International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6313
***Article Received**: 24 December 2022 **Revised**: 28 January 2023 **Accepted**: 08 February 2023*

Detecting the Anti-Social Activity on Twitter using EGBDT with BCM

Narender Chinthamu[1*], Satheesh Kumar Gooda[2], Dr. P. Shenbagavalli[3], Dr. N. Krishnamoorthy[4], Dr. S.Tamil Selvan[5]

[1] Enterprise Architect,
MIT CTO Candidate
e-mail: Narender.chinthamu@gmail.com

[2] Senior Manager, WESCO International,
USA, Student of Osmania University,
Hyderabad, India.
e-mail: skgooda@gmail.com

[3] Assistant professor, Department of Digital Sciences,
Karunya Institute of Technology and Sciences,
Coimbatore-641114, India
e-mail: shenbagavalli@karunya.edu

[4] Associate Professor Grade 2, School of Information Technology and Engineering,
Department of Software Systems and Engineering,
Vellore Institute of Technology, Vellore Campus,
Thiruvalam Road, Katpadi, Vellore, Tamil Nadu, India
Email: krishnamoorthyer@gmail.com

[5] Assistant Professor, Department of Computer Science and Engineering,
Erode Sengunthar Engineering College, Thudupathi,
Perundurai - 638 057, India
Email: Stamilselvancse@gmail.com

Abstract— The rise of social media and its consequences is a hot topic on research platforms. Twitter has drawn the attention of the research community in recent years due to various qualities it possesses. They include Twitter's open nature, which, unlike other platforms, allows visitors to see posts posted by Twitter users without having to register. In twitter the sentiment analysis of tweets are used for detecting the anti-social activity event which is one of the challenging tasks in existing works. There are many classification algorithms are used to detect the anti-social activities but they obtains less accuracy. The EGBDT (Enhanced Gradient-Boosted Decision Tree) is used to optimize the best features from the NSD dataset and it is given as input to BCM (Bayesian Certainty Method) for detecting the anti-social activities. In this work, tweets from NSD dataset are used for analyzing the sentiment polarity i.e. positive or negative. The efficiency of the proposed work is compared with SVM, KNN and C4.5. From this analysis the proposed EGBDT and BCM obtained better results than other techniques.

Keywords- Twitter, Anti-social activity, Social media, Gradient-Boosted Decision Tree, Bayesian Certainty Method

I. INTRODUCTION

Knowledge sharing and availability has been accelerated and simplified by social media. In disaster situations, social media is increasingly being used to raise awareness [1]. Twitter has drawn the attention of the research community in recent years due to various qualities it possesses. The open nature of Twitter, for example, enables visitors to see posts produced by Twitter users without registering. In addition, Twitter users don't always have mutually beneficial relationships. The process of sentiment analysis involves categorizing the opinions expressed about a specific object. Technology has helped us gain a better understanding of the general public's opinions about businesses, products, and general preferences. Understanding the sentiment of social media posts can provide insight into how users should react and proceed. In spite of this, the content of tweets (microblogs produced or published by Twitter users) is likely what makes this microblogging network so intriguing. 140 characters is the maximum length of a tweet, making them concise and easy to read. There are many slang terms, acronyms, and emoticons used in this situation. As well as distributing news and conversations, hashtags (words or phrases followed by the symbol '#') can also be used for forming micro-celebrities and sub-communities.

Although hashtags are not moderated and unrestricted, they are widely used, particularly on Twitter. The sentiment on Twitter can be analyzed using several methods [2]. One of the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6313
***Article Received**: 24 December 2022 **Revised**: 28 January 2023 **Accepted**: 08 February 2023*

most effective methods for detecting current topics and obtaining tweets about specific topics, products, services, and more has been utilizing these techniques. The purpose of sentiment analysis is to classify opinions expressed about a specific object. With the development of various technologies, it has become a vital measure to be aware of the general public's opinion on business, products, or general likes and dislikes. Monitoring the sentiment behind social media posts can provide the context in which the user should react and develop [3]. The following anti-social activities are illegal: online threats, stalking, and cyberbullying, Fraud and hacking, Purchasing Illegal Items, Uploading Criminal Activity Videos, and Vacation robberies. In this context, use Machine Learning (ML) to analyse Twitter sentiment in order to detect anti-social behaviour. A dataset of 5453 tweets was used to identify cyberbullying based on user characteristics such as attitude and emotion [4].

The detection of cyberbullying was significantly improved when user personalities and attitudes were incorporated into J48. It was found that there were ten major elements, which were then combined into a single model. Online bullying patterns were found to be less likely to be spotted when emotions were present. In terms of both size and category (i.e. role), the dataset is constrained. The preprocessing produced an unbalanced dataset of 5453 tweets, the majority of which were normal and spams. In [5,] there are primarily two categories of ML approaches that are commonly employed in sentiment analysis: supervised and unsupervised learning. During the analysis phase of supervised learning, the dataset is labelled and trained to produce a suitable output that aids decision making. Unsupervised learning, unlike supervised learning, does not supply any label data, making the process exceedingly tough. Towards the 2013 elections, an attempt was made by [6,] to measure the popularity of Pakistani political parties based on keyword-driven tweets. Machine learning techniques were used to analyze this dataset. The classification methods used were Prind, K Nearest Neighbors (KNN), and Naive Bayes based on unigram data. Based on the same dataset, four supervised classification methods were tested: Support Vector Machines (SVM), Naive Bayes Multinomial (NB), Random Forest (RF), and Naive Bayes Multinomial (NBMN). In twitter analysis, they are not focused on hot topics such as anti-social behaviour. Twitter categorization is used to target anti-social behaviour.

This study makes use of the Twitter dataset to identify anti-social behaviour using enhanced classification approaches. Contribution of the proposed work, in the first stage, an NSD data set including (n) features (i.e. tweets) is employed in this study to assure improper activity. Following that, the dataset is filtered, the relevant data is extracted, and the extraneous data from the data preparation is eliminated. The normalization approach is then applied to scale feature data into a fit. The feature selection approach (EGBDT) is given the normalized data fields as input. Finally, using the Improved GBDT algorithm, the input features are optimized. The unimportant feature is eliminated, and the optimal features are extracted using feature weighting and correlation-based feature selection (CFS) (FW). Finally, sentimental analysis employing a class based on CFS & FW in EGBDT is used to identify the anti-social behaviours. By employing the Bayesian Certainty Method (BCM), the remaining features are provided as input to make sure that a certainty factor is utilised to classify the data. The flow diagram of the proposed model was shown in the following figure 1.

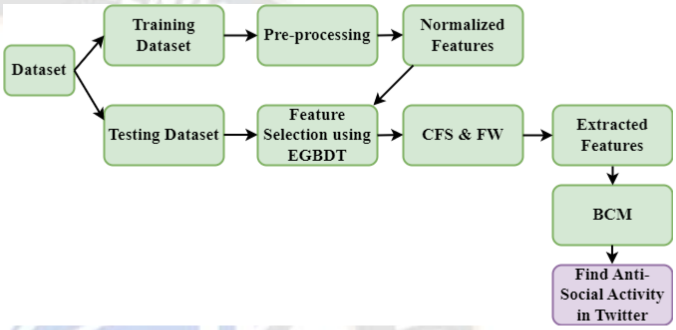

Figure 1. Flow diagram of the proposed work

II. LITERATURE SURVEY

Rogstadius et al. [7] introduced a Crisis Tracker, a web-based application that mixes automated real-time analysis with crowd sourcing, to annotate fast streams of unstructured tweets with metadata and aggregate relevant information, with the purpose of boosting situational awareness during crisis. They also demonstrated how combining crowd endurance with automated data collecting and language-independent real-time text clustering may improve robustness, accuracy, timeliness, and efficiency.

Using the literature on protest participation theory, Wu et al. [8] constructed a predictor variable for individual protest decisions. We then put these factors to the test using Twitter and the Egyptian revolution of 2011. They discover statistically major positive associations among the number of upcoming-protest explanations on Twitter and the start of protests. Then, using future-protest descriptions, construct prediction models and compare them to benchmarks generated by daily activity estimates from the Global Database of Events, Location, and Tone (GDELT).

Kanjo et al. [9] investigated how ordinary people engage with alerts and their influence on users' emotions. 50 operators were appointed to install and utilize their program NotiMind for five weeks. Participants' phones recorded hundreds of social and system alerts, as well as affect data through self-reported Positive and Negative Disturb Agenda assessments 3

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6313
Article Received: 24 December 2022 Revised: 28 January 2023 Accepted: 08 February 2023

times a day. These findings demonstrate that it is feasible to anticipate when people are in emotional states based on interactions with alerts.

Roy et al. [10] presented Attention Systems, a statistical bounded automaton that can measure the collective attention of a user group. Clubs are created on Twitter depending on individuals' geographical closeness or common preferences. They find two major elements that initiative combined user attention: (1) the community's attention instability (frequency of change of current subjects), and (2) the user group's choosy category affinity towards particular trends.

Subramani et al. [11] suggested a method for multi-class categorization from DV social media postings using cutting-edge Deep Learning models for DVCS group assistance. Domestic Violence Crisis Support (DVCS) institutions working on social networks have become essential in recent years in providing assistance to sufferers and their caregivers. The primary problem for DVCS organizations is to manually discover the catastrophic condition in a timely manner in the midst of the deluge of digitally created material.

Subramani et al. [12] used Deep Learning as a method for automatically identifying DV sufferers in acute necessity. Actual proof on a ground truth records set reached up to 94% accuracy, outperforming typical ML approaches. The investigation of useful properties aids in the identification of key words that may signal critical steps in the categorization process. The experimental findings will aid researchers and practitioners in the development of approaches for identifying and helping DV victims.

Son et al. [13] focused on the most current advances in research concerning ML for big data analytics and diverse methodologies in the framework of up-to-date computer systems for several social applications. Their specific goal is to examine the potential and problems of ML on big data and how it impacts the world, as well as to cover discussions on ML in Big Data in certain socioeconomic domains.

It has been demonstrated by Azi Lev-On [14] that social network are exposed as central discourse fields that consist of a wide range of stakeholders, including opinion and political leaders, terrorists, international celebrities, and anonymous children. Even soldiers and civilians who died could express themselves on Facebook through their Facebook profiles, which became a memorial after they passed away.

III. METHODOLOGY

3.1 *Preprocessing*

A preprocessing procedure is necessary for efficiently removing noise data and extracting the noise-free information. It can also boost classification performance. The preprocessing stages are as follows:

- Delete any data that is null or incorrect.
- Sorting
- Standardization

The blank (null) information in the NSD database are deleted, and the complete data are obtained for efficient elimination of the undesirable data and retrieval of the complete bunch of bits. The filtering applications are then extracted to eliminate the noisy pieces of information and recover the noise-free information. The filtered information is then standardized using a normalizing algorithm, yielding efficient preprocessed information.

3.2 *Feature Selection using Enhanced GBDT Algorithm*

a) GBDT Algorithm

DT are learning techniques that operate by analyzing and resolving on data properties. In DT, characteristics are nodes, and every leaf node represents a categorization. DT methods start with a series of cases, each with its own set of characteristics. The characteristics are classified depending on their features. The characteristics are expressed by symbols or mathematical numbers. For the training step, all subsets of characteristics are arranged. GBDTs are a ML [15] approach for improving a model's prediction value over repeated phases in the learning procedure. For every cycle of the DT, the scores of the coefficient, weighting, or biases assigned to all of the input parameters used to forecast the final value are adjusted with the purpose of reducing the loss ratio. After selecting the characteristics, the quantity of data is evaluated, that is called as entropy. Entropy is an indicator of the level of unpredictability. This work optimized the features based on the Shannon Entropy which is the one type of entropy. It has been used to find information gain in GBDT algorithm. The Shannon entropy is represented in equation (1).

$$-\sum N_{I=1} P_i \; log_2 p_i + -\sum p_i log_2 p_i \qquad (1)$$

The GBDT approach recognizes the optimal features to be picked for categorization by using the entropy score. Generally, GBDT algorithm need to calculate the entropy of child node with weight average which ensure to determine the information gain.

b) Information Gain (IG)

IG is employed to quantify the "data" of a feature that provides information regarding the class and is shown in equation (2). IG is the primary key utilized by DT methods to build a DT. The DT method is continually trying to enhance IG. The characteristic with the greatest IG will be tested or divided first.

$$\text{Information gain} = \text{Entropy (parent)} - [\text{Weight average}] * \text{Entropy (children)} \qquad (2)$$

3.3 *Enhanced Gradient-Boosted Decision Tree (EGBDT)*

This method improves the GBDT measure by applying the CFS and FW calculation method.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6313
***Article Received**: 24 December 2022 **Revised**: 28 January 2023 **Accepted**: 08 February 2023*

a) Correlation based Feature Selection (CFS)

CFS technique is an important key feature identification method. It evaluates the subset characteristics used to determine the particular prognostic ability of the characteristics. The correlation coefficient is employed to connect the link between subset characteristics and factor categories, encompassing inter and intra-correlation amongst features. The relating characteristics group promotes correlation between characteristics and classes while eliminating inter-correlation. The CFS technique is mostly employed in the DT technique to identify the optimal characteristics. It may also be used with other finding methods such as forward selecting, bi-directional searching, and more. The CFS computation is represented by equation (3),

$$r_{ZC} = \frac{N\,\overline{r_{Zi}}}{\sqrt{N + N\,(N-1)\overline{r_{ii}}}} \quad (3)$$

Where, r_{ZC} signifies the association between the feature sum and subset sum. The amount of characteristics N, r_{Zi} denotes the middling value of the association and r_{ii}signifies the middling rate of the inter connection between the feature subgroups. This average correlation value helps to identify the six optimized features called labels or subset and two optimized features called class.

b) Feature Weighting (FW)

FW is often seen as a subset of selecting features that might weaken the notion of feature conditioned independence. The EGBDT is another way to take the feature weighting for calculate the weight between two class features namely social and anti-social. Generally, FW calculation helps to know the relation and weight values between the two selected attribute variables. This work modified their average confidence score as two class feature weight that is represents in equation (4).

$$\mathrm{w}_1 = \frac{\mathrm{AC(i)}*\mathrm{N}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{AC(i)}} \quad (4)$$

Here AC(i) is the mean confidence of $\mathrm{F_i}$. In addition, as mentioned in equation (5), the Relief coefficient can be employed for FW.

$$\mathrm{w}_2 = \frac{\mathrm{RC(i)}*\mathrm{N}}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{RC(i)}} \quad (5)$$

To address the impact of the association value and mean confidence rating on classification outcomes, this study combines aforementioned coefficients to produce novel weighting coefficient, as well as its computation is shown in equation (6).

$$\mathrm{w_i} = \frac{(\mathrm{w}_1 + \mathrm{w}_2)}{2} \quad (6)$$

The above feature weighting value is used to measure the weight between two class features.

3.4 *Tweet Analysis*

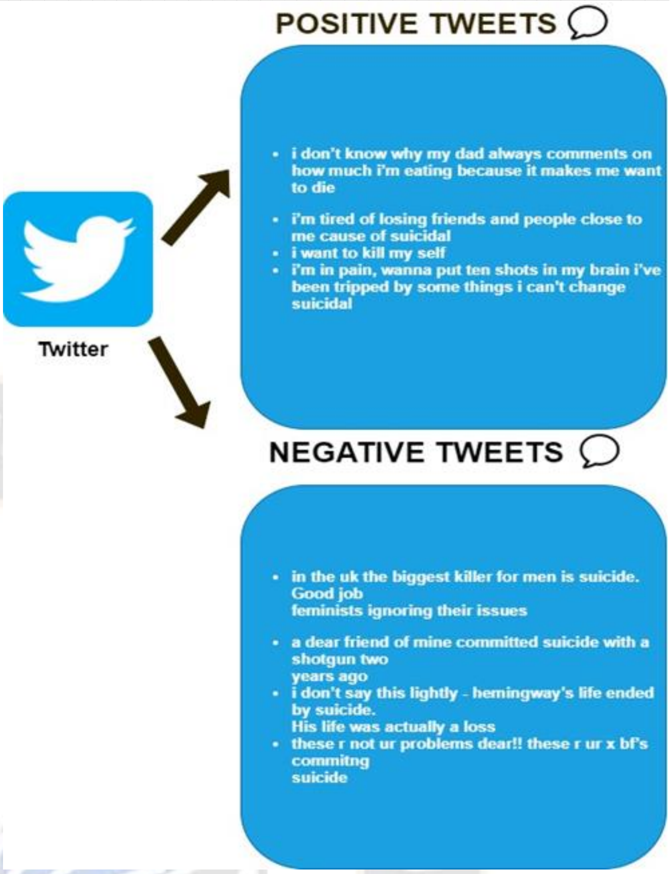

Figure 2. Positive and negative tweets

Figure 2 represents the tweet analysis of twitter which contains the positive and negative words. The above tweets consider the example of this work.

3.5 *Bayesian Certainty Method*

The Bayesian system is built on the probability-based Bayes statement. The Bayes statement is stated mathematically in the equation (7).

$$P\left({}^{a}/_{b}\right) = \frac{P({}^{b}/_{a})P(a)}{P(b)} \quad (7)$$

The certainty factor is one method to improving the Bayesian network accuracy and to solve the uncertainty problem. In general, Bayesian network show the result as probability. Bayesian network is not sufficient in sentiment analysis which means analysis need to show the result as concrete. So, that way this work use the CF is combined with Bayesian network. This work calculate the CF which contains the two kinds of certainty factors, namely social and anti-social to classify the features and make the result as certainly. Certainty factor is defined as CF (S, A). Where CF (P, N) is positive and negative. The CF ratio ranges from -1 to 1. A number of -1 indicates total doubt, whereas a value of 1 indicates perfect assurance. The positive and negative features of certainty filled by the subsets based on tweets such as social, probably safe, highly social, anti-social, highly anti-

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6313
***Article Received**: 24 December 2022 **Revised**: 28 January 2023 **Accepted**: 08 February 2023*

social and suspicious. Every subset obtained from the preceding approach or user-supplied confidence factor would be converted or transformed into a specific value ranging from 0 to 1. The below table 1 jargon is used to translate factor certainties to digits.

Table 1. Certainty Factor

Feature no	**Uncertainty terms**	**CF**
1P	social	0.4
2P	Probably Social	0.6
3P	Highly Social	1.0
1N	Anti-Social	-1.0
2N	Highly Anti-social	-0.8
3N	Suspicious	-0.6

Every element is a certainty of the structure in a premise, and every premise will be computed using the method below.

- **For solo premises**

Whenever there are 2 occurrences, p and q,

$$\mathrm{CF}(\mathrm{P},\mathrm{N})_{(\mathrm{i})} = \mathrm{CF}(\mathrm{P})_{(\mathrm{i})} * \mathrm{CF}(\mathrm{N})_{(\mathrm{i})} \quad (8)$$

- **For Shared Premises**

$$\mathrm{CF}_{(\mathrm{cmb})}(\mathrm{P},\mathrm{N})_{(\mathrm{i,j})} = \mathrm{CF}(\mathrm{P},\mathrm{N})_{(\mathrm{i})} + \mathrm{CF}(\mathrm{P},\mathrm{N})_{(\mathrm{j})} * (1 - \mathrm{CF}(\mathrm{P},\mathrm{N})_{(\mathrm{i})}) \quad (9)$$

1P= 0.4
2P= 0.6
3P= 1.0
1N= -1.0
2N= -0.8
3N=-0.6

As the above formation first three are anomaly feature set otherwise. The categorization outcome is stated utilizing certainty factor approach as follows:

3.6 *Compute the likely range of every premise*

$$\mathrm{CF}(\mathrm{P},\mathrm{N})_{(\mathrm{i})} = \mathrm{CF}(\mathrm{P})_{(1\mathrm{P})} * \mathrm{CF}(\mathrm{N})_{(1\mathrm{N})}$$
= 0.4* (-1.0)
= -0.6

$$\mathrm{CF}(\mathrm{P},\mathrm{N})_{(2)} = \mathrm{CF}(\mathrm{P})_{(2\mathrm{P})} * \mathrm{CF}(\mathrm{N})_{(2\mathrm{N})}$$
= 0.6 * (-0.8)
= -0.2

$$\mathrm{CF}(\mathrm{P},\mathrm{N})_{(3)} = \mathrm{CF}(\mathrm{P})_{(3\mathrm{P})} * \mathrm{CF}(\mathrm{N})_{(3\mathrm{N})}$$
= 1.0*(-0.6)
= 0.4

To carry out a mixture of a premise, take the subsequent stages:

- The combination premise is computed by combining premise 1 and 2, and the outcome is known as the Old1 hypothesis.
- The old1 hypothesis will be reassembled into premise3 utilizing the same combining procedure, except that the old1 hypothesis will substitute the first premise, and the resulting combo will be known as the old2 premise.
- If the resultant value is positive that is social activity and if the resultant value is negative that is anti-social activity. This feature results based on its subset features which are described in table 2 & 3.

Table 2. $\mathrm{CF}(\mathrm{P},\mathrm{N})_{(1)}$& $\mathrm{CF}(\mathrm{P},\mathrm{N})_{(2)}$

Tweet	Probably social	Anti-social	Suspicious	Highly Social	Positive/ Negative
Suicide is the leading cause of death among men in the UK. Well done, feminists, for disregarding their problems.	0.6	-1.0	-0.6	1.0	Positive

$$\left.\begin{matrix}\mathrm{CF}(\mathrm{P},\mathrm{N})_{(\mathrm{i})}\\ \mathrm{CF}(\mathrm{P},\mathrm{N})_{(2)}\end{matrix}\right\} result, result + \mathrm{CF}(\mathrm{P},\mathrm{N})_{(3)}$$

$$\mathrm{CF}_{cmb}(\mathrm{P},\mathrm{N})_{(1,2)} = \mathrm{CF}(\mathrm{P},\mathrm{N})_{(1)} + \mathrm{CF}(\mathrm{P},\mathrm{N})_{(2)} * (1 - \mathrm{CF}(\mathrm{P},\mathrm{N})_{(1)})$$
= -0.6 + -0.2 *(1-(-0.6))
= -0.8*(0.4)

$$\mathrm{CF}_{cmb}(\mathrm{P},\mathrm{N})_{(1,2)} = -0.8$$

Table 3. $\mathrm{CF}(\mathrm{P},\mathrm{N})_{(1)}$& $\mathrm{CF}(\mathrm{P},\mathrm{N})_{(2)}$

Tweet	Social	Anti-social	Highly anti-social	Probably social	Positive /negative
I'm not sure why my father continually remarks on how much I eat since it makes me want to die.	0.4	-1.0	-0.8	0.6	Negative

$$\mathrm{CF}_{cmb}(\mathrm{P},\mathrm{N})_{(\mathrm{result},2)} = \mathrm{CF}(\mathrm{P},\mathrm{N})_{(\mathrm{result})} + \mathrm{CF}(\mathrm{P},\mathrm{N})_{(2)} * (1 - \mathrm{CF}(\mathrm{P},\mathrm{N})_{(\mathrm{result})})$$
= -0.32 + 0.4 *(1-(-0.32))
= 0.08*0.68

$$\mathrm{CF}_{cmb}(\mathrm{P},\mathrm{N})_{(1,2)} = 0.054$$

IV. RESULT AND DISCUSSION

The proposed system was implemented using Python with 4 GB RAM, 1 TB hard disk, 3.0 GHz Intel i5 processor. The effectiveness of the suggested work is comparing with different existing classification techniques and algorithms. The following terms which are used for efficiency analysis,

- Accuracy
- Precision
- Recall

Table 4. Accuracy analysis between classification techniques

Data size (MB)	**Accuracy (%)**			
	SVM	**ID3**	**C4.5**	**EGBDT**
100	80	83	87	95

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6313
***Article Received**: 24 December 2022 **Revised**: 28 January 2023 **Accepted**: 08 February 2023*

250	78	75	80	87
500	67	71	78	80
750	61	64	70	75

Table 4 contains the accuracy of different algorithms such as SVM, ID3 & C4.5 are compared with the proposed EGBDT algorithm.

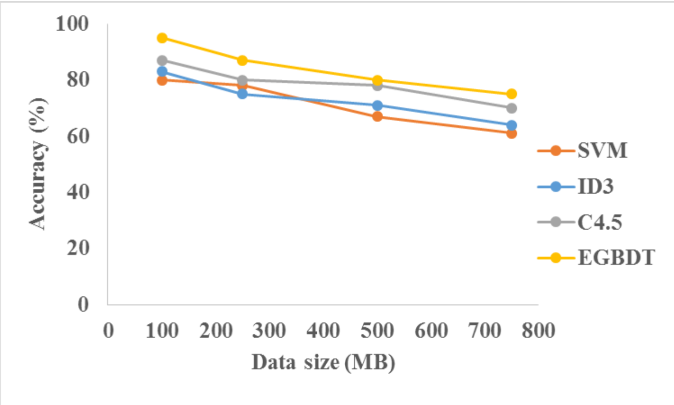

Figure 3. Accuracy Comparison

Figure 3 shows the feature prediction accuracy of SVM, ID3, C4.5, & EGBDT. This analysis EGBDT attains high prediction accuracy than others. Because EGBDT contains the CFS and FW feature based calculation which makes the prediction in better way.

Table 5. Precision analysis between classification techniques

No. of Tweets	Precision		
	SVM	KNN	BCM
100	0.53	0.59	0.80
500	0.38	0.55	0.74
1000	0.35	0.43	0.68
1500	0.25	0.36	0.57
2000	0.14	0.24	0.46

Table 5 contains the Precision values of different algorithms such as SVM, & KNN are compared with the proposed BCM.

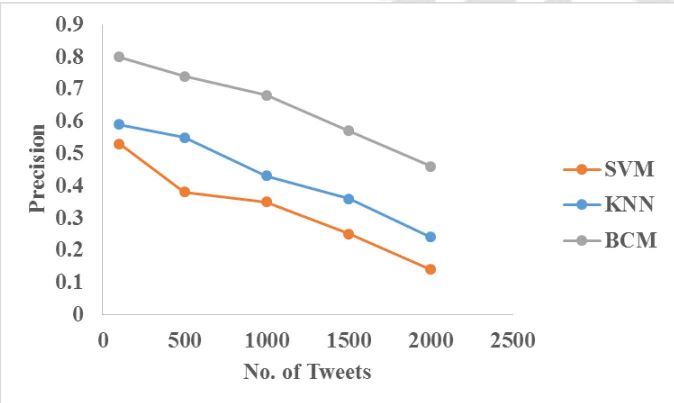

Figure 4. Precision Comparison

Figure 4 shows the precision value comparison of SVM and KNN is lower than the BCM which contains the certainty factor values to identify the tweet decision is easier than others with short time.

Table 6. Recall analysis between classification techniques

No. of Tweets	Recall		
	SVM	KNN	BCM
100	0.65	0.77	0.91
500	0.57	0.71	0.85
1000	0.49	0.66	0.79
1500	0.40	0.59	0.64
2000	0.33	0.61	0.57

Table 6 contains the Recall values of different algorithms such as SVM, & KNN are compared with the proposed BCM.

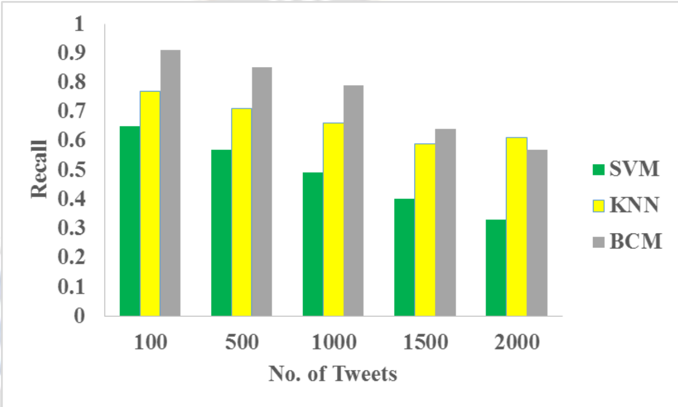

Figure 5. Comparison of Recall

Figure 5 shows the recall value comparison of SVM and KNN is lower than the BCM. It optimized the six features using CFS and FW in EGBDT section for calculate the positive and negative tweets.

V. CONCLUSION

The problem of sentiment classification on Twitter is presented in this research: for a certain tweet, sentiment orientation (if beneficial or harmful), and the features are taken from the NSD database and afterwards preprocessed. After that, features are selected and then weights are calculated using EGBDT algorithm. Some of the tweets are taken and analyzed by the BCM technique. It provides the certainty value for selected features to find the sentiment in twitter. Finally, classified the analyzed sentiment which is either positive or negative. The suggested work's effectiveness is evaluated against the previous works employing performance criteria such as correctness, precision, and recall. The proposed work obtained better accuracy, precision and recall than the existing works.

REFERENCES

[1] Qixuan Hou, Meng Han, and Zhipeng Cai (2020), “Survey on Data Analysis in Social Media:A Practical Application Aspect”, Big Data Mining and Analytics, 3(4), pp259–279.

[2] Mondher Bouazizi and Tomoaki Ohtsuki (2018), “Multi-Class Sentiment Analysis in Twitter: What If Classification Is Not the Answer”, IEEE access.

[3] Koyel Chakraborty , Siddhartha Bhattacharyya , Senior Member, IEEE, and Rajib Bag (2020), “A Survey of Sentiment Analysis from Social Media Data”, IEEE Transactions on Computational Social Systems.

[4] Balakrishnan, Vimala & Khan, Shahzaib & Arabnia, Hamid. (2020). Improving Cyberbullying Detection using Twitter Users’ Psychological Features and Machine Learning. Computers & Security. 90. 101710. 10.1016/j.cose.2019.101710.

[5] Abinash Tripathy (2017), “Classification of Sentiment of Reviews using Supervised Machine Learning Techniques”, International Journal of Rough Sets and Data Analysis, 4 (1).

[6] Ali Hasan , Sana Moin , Ahmad Karim and Shahaboddin Shamshirband (2018), “ Machine Learning-Based Sentiment Analysis for Twitter Accounts”, Math. Comput. Appl.

[7] J. Rogstadius, M. Vukovic , C. A. Teixeira ,V. Kostakos ,E. Karapanos&J. A. Laredo (2013). Crisis Tracker: Crowd sourced social media curation for disaster awareness, IEEE. VOL. 57 NO. 5 PAPER 4.

[8] Congyu Wu and Matthew S. Gerber (2017). Forecasting Civil Unrest Using Social Mediaand Protest Participation Theory. IEEE.

[9] EimanKanjo, Daria J. Kuss, And Chee Siang Ang (2017).NotiMind: Utilizing Responses to Smart Phone Notifications as Affective Sensors. IEEE access.

[10] Suman Deb Roy, GiladLotan, and Wenjun (Kevin) Zeng (2015).The Attention Automaton: Sensing Collective User Interests in Social Network Communities. IEEE, VOL. 2, NO. 1.

[11] SudhaSubramani, Sandra Michalska, Hua Wang, Yanchun Zhang, And HaroonShakeel, Jiahua Du (2019). Deep Learning for Multi-Class Identification from Domestic Violence Online Posts. IEEE access.

[12] SudhaSubramani , Hua Wang, HuyQuanVuand, and Gang Li (2018). Domestic Violence Crisis Identification from Facebook Posts Based on Deep Learning. IEEE access.

[13] L. H. Son et al. (2019). Machine learning on big data: A developmental approach on societal applications in Big Data Processing Using Spark in Cloud. Springer.

[14] Lev-On A. The Anti-Social Network? Framing Social Media in Wartime. Social Media + Society. July 2018. doi:10.1177/2056305118800311.

[15] Neelakandan, S., & Paulraj, D. (2020). A gradient boosted decision tree-based sentiment classification of Twitter data. International Journal of Wavelets, Multiresolution and Information Processing, 18(04), 2050027. https://doi.org/10.1142/s0219691320500277