

# Analyzing Techniques for Tweet Stream Processing for Different Applications

Ashwini Gedekar, Ashwini Zadgaonkar

**Abstract:** *Twitter has become an invaluable source of information to analyze user behaviour in real time. Data knowledge from twitter has surpassed the knowledge given by facebook posts, as people use twitter majorly to spread important information about their life, their community or their profession. In this paper, we analyze various applications which use tweet based data mining in order to produce some informative results about the user's personal behaviour. The applications mentioned in this text make use of certain techniques which range from sentiment analysis to critical event analysis in order to predict any outlier entries which might disrupt personal or public life in general. This paper can act as a stepping stone for early to moderate aged researchers by the introduction of different applications which are possible through tweet analysis. Some applications which are not yet developed are also mentioned in the paper, so that researchers can take advantage of our study and develop techniques on those applications based on data mining.*

**Index Terms:** *Critical ,event, knowledge, mining, sentiment ,tweet*

## I. INTRODUCTION

Social Computing is a creative and developing processing model for the examination and demonstrating of social exercises occurring on different stages. It is utilized to deliver scholarly and intuitive applications to infer productive outcomes [1]. The wide accessibility of internet based life destinations gives people to impart their opinions or insights about a specific occasion, item or issue. Mining of such casual and homogeneous information is exceptionally valuable to make inferences in different fields. However, the profoundly unstructured organization of the assessment information accessible on web makes the mining procedure testing [2].

There are four most generally talked about points in Twitter, for example, wellbeing, music, game, and innovation. In help of this, Honigman brought up that there are at any rate 40% of netizens around the world, who dilligently get to the heath data in the online life, are hugely impacted by the data that further influences the manner in which they manage their social insurance [2]. Music, on the hand, is the third of the best ten most every now and again

### Revised Manuscript Received on May 23, 2019.

Ashwini Gedekar, M.tech Scholar, Affiliated to Nagpur University, Ramdeobaba college of engineering & management, Nagpur, India.

Ashwini Zadgaonkar, Assistant professor, Affiliated to Nagpur University, Ramdeobaba college of engineering & management, Nagpur, India.

examined points in Twitter dependent on the examination done by Franklin and his group

[3]. Game is broadly talked about due to the substantial number of game darlings for various types of games the world over [6]. In the mean time, innovation is incredibly examined in any online life like Facebook, Twitter and Instagram to get the most recent data on innovation created at present. Innovation assumes a critical job on each part of our lives [3]. A hash tag is ordinarily used to distinguishing the theme of the

tweet. Having said that, numerous organizations can assemble tweets dependent on the hash tags when they are looking for reactions of netizens relevant to their items and administrations. Governments and other intrigue gatherings can utilize the gathered tweets to distinguish human reasoning examples, assemble recognizable proof and proposal, and furthermore feeling about a particular themes of interests [4]. All in all, imagine a scenario where the tweet has no hash tag. They at that point need to peruse the entire tweet to know the substance. It is most likely tedious and furthermore requires a gigantic measure of endeavors to comprehend the principle substance of the tweet.

Literary data present on web is significantly arranged into both of the two classifications: truth information and feeling information [5]. Certainty information are the target phrasings concerning distinctive substances, issues or occasions. While notion information are the abstract terms, that characterize person's assessments or convictions for a specific element, item or occasion. Assumption examination is the way toward perceiving and grouping diverse slants passed on online by the people to determine the author's methodology towards a particular item, point or occasion is certain, contrary or unbiased. Opinion investigation has three noteworthy segment of concentrate as pursues: conclusion holder for example subject, assumption itself for example conviction and article for example the point about which the subject has shared the assumption. An article is a substance that speaks to a positive individual, thing, item, issue, occasion, theme or any association [5]. Conclusion examination is done at various dimensions going from coarse dimension to fine dimension. The coarse dimension conclusion investigation decides the supposition of the entire original copy or report. The fine dimension estimation examination, though centers around the properties.



Notion examination of Twitter information is done on sentence level which comes in the middle of coarse dimension and fine dimension. In the slant investigation process, the notions present in the content are of two kinds: Direct and Comparative. The immediate feelings in content are autonomous from different items in a similar sentence [6].

In this paper, we compare techniques for processing which include but are not limited to sentiment analysis, summarization, event detection, prediction of events and classification of tweets based on user's online tweet patterns. The next section describes these techniques in details, followed by some sample results, finally we conclude the paper with some acute observations about these techniques and ways to improve them in order to build a more efficient system.

## II. LITERATURE REVIEW

The essential utilization of tweets is for investigating slants, and as per Thelwall, Buckley, Paltoglou, Cai, and Kappas [7] SentiStrength is an estimation examination independent instrument for short and even casual dialect. It has human dimension exactness for short English content. Senti Strength scales from 1 to 5 for both positive (+1 for powerless positive to +5 for outrageous positive) and negative (- 1 for feeble negative to - 5 for extraordinary negative) slant. Senti Strength exactness level is 60.6% for positive feeling and 72.8% for negative feelings. Expression ID was for just constrained words and furthermore uncertain component decided wrong extremity. Patricia L V Ribeiro, Li Weigang and Tiancheng Li [8] manufactured the supposition digging device for Twitter. This apparatus incorporates extended hashtag calculation and spam identification calculation. Hashtag calculation is utilized to gather numerous tweets and for better outcome spam identification calculation is utilized. The spam discovery calculation breaks down spam tweets and dispenses with every single loud tweet. It likewise utilizes diagram developer calculation which makes charts of significant relations with hubs. Diagram spread calculation is utilized to follow chart and Lexicon Building calculation which handles the nullification and lengthening highlights.

Linus Philip Lawrence [9] has looked at two assessment devices, Semantria and Social Mention. Semantria is a record level estimation examination instrument. Social Mention can gather data from in excess of 80 interpersonal interaction destinations. Investigation was done dependent on 12 diverse vehicle models and measurements. Semantria utilizes the calculation which recognizes feeling expressions and after that scales assessment esteems from +10.0 to - 10.0 and later scales from +2.0 to - 2.0 to decide extremity. Akshat Bakliwal, et.al [10], decided score by considering assessment bearing expressions by utilizing the corpus. The creators utilized Stanford Dataset and accomplished exactness of 87% and utilizing Mejj dataset accomplished 88%. The

creators likewise utilized directed machine learning strategy with Stanford Dataset and accomplished 88% precision. Tiara, Mira Kania Sabariah, Veronikha Effendy [11] created assessment instrument dependent on TV programs. It decided the supposition of every TV program, utilizing a word reference. The creators utilized the Lexicon based way to deal with assess the estimation score for the expressions including nullification taking care of. Xujuan Zhou, Xiaohui Tao, Jianming Yong, Zhenyu Yang [12], examine about the apparatus named TSAM (Tweets Sentiment Analysis Model). TSAM decides the feeling on a specific occasion, which has estimation bearing words. The creators have utilized stemming, POS labeling, pack of words systems to investigate the assessment. Earlier extremity is determined and utilizing most expressive words and the valence shifter vocabulary substances will move to either positive or negative including refutation.

A significant part of the exploration in unsupervised assessment order utilizing representative strategies makes utilization of accessible lexical assets. Turney [13] utilized pack of-words approach for opinion investigation. In that approach, connections between the individual words are not considered and a record is spoken to as a negligible gathering of words. To decide the general conclusion, opinions of each word is resolved and those qualities are joined with some conglomeration capacities. He found the extremity of a survey dependent on the normal semantic introduction of tuples extricated from the audit where tuples are phrases having modifiers or verb modifiers. He found the semantic introduction of tuples utilizing the web crawler Altavista. Kamps et al. [14] utilized the lexical database WordNet [15] to decide the enthusiastic substance of a word along various measurements. They built up a separation metric on WordNet and decided the semantic introduction of descriptive words. WordNet database comprises of words associated by equivalent word relations. Baroni et al. [16] built up a framework utilizing word space show formalism that beats the trouble in lexical substitution undertaking. It speaks to the neighborhood setting of a word alongside its general dissemination. Balahur et al. [17] presented EmotiNet, an applied portrayal of content that stores the structure and the semantics of genuine occasions for a particular space. Emotinet utilized the idea of Finite State Automata to recognize the passionate reactions activated by activities. One of the member of SemEval 2007 Task No. 14 [18] utilized coarse grained and fine grained ways to deal with recognize assumptions in news features. In coarse grained approach, they performed twofold arrangement of feelings and in fine grained methodology they grouped feelings into various dimensions. Learning base methodology is observed to be troublesome because of the necessity of an immense lexical database. Since informal community produces colossal measure of information consistently, now and then bigger than the extent of accessible lexical database, opinion examination ended



up monotonous and mistaken.

Machine Learning methods utilize a preparation set and a test set for arrangement. Preparing set contains input highlight vectors and their comparing class names. Utilizing this preparation set, a grouping model is created which attempts to arrange the info include

vectors into relating class marks. At that point a test set is utilized to approve the model by foreseeing the class names of concealed element vectors. Various machine learning methods like Naive Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM) are utilized to group audits [19]. A portion of the highlights that can be utilized for assessment order are Term Presence, Term Frequency, nullification, n-grams and Part-of-Speech [20]. These highlights can be utilized to discover the semantic introduction of words, expressions, sentences and that of records. Semantic introduction is the extremity which might be either positive or negative.

Domingos et al. [21] found that Naive Bayes functions admirably for specific issues with exceptionally subordinate highlights. This is astounding as the fundamental suspicion of Naive Bayes is that the highlights are free. Zhen Niu et al. [22] presented another model in which effective methodologies are utilized for highlight choice, weight calculation and grouping. The new model depends on Bayesian calculation. Here loads of the classifier are balanced by making utilization of agent highlight and one of a kind element. 'Agent highlight' is the data that speaks to a class and 'Novel element' is the data that helps in recognizing classes. Utilizing those loads, they determined the likelihood of every arrangement and in this way enhanced the Bayesian calculation.

Numerous works have been given to break down what makes web substance well known. The overview [23] demonstrates that the equation is as yet obscure yet a few highlights have been recognized, which in the Twitter case is by and large related to being retweeted. Specifically, the creators show that media content is the component all the more reliably found in prevalent tweets. Other than media content, the paper records as significant: (I) the maker is a notable identity in the field,

(ii) the content alludes to hot or dubious points, (iii) the content alludes to well known areas or occasions, (iv) it triggers feelings, and (v) it cooperates with different systems (for example a video prominent in Twitter will expand YouTube visits also). We can rough a portion of these highlights from our dataset; for instance, for (I) we will think about that confirmed records relate to wellknown identities and we limit (iii) by concentrate a specific occasion. Notwithstanding, whatever remains of highlights would require an examination that removes us from our dialect free objective. Something comparable occurs in [4], where the creators anticipate whether a tweet will be retweeted joining

data specifically accessible from the tweet (number of adherents, tweet length, hashtags, and notices) and assumption investigation. In [24] the creators explore the issue of foreseeing the notoriety of messages estimated by the quantity of future retweets. They perform characterization on various gatherings relying upon the quantity of retweets a tweet creates and utilize a few highlights like the content of the tweet, the diagram auxiliary properties of clients, the worldly elements of retweet chains, and other meta-data of clients. A vital research line toward this path is depicted in [25]. There, the creators present how to utilize neural systems, refining the work for content grouping in [22], [23], to perform notion examination of tweets autonomously of the dialect. Specifically, they utilize neural systems to derive the assessment of tweets in a dialect free manner, working at character level and supporting four unique dialects. A comparative methodology is followed in [24] for deriving semantic data from writings utilizing character-level convolutions. It is intriguing to join this methodology with our own, so we would have the capacity to utilize the assessment in our models. Be that as it may, this joining would confront a few troubles because of the fundamental advances in light of the fact that there is no official execution of convolutional neural systems in Spark. At long last, there are increasingly broad ways to deal with foresee the development of ubiquity in client produced content like the one in [25], tried on recordings (YouTube and Vimeo) and news (Digg). The creators classify the conduct of substance after some time to recognize examples of ubiquity development, and afterward foresee the future prevalence dependent on the genuine notoriety development and the distinguished examples. The main highlights utilized are the quantity of hits after some time and its change. Subsequently, it is a substance skeptic approach that can be connected to various applications, however it depends on the past prominence development and can't be utilized to recognize what makes a tweet effective.

Another zone of research identified with twitter information mining is occasion distinguishing proof. In [26], the creators proposed a general system for occasion distinguishing proof in internet based life reports. They utilized similitude metric learning ways to deal with create excellent bunching results. They announced that similitude metric learning methods yield preferred execution over conventional methodologies that considers content based comparability. Sakaki et al. [27] proposed a strategy to distinguish constant occasions. They proposed a calculation to recognize target occasions progressively and considered tweets-related highlights like catchphrases, number of words and their setting for distinguishing the objective occasions. The fundamental focal point of their investigation is to recognize quake occasion. In [28], the creators built up a framework to group tweets into genuine occasion tweets and non-occasion tweets.

Haddi et al. [29] investigated the job of



content pre-handling in motion picture audits assessment examination. The exploratory outcomes demonstrate that the precision of notion order might be altogether enhanced utilizing suitable highlights and portrayal after pre-handling. Saif et al. [24] contemplated the impact of various stop words evacuation strategies for extremity order of tweets and in the case of expelling stop words influences the execution of Twitter estimation classifiers. They connected six distinctive stop words recognizable proof techniques to six diverse Twitter datasets and saw how evacuating stop words influences two directed supposition characterization strategies. They evaluated the effect of evacuating stop words by watching variances on the dimension of information sparsity, the extent of the classifier's component space and its grouping

Execution. Utilizing pre-arranged arrangements of stop words contrarily affected the execution of Twitter feeling order approaches. Safi et al. [25] found taper-handling prompted a noteworthy decrease of the first element space. After pre-preparing, the vocabulary measure was decreased by 62%. Be that as it may, they didn't talk about the impact on the execution of Twitter opinion classifiers. Boa et al. [30] Investigated the impact fore-handling techniques on Twitter supposition order. They assessed the impacts of URLs, nullification, rehashed letters, stemming and lemmatization. Trial results on the Stanford Twitter Sentiment Dataset demonstrate that assessment order exactness increments when URL highlights reservation, nullification change and rehashed letters standardization are utilized, however diminishes when stemming and lemmatization are connected. Zhao [31] assessed the precision of URLs, stop words, rehashed letters, invalidation, abbreviation and numbers in the twofold Twitter slant arrangement errand. The analyses demonstrate that the exactness of estimation arrangement ascends in the wake of growing abbreviation and supplanting nullification, albeit scarcely change when expulsion URL, evacuation numbers and evacuation stop words are connected. A great deal of research has been done in the field of spam recognition. The creators have utilized the idea of social honeypots in [31], alongside machine learning for spam identification in OSNs. Social honeypots are phony profiles or records which are made intentionally to pick up the consideration of a spammer. The technique for distinguishing pharmaceutical spam in Twitter is talked about in [22]. This is finished by applying content mining systems and information mining devices. This paper is tending to how to arrange another approaching spam as pharmaceutical spam or not. The creators utilized choice tree (J48) calculation and Naïve-Bayes calculation. At long last they thought about they yield acquired by both these classifiers. A lot of 65 words (which were identified with pharmaceuticals) were utilized as the preparation set. In the event that somewhere around, single word out of these 65 were available in the tweets in the test set, at that point they will be named spam. Online Spam Filtering is displayed in [23]. This is a continuous framework. This can investigate a message (tweet on account of Twitter and post on account of

Facebook) and drop it in the event that it is observed to be a spam. The spam messages are dropped even before the expected beneficiary gets it. Everything occurs continuously. Such messages are not put away in the database. The paper utilizes machine learning methods. A great many tweets and posts are gathered from both Twitter and Facebook for datasets. In this paper the creators have utilized two regulated machine learning calculations to be specific Support Vector Machine (SVM) and Decision Tree. Assessment of the setting mindful spam that could result from data that is shared on the informal organizations is managed in [24]. The relief methods are likewise talked about here. The writers have done investigation on Facebook. The creators inferred that setting mindful email assaults have a high rate of achievement. The paper additionally specifies the protection systems taken by other interpersonal organizations like LinkedIn and MySpace. Gathered Twitter dataset and joins are analyzed in [26]. Here the creators have discovered highlights utilizing which content polluters can be effectively distinguished. The creators proposed a long haul investigation of ensuring informal communities utilizing honeypots. Right around 60 honeypots were sent for seven months which brought about the reaping of in excess of 30000 spam information. The spam order was finished utilizing machine learning calculations. Identify spam bots in Online Social Networks particularly in Twitter is the reason for [27]. Spammers use Twitter to post different copy refreshes. In this paper, suspicious conduct of spam bots are contemplated. They utilized order strategies like choice tree, neural systems, Support Vector Machine (SVM), k-closest neighbor and Naïve-Bayes calculation. The creators physically named 500 records as spam and non-spam for the preparation set. Everyone of the calculations utilized were contrasted and one another and in this way Naïve-Bayes was observed to be the best. The definition and recognizable proof of hazard factors and cautioning signs lie at the center of suicide counteractive action endeavors. In this paper, we have referenced the hazard factors characterized by the American Psychiatric Association (APA) [13] and the notice signs distinguished by the American Association of Suicidology (AAS) [14]. These assets speak to a dimension of accord between psychological wellness experts and furthermore give a rich exchange of the contrasts between suicide chance factors and cautioning signs. For further understanding, we guide the peruser to crafted by [14]. As featured by [14], cautioning signs mean expanded unavoidable hazard for suicide (i.e., inside minutes, hours, or days). As indicated by the APA, suicide cautioning signs may incorporate looking at biting the dust, noteworthy ongoing misfortune (demise, separation, detachment, or broken relationship), change in identity, dread of losing control, suicide plan, self-destructive contemplations, or no expectation for what's to come. As talked about in the accompanying sections, late research has given the development of such suggestions on long range informal communication locales. The greater part of the examination at the crossing point of conduct



wellbeing issue and web based life has concentrated on sadness location in online networks, explicitly Major Depressive Episodes (MDE). Nonetheless, the hazard factors for suicide characterized by the APA [13] go a long way past melancholy alone. Remember that misery does not really infer self-destructive ideation. Or maybe, suicide ought to be thought of as a potential end indication of sadness. While psychological well-being issues, for example, discouragement, self-destructive ideation, and self-mutilation are characterized restoratively as isolated ailments with covering side effects, the methodologies proposed to identify them online can be very comparative. The methodologies fluctuate in the information they are treating, i.e., Facebook posts, Twitter tweets, Reddit discussion strings, and so forth., and the particular occasion they are endeavoring to foresee. Moreno et al. [27] first exhibited that interpersonal interaction destinations could be a potential road for recognizing understudies experiencing despondency.

The commonness rates found for sadness uncovered on Facebook related to past works in which such data was self-revealed. On a bigger scale, Jashinsky et al. [15] indicated relationship between's Twitter-determined and genuine United States per-state suicide information. Together, these works set up the nearness of melancholy exposure in online networks and opened up another road for emotional wellness look into. De Choudhury et al. [6] investigated the possibility to utilize web based life to recognize and anticipate real burdensome scenes in Twitter clients. Utilizing publicly supporting strategies, the creators fabricated an accomplice of Twitter clients scoring high for misery on the CES-D (Center for Epidemiologic Studies Depression Scale) scale and for different clients scoring low. Concentrate these two classes, they found that what is known from conventional writing on burdensome conduct additionally means online networking. For instance, clients with a high CES-D score posted all the more every now and again late around evening time, collaborated less with their online companions, and had a higher utilization of first-individual pronouns. Also, online etymological examples coordinate past discoveries with respect to dialect utilization of discouraged people [16]. All the more as of late, De Choudhury et al. [10] have appeared phonetic highlights are critical indicators in recognizing people changing from mental talk via web-based networking media to self-destructive ideation. The creators demonstrated various markers describing these movements, including social commitment, appearance of misery, tension, and lack of caution dependent on a little subset of Reddit posts. Coppersmith et al. [17] analyzed the information distributed by Twitter clients before a suicide endeavor and gave an experimental investigation of the dialect and feelings communicated around their endeavor. One of the intriguing outcomes found in this investigation is the expansion in the level of tweets communicating misery in the weeks before a suicide endeavor, which is then trailed by a discernible increment in indignation and pity feelings the week following a suicide endeavor. In a similar line of research,

O'Dea et al.

[18] affirmed that Twitter is utilized by people to express suicidality and exhibited that it is conceivable to recognize the dimension of worry among suicide-related tweets, utilizing both human coders and a programmed machine classifier. These bits of knowledge have likewise been examined by Braithwaite et al. [19], who showed that machine learning calculations are productive in separating individuals who are at a self-destructive hazard from the individuals who are most certainly not. For a progressively itemized audit of the utilization of internet based life stages as an instrument for suicide aversion, the peruser may allude to the ongoing precise overview by Robinson et al. [20].

Related with highlights extraction from tweets, a portion of the specialists have concentrated on what sorts of highlights are utilized. Castillo et al. used four sorts of highlights, for example, 21 message-based, 7 client based, 35 theme based, and 5 proliferation based to make a classifier for assessing the believability of tweets [8]. In their exploration, they concentrated on the dimension of validity of each pattern on

Twitter. Gupta et al. utilized six sorts of highlights, for example, Tweet Meta-information Features, Tweet Content Features, User based Features, Network Features, Linguistic Features and External Resource Features for believability investigation [31]. Moreover, hashtags have been viably used as basic highlights for different assignments of content or internet based life examination, including tweet grouping [3]. In contrast to them, this framework centers around the substance based highlights, for example, Linguistic Features, NRC Sentiment hashtags dictionaries based Features and Disaster Lexicon based highlights as it were. Despite the fact that, the quick development of internet based life, it keeps on staying on the versatility issues of validity expectation or tweet characterization. In this way, another region of related research is checking and ordering for educational messages in microblog stages. The characterization of tweets as Credible or Not Credible is displayed in [8]. In addition, the data discovery and extraction framework for microblog posts was portrayed by Imran et al. [22]. In their work, Naïve Bayesian classifiers were utilized to group a tweet into one of the sorts, for example, Caution and Advice, Informative source, Donation, and Causalities and harm. Gupta et al. likewise gave a SVM-rank based framework, TweetCred to allot a validity score to tweets in a client's course of events [8]. As per the literary works, administered machine learning calculations have been connected by a large portion of the analysts to identify and arrange the substance in OSM. Innocent Bayes (NB) and Support Vector Machine (SVM) are utilized for tweets order in [28].

#### Final Stage

When you submit your final version, after your paper has been accepted, prepare it in two-column format, including figures and tables.



III. PROPOSED SYSTEM

We are developing a system based on event detection for calamities in real time. For this we are following the given steps,

- Data collection from real time tweet handles like BarackObama, NarendraModi and others, who tweet about happening of calamities
- Tweet classification process will classify the tweets into one of the given calamities
- Sentiment analysis module will be used to identify if the classified tweet is a positive or negative one, if the tweet is positive then there is a chance of calamity, if it is negative then there is little chance of the given calamity
- Post sentiment analysis, all these modules will be combined and aggregation engine will aggregate the tweets in order to obtain the final results about the given calamities

This system will be used to detect natural calamities and help in informing various departments about the disasters which might be happening around the world.

IV. ANALYSIS OF TECHNIQUES

Various techniques use various datasets for tweet analysis, some of these datasets are

- The data available directly from twitter is usually in four different types of format. They include HTML, XML, JSON and CSV
- Twitter API to obtain tweets containing specific keywords
- Some systems use people freely available annotated datasets in an year range, occurring in different geographies which affected a large set of users

The accuracy results of various techniques is given in the following table.

Technique	Application	Accuracy (%)
Neural words embedding	Natural disaster prediction	78
Bag of words	Natural disaster prediction	90
Feature selection	Natural disaster prediction	92
Sequential minimal optimization	Suicide prediction	66
Random forest	Suicide prediction	57
Simple logistic	Suicide prediction	83
Naïve bayes	Spam detection	94
Support vector machines	Spam detection	92
Linear regression	Sentiment analysis	94

Naive bayes	Sentiment analysis	93
Random forest	Sentiment analysis	96
Support vector machine	Sentiment analysis	93

I. Comparison of techniques based on applications

From the above table we can identify that Random forest, and support vector machine based techniques are better suited for tweet analysis as they give better accuracy across different applications. These techniques are generally used by various researchers in order to develop an optimum tweet processing system.

V. CONCLUSIONS

Tweet analysis can be used in multiple domains and random forest based techniques are generally identified to produce high level of accuracy for the system. Apart from the already mentioned techniques, researchers can apply deep learning based artificial intelligence techniques in order to improve the overall accuracy of the system. The tweet analysis systems can further be used for detecting global emergencies like volcanic eruptions, earthquakes and other natural calamities, so that humans can be used as sensors and develop a better and more optimum system for detection and prediction of events on a large scale.

per, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

REFERENCES

1. Z. Kechaou, B. M. Ammar and A. M. Alimi, "Improving e-learning with sentiment analysis of users' opinions", in Global Engineering Education Conference (EDUCON), 2011, pp. 1032-1038.
2. Weil, Kevin (VP of Product for Revenue and former Big Data engineer. Twitter Inc.), "Measuring Tweets." Twitter Official Blog, February 22, 2010. [Online]. Available: <http://www.internetlivestats.com/twitterstatistics>. [Accessed:19-Oct-2015].
3. M. Dredze , "How Social Media Will Change Public Health", IEEE Intelligent Systems, 2012, pp. 1541-1672.
4. A. Celikyilmaz, D. Hakkani-Tur and Junlan Feng, "Probabilistic modelbased sentiment analysis of twitter messages", IEEE Spoken Language Technology Workshop (SLT), 2010, pp. 79-84.
5. S. Bhuta, A. Doshi, U. Doshi and M. Narvekar, "A review of techniques for sentiment analysis Of Twitter data", Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014, pp. 583-591.
6. W. Medhat, A. Hassan and H. Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, vol. 5, no. 4, 2014, pp. 1093-1113.
7. Krikorian, Raffi (VP, Platform Engineering, Twitter Inc.), "New Tweets per second record, and how!" Twitter Official Blog. August 16, 2013.[Online]. Available: <https://blog.twitter.com/2013/new-tweetsper-second-record-and-how>. [Accessed: 19-Oct-2015].
8. "Twitter turns six." Twitter Official Blog. March 21, 2012. [Online]. Available: <https://blog.twitter.com/2012/twitter-turns-six>. [Accessed:19-Oct-2015].
9. "Three Cool and Inexpensive Tools to Track Twitter Hashtags", June 11, 2013. [Online]. Available



<http://dannybrown.me/2013/06/11/threecool-toolstwitterhashtags/>  
[Accessed: 19-Oct-2015].

10. X. Chen, M. Vorvoreanu and K. Madhavan, "Mining Social Media Data to Understand Students' Learning Experiences", IEEE Transaction, 2014, vol. 7, no. 3, pp. 246-259.
11. B. Gokulakrishnan, P. Plavnathan, R. Thiruchittampalam, A. Perera and N. Prasath, "Opinion Mining and Sentiment Analysis on aTwitter Data Stream", in Int. Conf. on Advances in ICT for Engineering Regions, 2012, pp. 182-188.
12. Z. WANG, V. J. Chuan TONG, X. XIN and H. C. CHIN, "Anomaly Detection through Enhanced Sentiment Analysis on Social Media Data", in 6th International Conference on Cloud Computing Technology and Science, 2014, pp. 918-922.
13. A.C.E.S Lima. and L.N. de Castro, "Automatic sentiment analysis of Twitter messages", in 4 th Int. Conf. on Computational Aspects of Social Networks (CASoN), 2012, pp. 52-57.
14. K. Ghag and K. Shah, "Comparative analysis of the techniques for Sentiment Analysis", in Int. Conf. on Advances in Technology and Engineering, 2013, pp. 1-7.
15. V. Singh and S. K. Dubey, "Opinion mining and analysis: A literature review", in 5th Int. Conf. on Confluence The Next Generation Information Technology Summit (Confluence), 2014, pp. 232-239.
16. Bing Liu, N. Indurkha and F. J. Damerau, Handbook of Natural Language Processing, Second Edition, 2010, pp. 1-3860-68.
17. M. S. Neethu and R. Rajasree, "Sentiment Analysis in Twitter using Machine Learning Techniques", in 4th Int. Conf. of Computing, Communications and Networking Technologies (ICCCNT), 2013, pp. 1-5.
19. P. Saloun, M. Hruzik and I. Zelinka, "Sentiment Analysis – e-Bussines and e-Learning Common Issue", in 11th IEEE Int. Conf. on Emerging eLearning Technologies and Applications, 2013, pp. 339-34.
20. J. Khairnar and M. Kinikar, "Machine Learning Algorithms for Opinion Mining and Sentiment Classification", in International Journal of Scientific and Research Publications, vol. 3, no. 6, June 2013.
21. K. Khan, B. Baharudin, A. Khan and F. Malik, "Mining Opinion from Text Documents: A Survey", Digital Ecosystems and Technologies, 2009, pp. 217-222.
22. Twitter Engineering, "200 million Tweets per day." Twitter Official Blog. June 30, 2011. [Online]. Available: <https://blog.twitter.com/2011/200-million-tweets-per-day>. [Accessed: 19-Oct-2015].
23. I. King, J. Li and K. T. Chan, "A Brief Survey of Computational Approaches in Social Computing", in Proc. of Int. Joint Conf. on Neural Network, 2009, pp. 2699-2706.
24. S. R. Barahate and V. M. Shelake, "A Survey and Future Vision of Data mining in Educational Field", in Proc. 2nd Int. Conf. on Advanced Computing & Communication Technology, 2012, pp. 96-100.
25. R. Batool, A. M. Khattak, J. Maqbool and S. Lee, "Precise tweet classification and sentiment analysis", in 12th Int. Conf. on Computer and Information Science (ICIS), 2013, pp. 461-466.
26. C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," in Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, 2010, vol. 40, no.6, pp. 601-618.
27. N. Kasture and P. Bhilare, "An Approach for Sentiment analysis on social networking sites", Computing Communication Control and Automation (ICCUBE), 2015, pp. 390-395.
28. A. Sarlan, C. Nadam and S. Basri, "Twitter Sentiment Analysis", in Int. Conf. on Information Technology and Multimedia, 2014, pp. 213-216.
29. G. Gautam and D. Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis", in 7 th Int. Conf. on Contemporary Computing, 2014, pp. 437-442.
30. G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education", Educause Review, 2011, vol. 46, no. 5, pp. 30-32.
31. V. Sehgal and C. Song, "SOPS: Stock Prediction Using Web Sentiment", in 7th IEEE Int. Conf. on Data Mining Workshop, 2007, pp. 21-26.
32. N. Altrabsheh, M. Cocea and S. Fallahkhar, "Sentiment analysis: towards a tool for analysing real-time students feedback", in 26th International Conference on Tools with Artificial Intelligence, 2014, pp. 420-423.
33. S. Bahrainian and A. Dangel, "Sentiment Analysis using Sentiment Features", in Int. joint Conf. of Web Intelligence and Intelligent Agent Technologies, 2013, pp. 26-29.



**Ashwini Zadgaonkar** received bachelor of Engineering Degree in Computer Technology from PCE , Nagpur university and Masters of Technology in Computer science and Engineering from RCOEM ,Nagpur university, India in 2012 and is currently working as an assistant professor in Computer science and engineering department of Shri Ramdeobaba College of Engineering and Management, Nagpur.

## AUTHORS PROFILE



**Ashwini Gedekar** received bachelor of engineering degree in information technology from KITS Ramtek, Nagpur university and currently pursuing M.tech from shri ramdeobaba college of engineering and management.

