Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

Efficient Spam Detection on X (formerly Twitter) : A Hybrid Artificial Neural Network and Fuzzy Decision Tree Approach

[1]M. Arunkrishna, [2]Dr. B. Senthilkumaran,

[1]Research Scholar,

PG & Research Department of Computer Science,

Christhu RajCollege (Affiliated to Bharathidhasan University),

Tiruchirappalli, Tamilnadu, India. arunkrishna.murugan@gmail.com

[2]Research Advisor,

PG & Research Department of Computer Science,

Christhu RajCollege (Affiliated to Bharathidhasan University),

Tiruchirappalli, Tamilnadu, India. skumaran.gac16@gmail.com

Abstract

These days, there are a plethora of online social media sites that bring people together, such as Instagram, X (Twitter), and Facebook. The abundance of user-generated content on X Platform has made it a leading social media platform. Users are able to connect with one another, share what they're up to, and discover new pals. Twitter detects spam URLs and blocks them using Google Safe-browsing. X (Twitter) attracts several types of spammers since it has a sophisticated API that allows users to read and publish data. Many previous studies have used different machine learning algorithms to identify spam on X. On the other hand, their methods have not been well tested, and they have shown to be inaccurate when applied to huge datasets. A hybrid approach is created by integrating Artificial Neural Networks along with Fuzzy Decision Trees was proposed as a solution to these problems in this study. According to the labels, the suggested classifier distinguished between span and non-span tweets. A massive dataset consisting of 600 million public tweets was used to test the suggested classifier. To assess the new calculation's presentation, metrics such as accuracy, F-measure, TPR and FPR are employed. The results shows that the our novel strategy is more reliable and powerful.

Keywords: *Online Social Media Networks,X(Twitter),Spam Detection,Hybrid Spam Detection Approach,Spam vs Non-Spam Tweets,*

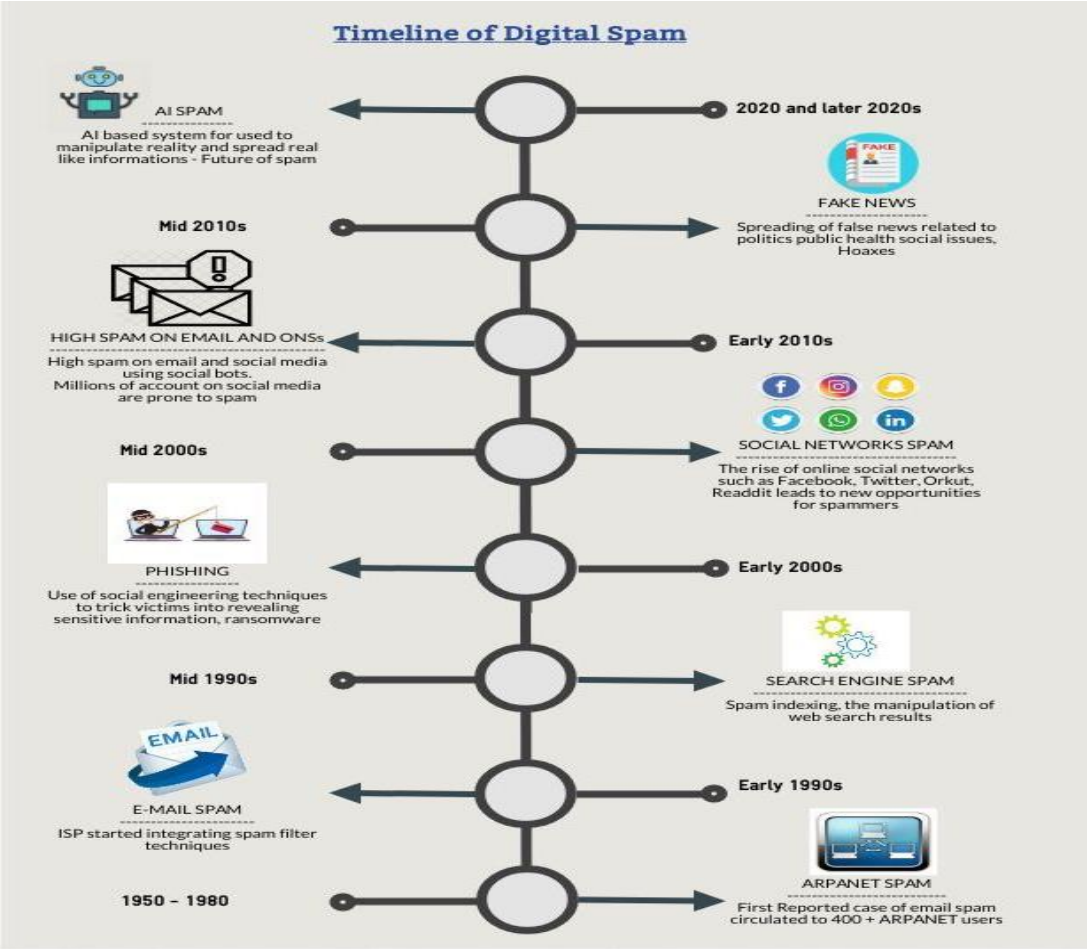
1. Introduction

Online Social Networks (OSNs), characterized by informal communities, have become a critical platform for communication in the digital age. These platforms enable vast numbers of individuals to connect and collaborate across geographical boundaries [1]. Among these, X(Twitter) has emerged as a leading social media platform due to its unique features. It allows users to engage in microblogging with concise messages, typically limited to 140 characters originally. Additionally, X facilitates following individuals of interest, including celebrities and public figures, all at no cost. This accessibility and ease of use contribute to its global appeal, accessible from various devices like smartphones, tablets, and computers.

However, the very features that make X platform attractive to a broad audience also present vulnerabilities. The platform's open nature can attract malicious actors, including cybercriminals and unauthorized users like

spammers. These actors can leverage X platform to launch various attacks, such as disseminating scams, phishing attempts, and unsolicited commercial messages (spam)[23]. Figure 1 provides examples of such

Figure1: suspicious online behavior timeline



suspicious online behavior timeline. Spam pervades online platforms, including social media. It manifests in various forms, echoing the tactics used in email, text messages, and web attacks. This includes unsolicited surveys, misleading messages, artificial likes, follower inflation (Sybils), and page postings by suspended spam accounts [2]. These activities disrupt the user experience and damage the platform's integrity.

Supervised learning, a common approach for combating spam, requires a substantial amount of labeled data for effective training [3]. This data labeling process, crucial for tasks like object detection and document classification, is often time-consuming and expensive. It necessitates human expertise or dedicated experimentation, hindering development speed.

Deep learning, a more advanced form of supervised learning, offers a potential solution. By leveraging powerful algorithms and vast datasets, deep learning models can potentially automate spam detection with greater accuracy and efficiency.

Decision tree classifiers estimate probabilities by assigning class labels to the leaves (terminal nodes) of the tree [4]. This class distribution is often used as a selection criterion for self-training algorithms. However, our research indicates that this approach has limitations. While the class distribution enables predictions, it doesn't improve the overall classification performance of the system or effectively utilize unlabeled data for training. In essence, the self-training algorithm fails to leverage the potential of unlabeled data due to the limitations of class distribution in ranking predictions. In this paper, we propose a new solution to this problem by integrating a deep learning method with a decision tree classifier

1.1. Objectives

- Implement the Innovative ANN-FDT Classifier for Spam Tweet Detection
- Minimize False Positive Rate
- Achieve High Accuracy in Classifying Spam and Non-Spam Tweets

2. Related Works

[5] In this regard, the study recommended the use of a semi supervised spam detection framework known as S3D in the social media network of X (Twitter). The suggested system had two modules: one for real-time spam detection and another for batch model updates. Four new, lightweight spam detectors are included in the module. Tweets that included URLs or connections that were boycotted were set apart by the boycotted area finder, and tweets that were believed to be duplicates of the ones that were at that point checked were set apart by the close to copy locator. First, the trustworthy, dependable users' tweets were tagged using the dependable nonspam (ham) detector to ensure they did not include any spammy or undesired terms. Then, detectors based on multi classifiers were used to classify the rest of the tweets. The data that is crucial for the detection module is updated depending on the tweets that were tagged in the preceding time frame. Experiments conducted on larger datasets show that the suggested system can learn new patterns of unwanted activities and maintain improved accuracy in identifying spam in tweet streams.

[6] Experimental results showed that the spam detection rate is affected by an unequal distribution of spam and ham (non-spam) categories. To take care of this issue of inconsistent conveyance, we introduced a fluffy rationale based oversampling approach that, as indicated by the possibility of fluffy based deterioration of data, may deliver manufactured information tests from moderately little restricted examples. That, yet we additionally made a gathering learning technique that can fabricate more precise classifiers in just three phases, even with uneven information. The underlying stage was to utilize a few methodologies, including as FOS, irregular undersampling, and irregular oversampling, to change the class dissemination inside the unequal information. The following stage included building a model for order utilizing every one of the reallocated informational collections freely. Finally, in order to merge the predictions from each of the classification models, a well-known vote system was utilized. Information gathered from late tweets on Twitter was utilized for the examination. The exploratory outcomes show that the learning technique depicted in the paper may essentially further develop the spam location rate in datasets with lopsided dissemination.

Our examination of a massive dataset containing roughly 500 million tweets identified that approximately 6% consisted of unwanted spam content [7]. This analysis revealed a concerning trend: a significant portion of X (Twitter) spam incorporates misleading information. This deceptive content serves various purposes, often attempting to lure victims to malicious websites. There was also acknowledgment of periodicity in the distribution of the X (Twitter) spam with the affected areas having a higher response rate to the malicious content. First of all, it must be noted that some of the conclusions made in the framework of the present research have positively affected the efficiency of several kinds of spam identification.

[8] Our aim was to identify the best classical machine learning algorithm for real time identification of tweeter spams. A lot of ground truth data was in abundance hence it was easy to test the stability of the devised algorithm. We also ensured this way that these solutions were capable of being upscaled to better their performance in the real sense. When it comes to the aspects of evaluation, in the course of the consideration we used several parameters – F-measure, detection accuracy, TPR, and FPR. Thus, the training samples of different sizes selected randomly were used to investigate the stability of the algorithms. In addition, the paper's "versatility" which described the fact that preparation and tests for various AI approaches diminish as the amount and distribution dependable on parallel computing grow, was helpful to assess the value of such an environment.

[9] That is why, solving the problem of the X (Formerly Twitter) Spam Drift, statistical consideration of about one million of spam and the same amount of non-spam tweets were made. Subsequently, the study report introduced a new Lfun approach to the audience. If the employed methods would be applied on the unlabeled

tweets, the updated spam tweets could be discovered. The next process was to submit the found tweets to the training phase of the classifier. Many experiment runs were performed to compare the efficiency of the proposed method. In certifiable situations the exploratory outcomes exhibit the way that the Lfun adaptation enhance the spam identification precision.

[10] Self-training approach using multiple classifiers, that incorporated four different classifiers was proposed as a result of the literature review based on the subject of X (Twitter) spam. The OFQ Filter is one of the numerous Probabilistic Data Structures (PDS) applied in this approach. OFQ Filter enables the performance of queries on databases storing URLs, spam details, unwanted users, and LSH. LSH, in return, can use the Extend Similarity Searching methods in attempt to detect potential spam and classify it at different stages seeking for nearly instant results with low computational complexity. A comparative analysis of PDS with other structures containinglike data was also done in order to assess the effectiveness of the suggested technique. The performance of the system was measures using standard parameters of a recommendation system such as recall, precision, and F-score.

[11] The investigated problem area is classified as class imbalance in the context of Twitter spam detection. Imbalance of classes is a serious problem in the case of a large number of samples of one class (say, spam) as compared to the other (say, non-spam). In order to solve this problem, the authors carried out the comparative analysis of several known techniques for processing the datasets with the imbalance. Then, they compared the performance of these methods with standard approaches typically utilized for identifying spams on X (Twitter). Comparing the results with ground truth of Twitter data, it is found that, by using ensemble learning along with well-designed technique, the classification performance can be improved significantly.

[12] recommends integrating bicoastal computing and social media analysis as a way of identifying the spamming Twitter accounts. The method expands existing algorithms that innegrate the K-means clustering with LFFA as well as a modified version of the FA with chaotic maps. To summarize, the proposed frameworks were applied to a dataset of 14,235 Twitter accounts, which collected 18,44,701 tweets. The data classification was based on 13 selected variables which was statistically significant and obtained from the results of social media analysis. Also, to identify users that shared similarities with the spammer as well as the non-spammer class, K-Means Fuzzy clustering was applied. There were six experiments with six different forms of FA; all of them used both K-means and Chaotic maps with Levy flights. The results have demonstrated that with help of the Gauss map method, the suggested chaotic FA converges to the optimal results much faster.

[13] proposed an interesting procedure for spam detection on the trial on the Twitter platform. The approach proposed in this paper is based on the assumption that Twitter is different from other platforms, therefore, traditional solutions aimed at spam detection do not work on it. To address this problem the recommended solution builds on Twitter’s strengths to detect spam on the platform. To gather the data we have employed the Twitter Application Programming Interface (API) to pull 77,033 tweets from 50,490 users on a wide range of topics.

Naïve Bayes was used to train the Twitter Spam Detector algorithm and this made it possible to have different categories for spammers and the legal users. Concerning the results of the assessment procedure, it is crucial to note that sensitivity equated to 0. 913 and accuracy value was zero. 943.The reported study used K-L divergence to report the spread of the spams and the Multi-scale Drift Detection Test (MDDT) to flag possible drifts [14].

Thus, using the detection findings for the purpose of re-training the basic classifier, enhances its performance. Thus, based on the experimental results indicated in this paper, it can be seen that the K-L divergence approach will show a continually varying feature pattern whenever such a drift occurs. The performances attained through the classifications were further recognized to be more efficient and enhanced by the use of performance metrics such as F-measure, recall, and accuracy.

As a mixture of DenStream and INB which stands for Incremental Naïve Bayes, the INB-DenStream was proposed in [22].We demonstrated INB-DenStream's efficiency by evaluating its performance on a variety of metrics, including computational complexity, parameter sensitivity, purity, general recall, and general accuracy. We compared the suggested method's performance to that of other popular alternatives, including CluStream,

DenStream, and StreamKM ++. The proposed strategy performs better than the other well-known approaches, according to the comparative results.

[16] Outline a straightforward method for a system that can identify and prevent spam. Both classical and mathematical methods may be used to anti-spam systems. Methods that fall under the category of "traditional approaches" include content screening and blacklisting. Methods based on mathematics are incorporating statistical ML and AI. A number of algorithms, collectively referred to as filters, have been developed to achieve this goal. It is possible for the filters to learn how to be more precise. The algorithm has to be given some basic instructions before it can learn from its errors. Analyzing, cleaning, visualizing, implementing, rating, and studying data are all part of the standard procedure.

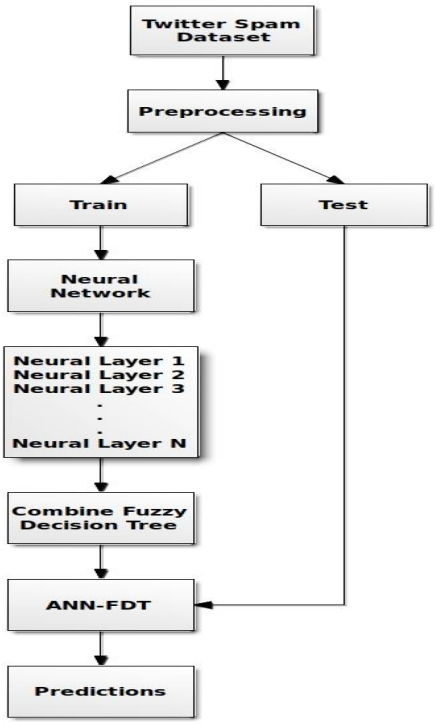
[19] Building a Clever Twitter Spam Location Framework that can give exceptionally exact subtleties on spam profiles is the way Twitter spam is distinguished. A single hybrid classifier checks the links using the API in Google Safe Browsing to provide further security, and the framework considers particular capabilities prior to assessing the gathered tweets. This led to better categorization of the collected tweets and smarter spam identification from the suggested system.

A approach for identifying distinct patterns using AI was given in [17][18]. The device is equipped with a neural fuzzy pattern recognition processor. Contains three generators: one for normalization and one for artificial neural network (ANN) learning. The match capability generator utilizes the watchfulness boundary to pick the best not really for yield creation, while the enactment capability generator initiates the result hub as indicated by the related information loads.

[23] We might break down the information with the arbitrary element, multi-facet perceptron on client object correspondence capability, and the fake brain BSF channel replaces the code plan. This framework is presented to recover the hidden capacity of the supplied issue. To ensure that the suggested ANN performs adequately. Prior to Matrix factorization, the hidden data and ANSF need to be trained. To explore the interactivity feature and for flexible factorization, CMfact MPeep hybrids are used.

[15] One method for reducing dimensionality in fuzzy image collections is Principal Component Analysis (PCA), which is non-parametric and unsupervised. It can be valuable when dealing with large volumes of fuzzy sets, as it aims to identify the underlying structure and reduce the data to its most significant components.The suggested machine learning method finds a great use case in identifying similarities between collections of fuzzy images by leveraging PCA for dimensionality reduction.

Figure 2: The Proposed Flow



3. Proposed Work

To sort tweets into spam and non-spam classes, an enormous number of tweets are gathered from Twitter. A pre-processing step is taken after data extraction from the Twitter Spam Dataset. Following the pre-handling step, the tweets go through the preparation and testing stages. The learned tweets are handled by the neural networks when data training is finished. The neural networks are constructed by combining n different neural networks. A fuzzy decision tree classifier is then applied to the resulting multi-layer model. After that, ANN-FDT is suggested as a way to combine the classifier with ANNs. The suggested ANN-FDT classifier processes the tested tweets. We then make educated guesses as to whether or not the collected tweets constitute spam. Pictured in Figure 2 is the flow.

3.1 Proposed ANN-FDT Algorithm

This paper proposes an ensemble model for identifying spam tweets. The model combines two powerful techniques: Artificial Neural Networks (ANNs) and Fuzzy Decision Trees (FDTs).

The Artificial Neural Network (ANN) Component:

The ANN component employs a multi-layered architecture with input,hidden, and output layers.Words from the tweets serve as the input data for the network.The network utilizes various activation functions, either individually or in combination, to process this information.

The Fuzzy Decision Tree (FDT) Component:

The FDT component leverages the same training dataset as the ANN, but the data is first normalized to a range between 0 and 1.This approach utilizes robust fuzzy sets to create more versatile membership functions.Each attribute in the dataset is assigned a value based on a fuzzy membership function within the FDT method.

Strengths of the Hybrid Approach:

Combining these two techniques (ANN and FDT) leads to a more effective and accurate method for spam tweet identification compared to individual approaches.

The pseudo-code below describes the proposed Hybrid Artificial Neural Network - Fuzzy Decision Tree algorithm.

1. Calculate the maximum depth of a branch (Bd).
2. The number of branches at every level is counted as numb 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 depth B_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 $h(l)=\text{h}(l-1)+\text{num}_b(l)$ arrays are created and the dimensions of the created arrays are $h(l)$ x $h(l-1)$,l=1,…,B_d.
6. The initial weights wei^{B_d+1} are stored with the dimension of $h(B_d)$ x $\text{out}_{\text{Ne}}(\text{Class}_N)$. The arrays are initialized to 0.
7. For each input a=0,1,…,$\text{inp}_{\text{Ne}}-1$:
8. Set $\text{wei}_{\text{a,a}}^l \rightarrow 1$ for l<$\max_a^L$

3.2 Challenges and Fuzzy-based Improvements in Decision Trees

A common challenge with decision trees is that after processing input data through the hidden layers, the branches can grow excessively long, potentially hindering performance. The proposed approach addresses this issue by leveraging fuzzy logic within the decision tree framework. While the overall construction process remains similar to traditional decision trees, the calculation of information gain (denoted as G_gain) differs in fuzzy ID3 algorithms.

3.3 Fuzzy Information Gain Calculation:

Specifically, fuzzy ID3 employs membership functions associated with fuzzy sets to calculate information gain. The suggested approach in [20] utilizes fuzzy sets and the intersection operator (denoted as ∩) along with the arithmetic product operator (denoted as *) to compute information gain.

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

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

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

In the set of equations represented by Hu, Hv, and EAG, The membership functions for attributes a and b are represented as Oab. Here, the value of attribute b varies in proportion to attribute a. Here, Q stands for a collection of pairs (a, b) and the branches that go from the root to node j. he calculations performed in this method closely resemble those utilized in the original ID3 algorithm. The set of leaf nodes is denoted by L_node, with individual leaf nodes represented by the symbol l. The following formula estimates the posterior probability of a given sample belonging to class t

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

The conditional probability, denoted by p_t^l, represents the likelihood of class t occurring at a specific leaf node l within a decision tree framework. This probability is determined by a multidimensional membership function, Q_l, which captures the relative contribution of each attribute along the path from the root node to l. The specific form of Q_l depends on the chosen decision tree algorithm and the underlying data distribution

4. Evaluation of Results

A few presentation measurements, including exactness, TPR, FPR, and F-measure, were point by point in this part. The proposed system was tried on a huge dataset to get exploratory discoveries [18]. From 600 million tweets, we recovered the URLs in general. We can determine whether the URL is malicious by using the WSR service. Our analysis of this massive dataset revealed 6.5 million malevolent tweets, or around 1% of all tweets.4.1 KPIs for Performance

4.1 Performance metrics

4.1.1. Precision

Accuracy, the inverse of precision, is a standard value measurement that is also known as the weight arithmetic mean. The recipe displayed underneath might be utilized to decide the exactness.

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

4.1.2. TPR or True Positive Rate

An examination's True Positive Rate (TPR) is the sum of all the significant positive discoveries made for every accessible positive case.

4.1.3. FDR or False Positive Rate

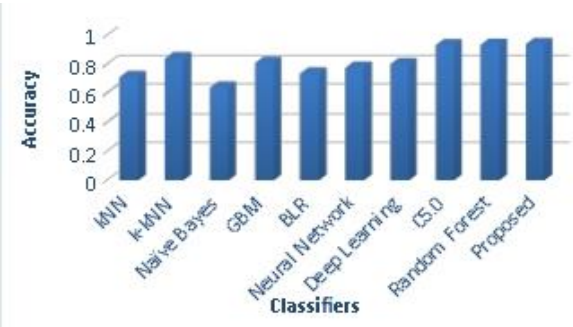
Like before, the False Positive Rate (FPR) is calculated by adding up all the incorrect positive results for every available negative sample.

4.1.4 F-measure

F1 scoring is a technique that strikes a balance between accuracy and recall. The following formula may be used to calculate this: F1-Score= (2*Precision*Recall) / (Precision + Recall).

5. Results

The effectiveness of the proposed approach was assessed by calculating performance metrics such as precision, F-measure, TPR (True Positive Rate), and FPR (False Positive Rate). Subsequently, the results were benchmarked against those obtained from established classification algorithms to demonstrate the method's competitiveness. The levels of accuracy, particularly for datasets, are shown in the figure below (Fig. 3).



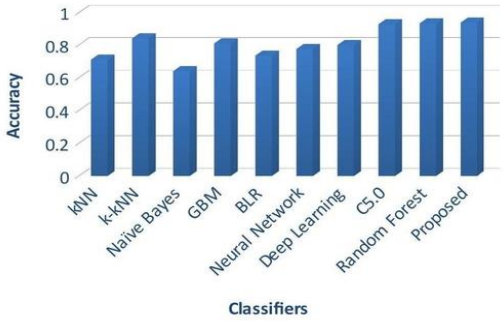


Figure3: Accuracy for Dataset 1 **Figure 4: Accuracy for Dataset 2**

This study compares the proposed method to several existing methodologies, evaluating their accuracy in achieving the desired task. The proposed method achieves the highest accuracy among the compared methods, reaching a score of 0.96. Fig. 4 displays the accuracy values, with dataset 2 being the exception. We check the suggested work for correctness and compare it to other methods that are already out there. The comparison clearly shows that the given strategy achieves the maximum accuracy of 0.95.

As shown in Figure 5, the proposed method achieves demonstrably lower False Positive Rates (FPR) on both datasets 1 and 2 compared to existing techniques. The reductions in FPR are 6 and 5 for datasets 1 and 2, respectively.

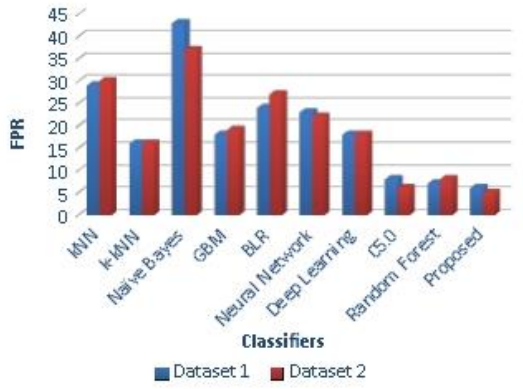


Figure5: False Positive Rates (FPR) for Datasets 1&2

Figure 6 displays the True Positive Rates (TPR) for datasets 1 and 2, a commonly used performance statistic. We compare the TPR of the suggested method to that of well-established methods. Figure 1 shows that the suggested strategy obtains TPR values of 93 and 94 for datasets 1 and 2, respectively, which are superior.

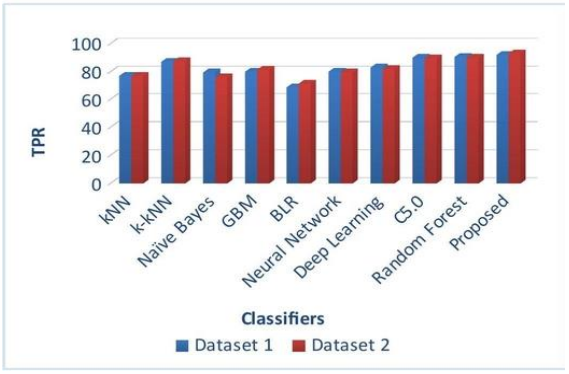


Figure6: True Positive Rates (TPR) for Datasets 1&2

Figure 7 displays the results of the F-measure for both the first and second dataset. In comparison to previous methods, we assess the suggested method's F-measure. While the proposed method achieves F-measure scores of 88.57 and 93.90 for datasets 1 and 2, respectively, it's important to consider whether these represent improvements or drawbacks depending on the specific task and desired balance between precision and recall

Figure7: F-measure for Datasets 1&2

6. Conclusion

The rise of social media platforms has revolutionized online communication, connecting people across the globe. However, this open environment also attracts malicious actors like spammers. This paper proposes the Innovative hybrid classification method which combines the strengths of Artificial Neural Networks (ANNs) and Fuzzy Decision Trees (FDTs) to effectively categorize tweets as spam or legitimate content. We assess the efficacy of this approach by utilizing well-established criteria including accuracy, True Positive Rate (TPR), False Positive Rate (FPR), and F-measure. Our analysis demonstrates that the Hybrid Artificial Neural Network and Fuzzy Decision Tree approach significantly improves performance in combating online social network spam.

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.

[24] SocketLabs. (2016, October 21). A Brief History of Email Spam | Infographic. SocketLabs. https://www.socketlabs.com/blog/a-brief-history-of-spam-infographic/

[25] Zhouxiang Fang, et al.,"How to generate popular post headlines on social media?".AI Open,Volume 5,pp. 1-9,2024.