A MULTI-CLASSIFIER APPROACH FOR TWITTER SPAM DETECTION USING INNOVATIVE ANN-FDT ALGORITHM

M.Arunkrishna
Research Scholar, Department of Computer Science,
Jairams Arts and Science College (Affiliated to Bharathidhasan University),
Karur - 639003, Tamilnadu, India.
arunkrishna.murugan@gmail.com
https://orcid.org/0000-0001-9310-9299

B.Mukunthan
Research Supervisor, Department of Computer Science,
Jairams Arts and Science College (Affiliated to Bharathidhasan University),
Karur - 639003, Tamilnadu, India.
dr.mukunthan.bmk@gmail.com.
https://orcid.org/0000-0001-8452-3164

Abstract - Nowadays, various social media platforms are available in Internet like Facebook, Twitter and Instagram for uniting the people. Twitter is one among the most famous platform in social media due to its available information among users. Users allows to find new friends and update their latest information and activities. Twitter is using Google Safe-browsing to detect the spam URL and block spam links. Due to the presence of advanced API which enables to read and write the data in Twitter, different kinds of spammers are attracted in the Twitter. There are various existing researches applied various machine learning techniques to determine the twitter spam. However, there is no comprehensive evaluation on their algorithms and lack of accuracy in large dataset. To rectify these issues, this research proposed hybrid method with the combination of Artificial Neural Networks with Fuzzy Decision Tree (ANN-FDT). The proposed classifier classified the span and non-span tweets based on the labels. For experimental analysis, the proposed classifier applied on large dataset of 600 million public tweets. The performance of proposed algorithm is evaluated by means of measures like accuracy, TPR, FPR and F-measure. From the results it can be seen that the proposed technique has improved performance.

Keywords: Twitter Spam; Spam detection; ANN-FDT; Accuracy; TPR; FPR; F-measure.

1. Introduction

Nowadays, OSNs (i.e.) Online Social Networks acts as one of the most famous tools utilized by millions of peoples on the Internet for communicating and collaborating with each other [1]. Twitter is one of such a major platform of social networking that attracts the people around the world through granting free services like microblogging for its users, can able to determine the messages or tweets up to 140 characters, the users can able to follow other users and famous personalities, etc. It can be used across different types of devices like smartphones, personal computers, and tablets. Numerous users across the Twitter platforms share their expressions, feelings, news, and special moments in the form of tweets to their followers. Due to the easy access and convenience of using it in desired places, numerous criminals or unauthorized users such as spammers are attracted easily to Twitter. There are different types of attacks like a scam, phishing, and spam that are available, which possess the ability to attack the Twitter platform in an unauthorized manner. Different types of suspicious behavior available online have been depicted in Fig 1 [2].

The suspicious behavior includes different types of spams like email spam, SMS spam, and web spam. In addition to this, different kinds of reviews, messages, likes, followers, Sybils, pages posts, suspended spammers, etc. which are observed to be suspicious have been reported. Based on the availability of necessary labeled examples, the efficiency of supervised learning approaches can be improved [3]. This concept is applied to various purposes, like detecting the objects, documents, and categorization of web pages. However, the process of labeling the examples is very complex, requires high cost, and consumes more time as they need experimental analysis or experienced persons. Hence, a deep learning technique, which is an advanced form of supervised learning technique and the usage of deep learning results in resolving the issues listed above.

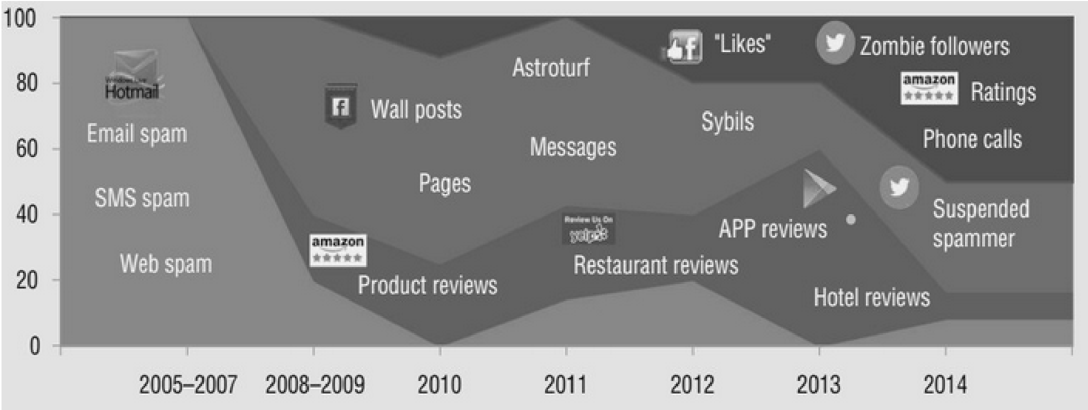

Fig. 1. Various types of suspicious behavior.

[4] The evaluation of probability is performed through the decision tree classifier, which distributes the class in the form of leaves in the branches of the tree. It can be seen that the class distribution is utilized in the form of selection measures for self-training, but ultimately the performance of classification carried out by the self-training algorithm is not improved. Hence the unlabelled data is not benefited by the algorithm. Predictions are made through the distribution of class at leaves of a tree's branches. And due to this, the decision tree classifier could not give away with improvement in the ranking of predictions. And in order to solve this issue, a deep learning technique is employed along with the decision tree classifier, and this is given as novelty in our proposed methodology.

1.1. *Objectives*

- To detect the spam tweets by using the Innovative ANN-FDT classifier.
- To reduce the false detection rate.
- To classify between the spam and ham (non-spam) tweets.

2. Related Works

[5] For detecting the spam existing in the social media platform of Twitter, a framework of semi-supervised spam detection (i.e., S3D) was proposed in the research work. Two different modules namely spam detection module and model update module were present within the proposed framework, in which the spam detection module work in real-time mode and the model update module that can work in the batch mode. In the module, there are four lightweight detectors introduced for detecting spams. There are tweets which possess blacklisted links or URLs were labeled by the blacklisted domain detector, and the tweets which were suspected as duplicates of the pre-labeled tweets were labeled through the near duplicate detector. The reliable nonspam (ham) detector labeled the tweets which were posted by trusted, reliable users without any spammy or unwanted words, and finally, the remaining tweets were labeled with the aid of detectors based on multi classifiers. Depending upon the tweets which were labeled in the previous time window, the data which are essential for the detection module are got updated. The experimentations performed on the larger sized information exhibit that the proposed system was capable of acquiring knowledge regarding the new patterns related to the unwanted activities and can also be able to sustain enhanced accuracy of detecting the presence of spam in a stream of the tweet.

[6] The impact of uneven distribution among the spam and non-spam categories on the spam detection rate was demonstrated experimentally in the research work. This issue of uneven distribution was addressed by proposing an oversampling technique based on the fuzzy logic, and the proposed technique can able to generate synthetic data samples from very little restricted samples with respect to the concept of fuzzy-based decomposition of information. In addition to this, an ensemble learning technique was developed that possesses the ability to learn higher and accurate classifiers within three steps, even from the imbalanced data itself. The distribution of class within the imbalanced data was adjusted through employing different techniques like FOS, random under sampling, and random oversampling in the first step. After this in a model was separately constructed in the successive step for classification based on every data sets that are redistributed. And a popular voting technique

was presented in the final step for combining the predictions that are obtained from every classification model. The analysis was carried out in the data obtained from the recent tweets from Twitter. Resultants attained after the experimentation indicate that the learning technique presented in work can able to enhance the detection rate of spams within the datasets along with imbalanced distribution, in a significant manner.

[7] Deceptive information present within the unwanted Twitter spam was presented in the research work. Around 500 million tweets were gathered with about 6% spam, and the analysis was performed for the collected information. It was obtained from the analysis that numerous deceptive content of spam functions dissimilar in order to attract the victims to various malicious sites. The regional response rate for numerous outbreaks of Twitter spam was found to higher. The performance of various spam detecting approaches was enhanced based on the fact of obtaining results that were discussed above.

[8] Performance of a wider variety of conventional techniques of machine learning was compared to identify the most desired and effective performance. The stability was also determined depending upon the huge rate of available ground truth data. Further, the techniques were evaluated by means of scalability property in order to attain a better detecting capability of Twitter spam in real-time. Performance analysis was performed for various measures like TPR, FPR, detection accuracy, and F measure. The stability of the compared algorithms was evaluated through the aid of training samples with varying sizes that were selected randomly. The impact of the parallel computing environment was well understood through the property of scalability with respect to the minimization of training as well as the testing time of the compared machine learning techniques.

[9] For solving the issue of Twitter Spam Drift, a deep analysis was performed initially in the statistical features that were obtained of about one million spam and same amount of non-spam tweets. After this, an innovative scheme of Lfun was proposed in the research paper. The modified spam tweets were discovered from the unlabelled tweets by the usage of the proposed methodology. Then the discovered tweets were employed to the training stage of the classifier. The proposed technique was evaluated by performing numerous experiments. Outcomes of the experimentation indicate that the Lfun technique proposed in work can able to increase the accuracy of spam detection in real-world situations.

[10] A semi-supervised approach was proposed in work in order to detect the spam occurring in Twitter through the usage of a specialized ensemble-based technique that consists of four different classifiers. The proposed approach was developed with respect to the utilization of different PDS (i.e.) Probabilistic Data Structures such as OF (i.e.) Quotient Filter that can able to question the database of URL, unwanted users, a database containing spams and LSH (i.e.) Locality Sensitive Hashing. LSH possesses the ability to search the similarities and classifiers in different stages that can offer rapid outcomes with much-minimized effort for the computational process. The performance of the suggested technique was determined through performing the comparative study for the PDS that possess related structures of data. Performance measures like precision, recall, F score were calculated for the evaluation of the system's performance.

[11] investigated the class imbalance pertaining to the detection of spams on Twitter. Initially, a comparative investigation was performed concerning about few famous techniques that can handle the issue of imbalance for the determination of efficient technique, which can address the issue of class imbalance. Subsequently, another comparison was performed for the detection of spams on Twitter with respect to numerous conventional approaches. Results attained on performing the experimentation denote that ensemble learning based on a fuzzy approach can enhance the performance of classification significantly with respect to the gathered ground truth Twitter data.

[12] The spam profiles were identified by proposing a hybrid technique which combines the analytics of social media along with the bio-inspired computing. The spams in the Twitter market were detected by adopting a modified algorithm that integrates K-means with LFA (Levy flight Firefly Algorithm) and also along with the extension of chaotic maps (i.e.) FA which is abbreviated as Firefly Algorithm. Tweets of about 18, 44, 701 in count were gathered from a variety of 14, 235 Twitter profiles, which varies on 13 different statistically remarkable items, which were derivatives of social media analytics. Moreover, the overlapping of users in two different clusters of spammers and non-spammers was identified by the usage of the Fuzzy C-Means clustering approach. Tests were conducted for six different categories of FA integrated with K-mean's chaotic maps and levy flights. Resultants of the tests denote that the proposed FA with chaos converges to a working solution rapidly on employing the Gauss map technique.

[13] proposed a novel system for Twitter Spam Detection. in the experiment. The proposed method claims that the spam in Twitter could not be detected efficiently through the conventional spam detection technique due to the unique characteristics possessed by the Twitter. In order to combat the problem, proposed system makes use of the specific features of Twitter for the detection of spam in the Twitter. Through the usage of API, which was provided by Twitter, a wide range of 77, 033 tweets posted by about 50, 490 users on Twitter were gathered. The spammers were classified separately from the legitimate users through the employment of Naïve Bayes, which was utilized for training the Twitter Spam Detector system. The value of accuracy and sensitivity were found to be 0.943 and 0.913 correspondingly, depending upon the results that were attained after the evaluation.

[14] K-L divergence was adopted in the presented work for denoting the distribution of spams, and the possible drifts were localized through the usage of MDDT (i.e.) Multi-scale Drift Detection Test. An improvement in the performance is observed by retraining the base classifier depending upon the results of detection. Experimental results states that, whenever a drift occurs, the adopted K-L divergence method possess constant change in patterns among the features. Then the final results obtained for the classifications were found to be more efficient and enhanced by means of performance measures like F-measure, recall, and accuracy.

[22] proposed a hybrid system of INB-DenStream that combines the concept of DenStream along with INB (i.e.) Incremental Naïve Bayes. The efficiency of INB-DenStream was presented by measuring the performance by means of general precision, general recall, purity, parameter sensitivity, F1 measure, and complexity in the computational process. The performance of the proposed technique was compared with various other prevailing techniques like CluStream, DenStream, and StreamKM ++. Results obtained at the end of comparison implies that the proposed technique surpasses the performance of other prevailing techniques.

[16] Describe clear approach for spam detection and prevention system. Anti-spam systems may fall under either traditional approaches or mathematical approaches. Techniques like blacklisting content filtering comes under traditional approaches. Statistical ml and AI techniques are coming under mathematical approaches. There has been variety of algorithms to accomplish the task these are known as filters. The filters have the ability to to fine-tune itself by learning. An initial teaching is needed so that the algorithm can adopt based on previous experiences. Standard method includes exploring data, visualisation, cleaning, feature extraction, algorithm implementation, scoring and analysis.

[19] Detection of Twitter spam is done by developing an Intelligent Twitter Spam Detection System that is capable of granting very precise information regarding the spam profiles. The system is a single hybrid classifier that considers unique feature sets prior to the assessments of the tweets gathered and validates the links through employing API in Google Safe Browsing for granting additional security. Hence, an improved classification was done for the gathered tweets, and the intelligent detection of twitter spam was also provided by the proposed system.

[17][18] presented a method for identification of unique patterns using AI. It has the processor for neural fuzzy pattern recognition. Holds three generators namely input generator for ANN learning and also used for normalization. The activation function generator activates the output node based on corresponding input weights, the match function generator with vigilance parameter decides the better not for output generation.

[23] Presented a framework to restore the hidden capability of given problem the artificial neural BSF filter substitutes the code design so that we can analyse the data with the random feature, multilayer perceptron on customer object communication function. So that the proposed ANN gives satisfactory performance. The hidden data and ANSF must be trained before Matrix factorization. Hybrid of CMfact MPeep are used to investigate the interaction feature and for flexible factorization..

To work with large number of fuzzy sets [15] use principal component analysis (PCA) based on non-parametric statistical technique for dimensionality reduction with fuzzy image sets. The proposed machine learning approach is very useful in finding similarities between fuzzy image sets

3. Proposed Work

3.1. *Proposed Flow*

Numerous tweets are gathered from Twitter in order to classify them into spam and non-spam tweets. Data from the Twitter Spam Dataset are gathered and are subjected to pre-processing. After the pre-processing process, the tweets are subjected to training and testing. First data training is completed and after that the trained tweets are processed to the neural networks. In the neural networks, n number of neural networks are formed. Then the layers are combined with fuzzy decision tree classifier. Then the classifier is integrated with Artificial Neural Networks and ANN-FDT is proposed. The tested tweets are processed in the proposed ANN-FDT classifier. Then the predictions are made that whether the gathered tweets are spam or non-spam tweets. Figure2 shows the flow.

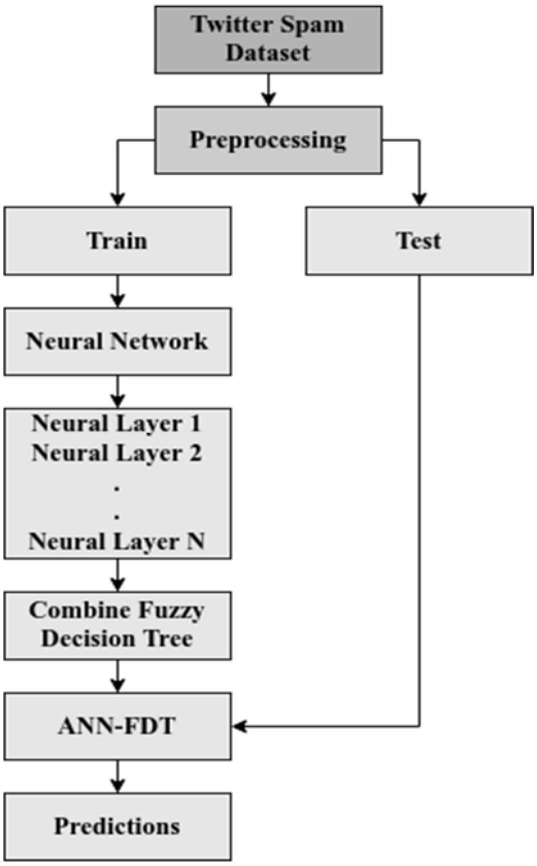

Fig. 2. The flow of the proposed work

3.2 *ANN-FDT Algorithm*

In this ensemble model, initially artificial neural network is used with different aspects to determine the spam tweets. In network, the input layer, hidden layer and output layer presents. Each twitter information are grouped into word. It performs different activation functions and different combinations of activation functions. In Fuzzy Decision tree, same training dataset is used that are normalized into real numbers as [0, 1]. In this method, the powerful fuzzy sets are used to attain more flexible membership functions. In this FDT technique each attribute has been associated with fuzzy membership function. This hybrid method provides more accuracy and efficiency for detecting the spam tweets. Artificial neural network can process information like biological neurons.[21] the learning process of a neural network is training and the problem-solving process with the acquired knowledge is known as inference. Inference is the mapping of input patterns to their correlated output pattern. Based on its nature, there are variety of ways in which ANN can be applied.

The pseudo-code for the proposed ANN-based Fuzzy Decision Tree (i.e.,) ANN-FDT is given as:

1. Calculate the maximum depth of a branch (B_d).
2. The number of branches at every level is counted as $num_b l$
3. Every input of maximum depth occurring in the branch is expressed as max_a^L.
4. Hidden layers will be present for every maximum branch depthB_d. Similarly, the input layer with neutrons (i.e.)inp_{Ne}, output layers with regression (i.e.) out_{Ne} and classification of neurons in the neural network (i.e.) $Class_N$ are also represented.
5. A number of neurons are present with every hidden layers $h_{(l)}$ and it is given by $h(l)=h(l-1)+num_b(l)$ arrays are created and the dimensions of the created arrays are $h(l)$ x $h(l-1), l=1,\dots,B_d$.
6. The initial weights wei^{B_d+1} are stored with the dimension of $h(B_d)$ x $out_{Ne}(Class_N)$. The arrays are initialized to 0.
7. For each input $a=0,1,\dots,inp_{Ne}-1$:
8. Set $wei_{a,a}^{l} \rightarrow 1 for l < max_a^L$

Input values are passed inside the hidden layers, till the branches of the decision tree splits into very long.

Recurse through decision-paths of the decision tree:

Similar to the creation of traditional decision trees, the proposed fuzzy-based decision trees are constructed, except for the measurement of entropy. Through the implementation of fuzzy sets membership functions, the gain of information G_{gain} in fuzzy ID3 is determined. Operator id of the arithmetic product is intersected with fuzzy sets, through the utilization of the proposed work [20].

$$b_t^p \rightarrow |b_t^d|/|b^d|$$

$$b^d \rightarrow \sum_{x \in d} \left(\prod_{(\text{a,b}) \in Q} o_{\text{ab}}(x) \right)$$

$$b_t^d \rightarrow \sum_{x \in d_k} \left(\prod_{(\text{a,b}) \in Q} o_{\text{ab}}(x) \right)$$

In H_u , H_v,EAG the H_u,H_v,E ∧ G_{gain} equations, the membership functions of attributes (a, b) is denoted as O_{ab}, where bth attribute value of ath attribute. A set of pair (a, b) is denoted as Q alongside the branches from the root to node j. Calculation are carried out similar to the calculation of original ID3 approach. Set of leaf nodes are indicated as L_{node} and element of a set is denoted as l. x – the sample data belongs to class t based on the probability, which is calculated based on the formula:

$$P_{(x)}^t \rightarrow \sum_{l \in L_{\text{node}}} p_t^l \cdot \left(\prod_{(\text{a,b}) \in Q_l} o_{\text{ab}}(x) \right) = \sum_{l \in L_{\text{node}}} p_t^l \,.o_{\text{ab}}(x)$$

The relative frequency of a class t at the lth leaf node is denoted as p_t^l and the set of (a,b) branches from the root to lth leaf node is represented as Q_l. Then the multidimensional membership function with respect to the lth leaf node is expressed as $o_{\text{ab}}(x)$.

4. Performance Analysis

This section described that the various performance measures such as accuracy, TPR, FPR and F-measure. For experimental results, the proposed method applied to large dataset [18]. We collected 600 million tweets with URLs. By using WSR service, we can identify whether the URL is malicious URLs. In this large dataset, we identified 6.5 million malicious tweets, which accounted for approximately 1 % of all tweets.

4.1. *4.1 Performance Measures*

Performance analysis is carried out by calculating the values of various performance measures like accuracy, sensitivity, recall, precision, F-score, Jaccard-coefficient and missed classification.

4.1.1. Accuracy

Accuracy is the measurement of standard values and known as the weight arithmetic mean and it is considered as the inverse value of the precision value. The accuracy can be calculated based on the formula given below.

$$\text{Accuracy}=\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$$

4.1.2. True Positive Rate (TPR)

Total Number of accurate positive results obtained for every positive sample that are available at the time of analysis is known as True Positive Rate (i.e.) TPR.

4.1.3. False Positive Rate (FDR)

Similarly, the total number of inaccurate positive results obtained for every negative sample that are available at the time of analysis is known as False Positive Rate (i.e.) FPR.

4.1.4. F-measure

The method of F1 score is referred to as the harmonious mean of accuracy and recall. This can be computed with the aid of the given formula

F1-Score= (2*Precision*Recall) / (Precision + Recall).

5. Results

Results are obtained for different performance measures like accuracy, True Positive Rate (TPR), False Positive Rate (FPR) and F-measure were obtained and the results of these measures are compared with numerous existing classifiers in order to prove the efficiency of the proposed work.

The Below Figure (Fig. 3) depicts the values of accuracy, especially for dataset 1. The precision of the presented work is determined, and it is compared with various existing techniques. From the comparison, it is noticed that the proposed method comes up with the highest accuracy value of 0.95.

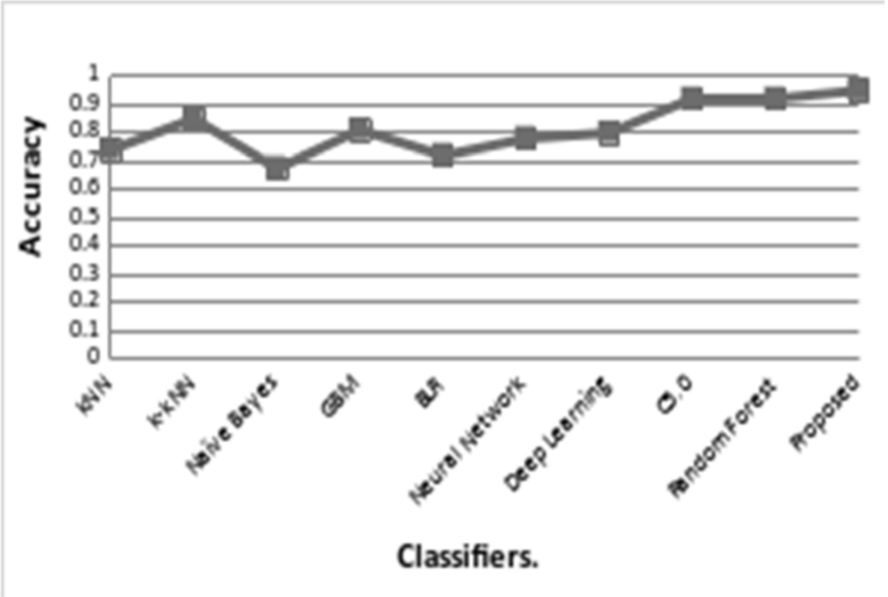

Fig. 3. Accuracy for Dataset 1

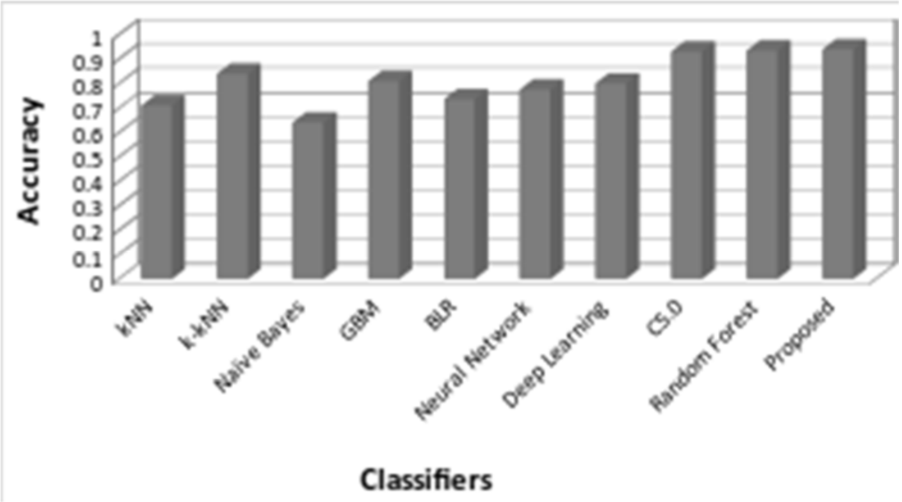

Fig. 4. Accuracy for Dataset 2

Fig. 4 exhibits the values of accuracy, especially for dataset 2. The accuracy of the proposed work is determined, and it is compared with various existing techniques. From the comparison, it can be noticed that the presented method comes up with the highest accuracy value of 0.94.

Fig. 5 indicates the FPR values for both datasets 1 and 2. The FPR of the proposed work is determined, and it is compared with already available techniques. From the comparison, it can be noticed that the proposed method comes up with a reduced FPR value of 6 and 5 for dataset 1 and 2.

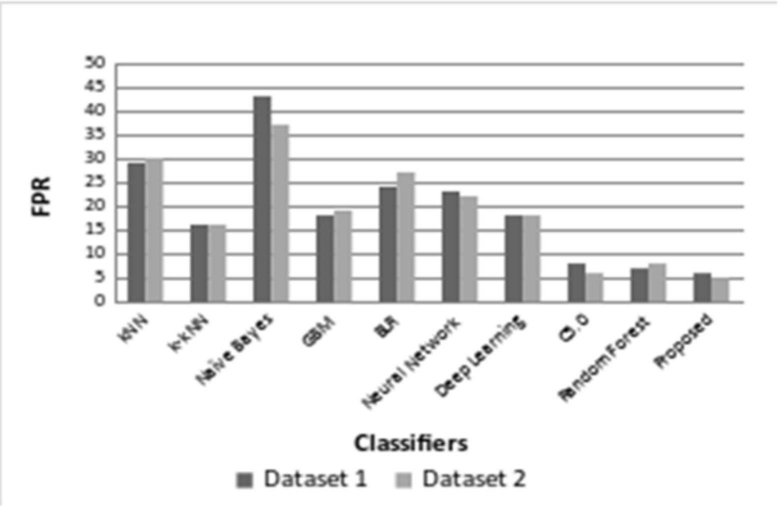

Fig. 5. Comparison of FPR

TPR values in Fig 6 indicates the evaluation criteria for the given datasets. (both 1 and 2). The TPR of the proposed work is determined and it is compared with various existing techniques. From the comparison, it can be noticed that the proposed technique comes up with a higher TPR value of 92 and 93 for dataset 1 and 2.

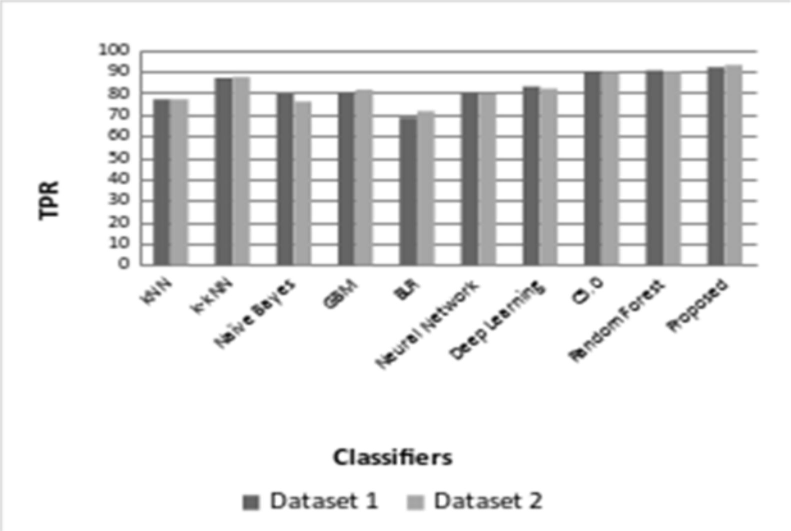

Fig. 6. Comparison of TPR

F-measure values (Figure 7) indicates t for both datasets 1 and 2. The F-measure of the proposed work is determined and it is compared with various existing techniques. From the comparison, it clearly states that the proposed method comes up with a reduced F-measure value of 93.89 and 88.56 for dataset 1 and 2.

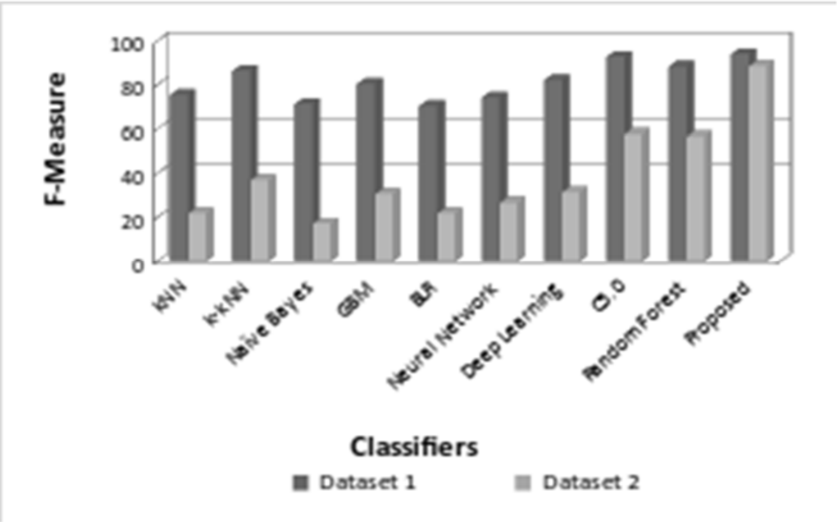

Fig. 7. Comparison of F-Measure

6. Conclusion

To establish a connection with other people, various social media platforms are now available on the Internet, such as Facebook, Twitter and Instagram. Among those, Twitter is one of the dominant social media platforms. Throughout Twitter, different users share their articles, tweets, thoughts, etc. Due to the presence of advanced APIs that allow Twitter to read and write the data, Twitter attracts various types of spammers. In this article, the tweets obtained are categorized by using the Innovative ANN-FDT algorithm. This makes the sorting of the tweets between spam and non-spam. The performance of the proposed algorithm is assessed using measures such as precision, TPR, FPR and F-measure. From the results, it can be seen that performance improved with the proposed technique.

References

[1] Wu, Tingmin, et al., "Twitter spam detection: Survey of new approaches and comparative study”. Computers & Security. 76. 10.1016/j.cose.2017.11.013.

[2] M. Jiang, et al., "Suspicious behavior detection: Current trends and future directions," IEEE Intelligent Systems, vol. 31, pp. 31-39, 2016.

[3] J. Tanha, et al., "Semi-supervised self-training for decision tree classifiers," International Journal of Machine Learning and Cybernetics, vol. 8, pp. 355-370, 2017.

[4] Y. Xia, et al., "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," Expert Systems with Applications, vol. 78, pp. 225-241, 2017.

[5] S. Sedhai and A. Sun, "Semi-supervised spam detection in Twitter stream," IEEE Transactions on Computational Social Systems, vol. 5, pp. 169-175, 2017.

[6] S. Liu, et al., "Addressing the class imbalance problem in twitter spam detection using ensemble learning," Computers & Security, vol. 69, pp. 35-49, 2017.

[7] C. Chen, et al., "Investigating the deceptive information in Twitter spam," Future Generation Computer Systems, vol. 72, pp. 319-326, 2017.

[8] G. Lin, et al., "Statistical twitter spam detection demystified: performance, stability and scalability," IEEE Access, vol. 5, pp. 11142-11154, 2017.

[9] C. Chen, et al., "Statistical features-based real-time detection of drifted twitter spam," IEEE Transactions on Information Forensics and Security, vol. 12, pp. 914-925, 2016.

[10] A. Singh and S. Batra, "Ensemble based spam detection in social IoT using probabilistic data structures," Future Generation Computer Systems, vol. 81, pp. 359-371, 2018.

[11] C. Li and S. Liu, "A comparative study of the class imbalance problem in Twitter spam detection," Concurrency and Computation: Practice and Experience, vol. 30, p. e4281, 2018.

[12] R. Aswani, et al., "Detection of spammers in twitter marketing: a hybrid approach using social media analytics and bio inspired computing," Information Systems Frontiers, vol. 20, pp. 515-530, 2018.

[13] A. T. Kabakus and R. Kara, "“TwitterSpamDetector”: A Spam Detection Framework for Twitter," International Journal of Knowledge and Systems Science (IJKSS), vol. 10, pp. 1-14, 2019.

[14] X. Wang, et al., "Drifted Twitter Spam Classification Using Multiscale Detection Test on KL Divergence," IEEE Access, vol. 7, pp. 108384-108394, 2019.

[15] B., Mukunthan. (2019). Improved Content Based Medical Image Retrieval using PCA with SURF Features. International Journal of Innovative Technology and Exploring Engineering. 8. 10.35940/ijitee.J1020.08810S19.

[16] M.Arunkrishna, B.Mukunthan “ Review on Classification of Anti-Spam Solutions : Approaches, Algorithms Demystified.” Studies in Indian Place Names Vol. 40 No. 60 (2020): Vol-40-Issue-60-March-2020 , vol. 40, no. 60, 6 Mar. 2020, pp. 4449–4458.

[17] Mukunthan B, Nagaveni N. Identification of unique repeated patterns, location of mutation in DNA finger printing using artificial intelligence technique. Int J Bioinform Res Appl. 2014;10(2):157-176. doi:10.1504/IJBRA.2014.059516

[18] A, Pushpalatha & B, Mukunthan. (2010). Automation of DNA Finger Printing for Precise Pattern Identification using Neural-fuzzy Mapping approach. International Journal of Computer Applications. 12. 10.5120/1761-2411.

[19] V. Vishwarupe, et al., "Intelligent Twitter spam detection: a hybrid approach," in Smart Trends in Systems, Security and Sustainability, ed: Springer, 2018, pp. 189-197.

[20] C.C. Wei and N.S. Hsu,"Derived operating rules for a reservoir operation system: Comparison of decision trees, neural decision trees and fuzzy decision trees,"Water resources research,vol.44, 2008.

[21] Mukunthan, B. & Nagaveni, N.Nagaveni. (2011). "Automating Identification of Unique Patterns, Mutation in Human DNA using Artificial Intelligence Technique". International Journal of Computer Applications. 25. 26-34. 10.5120/3003-4038.

[22] H. Tajalizadeh and R. Boostani, "A novel stream clustering framework for spam detection in twitter," IEEE Transactions on Computational Social Systems, vol. 6, pp. 525-534, 2019.

[23] B., Mukunthan. (2019). Improved Content Based Medical Image Retrieval using PCA with SURF Features. International Journal of Innovative Technology and Exploring Engineering. 8. 10.35940/ijitee.J1020.08810S19.