, various distinctive microblogging platforms are accessible, including Twitter,6 Tumblr,7 FourSquare,8 Google+,9 and LinkedIn.10 One of the most mainstream microblogs is Twitter, which was dispatched in 2006 and from that point forward has pulled in countless clients. At present, Twitter has 284 million clients who post 500 million messages for every day. Because of the way that it gives a simple method to get to and download distributed

posts, Twitter is viewed as one of the biggest datasets of client produced content.

Machine Learning Approach

Machine learning (ML) is a scientific discipline that investigates the development and the investigation of algorithms that can gain from information. Applied rule-based, supervised, and semi-supervised techniques, and gathered tweets about the president (Obama) to gauge the sentiment of individuals' opinion towards his occupation execution. Moreover, a cross-correlation analysis of time arrangement was explored to foresee sentiments by naming 2500 tweets to anticipate the test dataset of 550,000 unlabeled tweets. A mixture technique had been utilized since a serious classifier was utilized for sentiment analysis "The Latent Dirichlet Allocation Model", where a subject has probabilities of producing different words; first, the understood topical structure was separated from the tweets, second, 32 million tweets were broke down to anticipate the US official appointment of 2012. Semantic element by adding a semantic idea to every element in the tweet to anticipate the sentiments for the gathered dataset. The emoticons include alongside the twitter messages by utilizing The Distant Supervised Learning algorithm.

2. LITERATURE REVIEW

Naive bayes

Naive bayes theorem is a classification strategy with the autonomous supposition between the predictors. As such the approach of specific indicator of one class isn't associated with closeness of some different class. How about we take a case, an apple might be viewed as a natural product in the event that it is red in shading, and on the off chance that it is round fit as a fiddle and if its diameter is consider being three inches around. Notwithstanding of these component are reliant on each other or within the sight of another element. All these autonomous properties add to discover the likelihood of naïve classifier that this is an apple. Naive bayes is valuable for huge informational

collections and can be manufacture without any problem.

1. Huma Parveen , Prof. Shikha Pandey (2017) proposed a Hadoop Framework for processing film data set that is accessible on the twitter website as reviews, feedback, and comments. Hadoop works cluster without losing data or 'of any work. HDFS assists with overseeing cluster by breaking approaching records into little piece called block. Consequences of sentiment analysis on twitter data will be shown as various segments introducing positive, negative and neutral sentiments. It examines the extraction of sentiment from an acclaimed microblogging website, Twitter where the user posts their perspectives and opinion. We have done sentiment analysis on tweets which help to give some expectation on business intelligence. Comments, reviews and opinion of individuals assume a significant part to decide if a given population is happy with the item, administrations.

Merits:

Hadoop is the most ideal approach to settle Big Data challenges.
It aides in predicting the sentiment of a wide assortment of individuals on a specific function of intrigue like the survey of a film, their opinion on different subjects wandering the world over.

Demerits:

For real time analytics, the accessible framework is wasteful.
It is very time burning-through cycle to investigate the enormous measure of data in a short timeframe.

2. Ruchi Mehra , Mandeep Kaur Bedi, Gagandeep Singh, Raman Arora, Tannu Bala, Sunny Saxena (2018), presents analysis for sentiment lead of Twitter data. The proposed work utilizes the blameless Bayes and fluffy Classifier to organize Tweets into productive, antagonistic or neural direct of a particular person. It present preliminary appraisal of our dataset and request results which showed that united proposed technique

is progressively successful to the extent Accuracy, Precision and Recall. Sentiment analysis and evaluation mining is the field that investigate people's sentiments, conclusions, emotions from works delivered by the customers. It is the dynamic examination domains in like manner language planning and is similarly broadly inspected in web mining, data mining and web-based social networking examination as sentiments are basic influencers of practices of human.

Merits:
It can be consistently utilized as structures and applied for identification or classification errands.
It can be applied to enormous arrangements of data to set up membership, for this situation inspiration, antagonism and neutral.

Demerits:
If unmitigated variable has a classification in test data set, which was not seen in preparing data set, at that point model will allocate a 0 (zero) probability and will be not able to make an expectation.

Corpus-Based:
The corpus-based approach have goal of giving word references identified with a particular space. These word references are produced from a bunch of seed opinion terms that becomes through the hunt of related words by methods for the utilization of either statistical or semantic techniques.

3. Jaishree Ranganathan, Allen S. Irudayaraj, Angelina A. Tzacheva (2017)
proposed another improved and all the additionally promising framework, to the extent speed and profitability, for making meta-exercises by executing Specific Action Rule disclosure reliant on Grabbing methodology (SARGS) calculation. For this, we play out a general analysis of metaactions making algorithmic execution in Apache Spark driven framework and customary Hadoop driven framework using the Twitter long reach casual correspondence data and evaluate the results. We perform corpus based Sentimental Analysis of long reach relational

correspondence data, and test the supreme time taken by both the frameworks and their sub parts for the data getting ready. Results show faster computational time for Spark framework diverged from Hadoop MapReduce for realizing the meta-action age techniques. Data mining methods are used to analyze massive data sets, to perceive the crucial data plans and to reveal the disguised data. New algorithms have been proposed in the earlier decade to find some remarkable exercises subject to the discovered models as Action Rules. The Action Rules are express data plans eliminated from gigantic dataset which intends to change the current assessment of the versatile quality, reasonable, to an ideal worth.

Merits:
The credits in a dataset are isolated into adaptable ascribes, whose worth is changeable, and stable ascribes, whose qualities is unchanging.

Demerits:
No strong and computerized method of data testing.

Lexicon based method
Lexicon based method utilizes sentiment word reference with opinion words and match them with the data to decide extremity. They doles out sentiment scores to the opinion words depicting how Positive, Negative and Objective the words contained in the word reference are. Lexicon-based approaches essentially depend on a sentiment lexicon, i.e., an assortment of known and precompiled sentiment terms, phrases and even colloquialisms, produced for customary classes of correspondence, for example, the Opinion Finder lexicon;

4. Juan Guevara, Joana Costa, Jorge Arroba, Catarina Silva (2018)
proposed to convey a portable application that gives data zeroing in on zones, for example, Politics, Social, Tourism, and Marketing utilizing a statistical lexicon approach. The application shows the extremity of each subject as positive, negative, or neutral. In this work we

investigated the effect of sentiment analysis in certain points, for example, Politics, Social, Tourism, and Marketing. This work was tended to in Spanish language and with subjects from Ecuador. To this objective, lexicon approach was utilized notwithstanding a word reference of positive and negative words, which considered the extremity classification of each tweet removed by twitter API. The techniques utilized in the processing of text, for example, eliminate URLs, usernames and lower case, have permitted the decrease of highlights or words, just as processing time. Additionally the classification of tweets by the lexical method requests an extraordinary processing time, which implies that more data, all the more processing time.

Merits:
Advantage of being a portable application to introduce the outcomes on screen in a smaller and well disposed manner.
The primary bit of leeway of lexicon-based approaches is that it isn't needed to have a lot of data for preparing to be effortlessly utilized in numerous areas.

Demerits:
The constraints of cell phone processing and the energy utilization related with such processing in cell phones.

5. Hossam Elzayady Khaled M. Badran Gouda I. Salama (2018) proposed a proficient sentiment forecast strategy, using the Apache Spark's Machine Learning library to execute diverse classification algorithms. The outcomes demonstrate a huge upgrade in the accuracy of Naive Bayes and Logistic Regression regarding expanding the volume of dataset, while the improvement isn't solid in Decision Trees, additionally, test's outcomes reason that there is a reverse corresponding connection between running time and the quantity of machines in the Spark Cluster, So in the event of adding additional hubs in the cluster, better capacity will be gotten. From the previous results, our framework can be portrayed as compelling and adaptable. Sentiment analysis can be depicted as a significant part of Natural Language Processing (NLP), its expect to recognize the importance from a report so as to find the extremity of the content. For the sentiment analysis, we concentrate toward the Twitter, a microblogging social networking website, where users can speak with one another or share their opinions in short blogs.

Merits:
Apache Spark created to make processing and investigating the data simpler.

Demerits:
The social database the executives frameworks motors can't handle unstructured or enormous data any longer.

6. Snehal Kale, Vijaya Padmadas (2017) present online and the computerized analysis of such data holds an incredible guarantee in business analytics for offering a solid help in decision making. This paper looks at the very heart of the idea of sentiment analysis by grouping the tweets with the assistance of algorithms like Naïve Bayes, Maximum Entropy, and Negation. In this paper, we first preprocess the tweets to eliminate pointless substance in tweet; we at that point extricate the descriptors which frames the component vector, which are additionally used to discover equivalents utilized in additional semantic calculations in previously mentioned algorithms. Creator distinguishes tweets in a huge unstructured/structured data then the data gathered would be preprocessed. We will at that point dissect their extremity, and classification into various classifications premise the Naïve Bayes and Maximum Entropy, and Negation will be finished. In this, machine learning will be utilized to investigate the tweets and to tag acquired tweet in a predefined classification.

Merits:
To distinguish tweets in an enormous unstructured/structured data then the data gathered would be preprocessed.

Demerits:

It can't discover the progressions needed for the framework's improvement except if we make sense out of it.

Deep Learning:

Deep learning is one of the quickest developing fields of machine learning and is applied to tackle perceptual issues, for example, picture acknowledgment and understanding natural languages. Deep learning utilizes neural networks to learn numerous degrees of deliberation. In text-related tasks, deep-learning approaches ordinarily incorporate two stages. To start with, they take in word embeddings from the content assortment and these are then applied to deliver the portrayals of the records. According to sentiment analysis, deep learning is utilized to take in word embeddings from a lot of text data.

7. Sani Kamış , Dionysis Goularas (2019)

presents a correlation of different deep learning methodologies used for sentiment analysis in Twitter data. At this moment, learning (DL) methods, which contribute at the same time to the plan of a wide extent of issues, picked up reputation among scientists. Particularly, two classes of neural networks are utilized, convolutional neural networks (CNN), which are especially performant in the territory of picture taking care of and irregular neural networks (RNN) which are applied with accomplishment in like manner language getting ready (NLP) tasks. At the present time evaluate and think about outfits and mixes of CNN and a class of RNN the long shortterm memory (LSTM) networks. This paper adds to this zone by surveying the most notable deep learning techniques and game plans reliant on an embraced dataset about sentiment analysis which is produced using Twitter data under a solitary testing framework. we present the dataset, the word installing models with their plans, and the unmistakable deep neural framework arrangements that are used at this moment. GRU networks and RCNN's are rejected considering the way that they give near results with LSTM networks and CNN's. The makers proposed two unmistakable CNN arrangements using assorted word embeddings, Word2Vec and GloVe, independently.

Merits:

RNN-utilized as generative models that require a succession yield, with text, yet on applications, for example, producing penmanship.

Demerits:

It was moderately hard to assess the part of a dataset, a network design or a particular arrangement and tuning.

8. Naveenkumar K S, Vinayakumar R, Soman K P (2019)

presents the show of straight and nonlinear substance portrayal techniques for sentimental analysis. The assembled dataset "Amrita-CEN-SentiDB1" is presented to various non-straight substance portrayal techniques with the deep learning plan which performs better than the immediate substance portrayal with the AI calculations. The presentation of the proposed methodology can be extended probably by hyper boundary tuning the framework. Another reliable dataset is presented to various pre-getting ready methods and subsequently the part extraction frameworks result they are passed to the deep learning methodologies out of which by using the substance portrayal technique, worldwide vectors (glovec) with the long transient memory (lstm) has the most noteworthy precision of 75%, which is the benchmark precision for this dataset. In the current technique the AI is making room in the various fields, one such is the normal language planning, here the data used for the dealing with is the substance data.

Merits:

The capacity to comprehend the content through the algorithms that are only accomplished for them.

Demerits:

They don't give consecutive access; this might be worthwhile in certain algorithms however in others it is a disaster.

9. Yaser Maher Wazery, Hager Saleh Mohammed, Essam Halim Houssein (2018) proposed RNN-LSTM on twitter dataset to arrange evaluation of people into positive and negative and differentiating the results of precision and assorted AI calculations, for instance, Support Vector Machine, K-Nearest Neighbor, Na¨ıve Bayes, and Decision Tree. The disclosures of the paper showed that RNN-LSTM had achieved the most noteworthy precision at 88% for Amazon dataset, 87% for IMDB dataset and 93% for Airline dataset. Online journals and casual networks became huge assets for mining sentiments. They are taken consideration of by various people to reveal their points of view around a couple of subjects or things. In this manner, it is a useful resource for associations and relationship to accumulate experiences about their things and to improve results. Likewise, this massive proportion of data has become a wellspring imperative to examiners to hear the perspective of people. Starting late, Deep Neural Network calculations have demonstrated significant overhauls over existing AI approaches in various spaces, expressly talk acknowledgment and PC vision.

Merits:
To arrange the opinion of the clients which are mostly one of three prospects; positive opinion, negative opinion or somewhere close to positive or negative.

Demerits:
Large number of hyper parameters.
Training huge neural nets on enormous issues takes quite a while.
Non-raised misfortune capacities..

Hybrid Methods
Various specialists consolidated machine-learning and lexicon-based approaches.

10. Ketaki Gandhe , Aparna S. Varde , Xu Du (2018) proposed approach includes a creamer learning technique for portrayal of tweets subject to a Bayesian probabilistic procedure for sentence level models given not entirely named planning data. For execution, we use AWS to eliminate data from Twitter,

store isolated data in MySQL databases and code Python substance to complete the analyzer. The graphical models are seen using IPython Notebook. The makers depict existing methods and approaches for an end arranged data recuperation. In [4], the makers use Weblogs as datasets for sentiment analysis and use emoticons consigned to blog passages as pointers of clients' perspectives. The makers apply SVM (support vector machines) and CRF (unforeseen discretionary field) understudies to organize sentiments at the sentence level and a short time later exploration a couple of techniques to choose the sentiment. The dataset they use includes about a half million Web pages and a quarter million reports. They don't use Twitter datasets.

Merits:
Conduct assessment on real data applicable to item reviews and metropolitan approach.

Demerits:
It does exclude a blend of classifiers in a group.

CONCLUSION
Ongoing years have witnessed an expanding research enthusiasm for dissecting tweets as indicated by the sentiment they express. This intrigue is a consequence of the huge measure of messages that are posted regularly in Twitter and that contain important data for the public state of mind for various points. In this paper, we give a survey and relative investigation of existing techniques for opinion mining including machine learning, deep learning and hybrid method. This survey serves to specialists to upgrade the analysis of twitter data for explicit functions to quantify the impact and the conduct of users towards various functions classifications.

REFERENCES
[1]. Huma Parveen, Prof. Shikha Pandey (2017), Sentiment analysis on Twitter Data-set using Naive Bayes algorithm", **DOI:** 10.1109/ICATCCT.2016.7912034, **Electronic ISBN:** 978-1-5090-2399-8, IEEE.

[2]. Sani Kamış, Dionysis Goularas (2019), "Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data", 978-1-7281-2914-3/19/$31.00 ©2019 IEEE DOI 10.1109/Deep-ML.2019.00011, IEEE.

[3]. Ruchi Mehra, Mandeep Kaur Bedi, Gagandeep Singh, Raman Arora, Tannu Bala, Sunny Saxena, (2018), Sentimental Analysis Using Fuzzy and Naive Bayes", Proceedings of the IEEE 2017 International Conference on Computing Methodologies and Communication (ICCMC), 978-1-5090-4890-8/17/$31.00 ©2017 IEEE.

[4]. Naveenkumar K S, Vinayakumar R, Soman K P, (2019), "Amrita-CEN-SentiDB1:Improved Twitter Dataset for Sentimental Analysis and Application of Deep learning", 10th ICCCNT 2019 July 6-8, 2019, IIT - Kanpur, Kanpur, India.

[5]. Jaishree Ranganathan, Allen S. Irudayaraj, Angelina A. Tzacheva,(2017), "Action Rules for Sentiment Analysis on Twitter Data using Spark", 2375-9259/17 $31.00 © 2017 IEEE DOI 10.1109/ICDMW.2017.14, 2017 IEEE International Conference on Data Mining Workshops

[6]. Yaser Maher Wazery, Hager Saleh Mohammed, Essam Halim Houssein (2018), "Twitter Sentiment Analysis using Deep Neural Network", **DOI:** 10.1109/ICENCO.2018.8636119, **Electronic ISBN:** 978-1-5386-5117-9, IEEE.

[7]. Ketaki Gandhe, Aparna S. Varde, Xu Du (2018), "Sentiment Analysis of Twitter Data with Hybrid Learning for Recommender Applications", 978-1-5386-7693-6/18/$31.00 ©2018 IEEE

[8]. Juan Guevara, Joana Costa, Jorge Arroba, Catarina Silva (2018), "Harvesting Opinions in Twitter for Sentiment Analysis", **DOI:** 10.23919/CISTI.2018.8399226, **Electronic ISBN:** 978-989-98434-8-6, IEEE.

[9]. Hossam Elzayady Khaled M. Badran Gouda I. Salama (2018), "Sentiment Analysis on Twitter Data using Apache Spark Framework", **DOI:** 10.1109/ICCES.2018.8639195, **Electronic ISBN:** 978-1-5386-5111-7, IEEE.

[10]. Snehal Kale, Vijaya Padmadas (2017), "Sentiment Analysis of Tweets Using Semantic Analysis",

DOI: 10.1109/ICCUBEA.2017.8464011, **Electronic ISBN:** 978-1-5386-4008-1, IEEE.