INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

URL Redirection Attack Mitigation in Social Communication Platform using Data Imbalance Aware Machine Learning Algorithm

Sagargouda S Patil[1]*, **H A Dinesha**[2]*

1 Assistant Professor, Computer Science, Jain College of Engineering and Research, Belagavi, Karnataka, India
2 Professor& HOD, Computer Science, Nagarjuna College of Engineering and Technology, Bangalore, Karnataka, India

OPEN ACCESS

Received: 29.09.2021
Accepted: 10.02.2022
Published: 22.03.2022

Citation: Patil SS, Dinesha HA (2022) URL Redirection Attack Mitigation in Social Communication Platform using Data Imbalance Aware Machine Learning Algorithm. Indian Journal of Science and Technology 15(11): 481-488. https://doi.org/10.17485/IJST/v15i11.1813

* **Corresponding authors**.

sagar.cs.kle@gmail.com

sridini@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To present a model which can detect malicious attacks using the URL of the Social Communication Platform using the data imbalance machine learning algorithm. The main objective is to detect the attack and prevent it from happening. **Methods:** This study presents an efficient feature extraction and selection method addressing feature imbalance problems; and also presents an improved concept drift and machine learning-based classification. This paper extracts the URL of the undesired tweets, identifies them, and filters them for classification. **Findings:** The experiments have been conducted using the drifted twitter spam dataset. Our model DIA-XGBoost extracts the URL of the undesired tweets, identifies them, and filters them for classification. Further, the attack pattern varies with respect to time. Furthermore, the results show that our DIA-XGBoost attains higher accuracy performance by 1.254%, URL recall performance by 0.14%, and increased F-measure performance by 10% when compared with the existing ML techniques (Random Forest, K-Nearest Neighbour, XGBoost). Thus, the existing ML-based classification model achieves poor classification accuracy whereas our model solves this issue. **Novelty:** Various Machine Learning (ML) techniques have been applied for the classification of URL redirection attacks. However, the spam data generally exhibit feature imbalance. Further, the attack pattern varies with respect to time. Thus, the existing ML-based classification model achieves poor classification accuracy. Hence, our model solves the issue using the DIA-XGBoost algorithm, detects and prevents URL malicious attacks.

Keywords: Data Imbalance; Feature Extraction; Concept Drift; URL; Machine Learning; URL Redirection Attack

1 Introduction

Online Social platforms like Facebook, Instagram, WhatsApp, and Twitter have changed the way of sharing information between people[1]. Users join social networks usually to connect to their families, friends, or who they are interested in. The information being shared by the users can be of any type. Some of the information shared could be irrelevant to other users. Some of the users share the wrong information to gain the attention of the user. The best example of sharing the wrong information could be edited images and videos, fake news, and other kinds of irrelevant data that is being shared using the URLs in social platforms. To reduce these kinds of fake URLs, this model has been created to identify and classify the following URLs to reduce these kinds of attacks. All the existing models mainly focus on network security and try to identify using the four kinds of attacks: Dos (denial of service), probe, U2R (User to Root), and R2L (remote to local). Generally, the users are unaware that the attacker has been attacked and the information has been leaked leading to many frauds. Many models have been created to stop these kinds of attacks before happening. Many machine learning algorithms have been used in the model, to predict these attacks. Some of the models have predicted the attacks but the accuracy rate has been compromised. We have created a model that gives a correct identification of the given malicious URL and classifies the URLs into different classifications. The goal of this model is to check the malicious URLs and classify them into different types of attacks.

In [2], has explained how the attackers get the attention of the users and make them open the URL links. The attacker sends misleading information that makes the social network a plot for the media. Some of the unwanted text or data can be sent through these URLs. The paper explains the detection model that is used on social networks with the help of supervised machine learning algorithms like Naïve Bayes and Enhanced Random Forest classification and F1-scoring method to check the accuracy and precision of the model. In [3], the paper has classified spammer and non-spammer content in social networks using the Genetic Algorithm. The detection of the attacker in an online platform is a critical job. In [4], it mainly focuses on the Drifted Social Network malicious URL problems and overcomes this problem using the machine learning algorithm, Adaptive K-Nearest Centroid Neighbor Classifier (AKNCN). The model is trained using the spam and non-spam content that have been taken as the dataset for training the classifier. In [5], discusses the machine learning concepts and concept drift approach for detection of malicious URLs. They have collected the Lexical, host-based, and content-based features from the URL to train the machine learning algorithm. There are multiple machine learning algorithms used in this model such as Random Forests, Gradient Boosted Decision Trees, and Deep Neural Network classifier to check the accuracy and precision. In [6], explains the concept of data drifts and concept drift. They have proposed a model where they can have a solution for both data drifts and concept drifts in IoT-Driven Intrusion detection. They have a solution that stabilizes both of the problems and gives an efficient performance for the intrusion system. In [7], they have an approach for intrusion detection using stream-oriented learning for adaptation of the concept drift. Machine learning algorithm, Adaptive Random Forest classifier has been used to train and detect various intruders in a data stream model. In [8], they have a theory that the real-world data stream has a challenge in the implementation of the machine learning model. This paper proposes a method of active concept drifts detection algorithm for time series analysis in an even environment. In [9], they have explained the challenges faced by the intrusion detection system in IoT which is based on Machine learning. The challenges faced are concept drift, high dimensionality that is a large amount of data to be preprocessed before training and testing the model, and computational complexity. They have concluded that the three problems should be addressed in the neural network model for an intrusion detection system in IoT. The real-world data streams keep changing from time to time [10], known as the concept drift. As the data keeps changing according to time there are many challenges faced by the machine learning algorithms. A Recurrent Adaptive Classifier Ensemble [10] keeps the data of preciously trained models and always trains the model using both the new and existing classifiers. As there are many malware and malicious data that have been increasing over the years [11], the detection model has to be improved using the concept drift and train the model using the real-time data and different malware samples to achieve higher accuracy. The Heterogeneous Dynamic Ensemble Selection Base on Accuracy and Diversity (HDES-AD) [12], has been used to train the real-time datasets in nonstationary environments. In a review on Social Network Platform detection systems [13], the researchers have introduced various defense techniques to overcome the problem of Online Social network services. In [14], this paper discusses the CONFRONT method that is used to detect concept drift in botnets. Many approaches have been proposed for various types of detection of malicious URLs [15], a new system has been proposed to detect the URL where the URL classification model is trained, and using the concept drift detection the machine learning model is trained. In [16], this paper explains how the concept of drifts faces challenges in data imbalance. As the streams evolve, the trained model becomes outdated. To continuously run the model the paper has proposed the solution based on Restricted Boltzmann Machine. A framework called spam drift [17], was introduced to deal with the detection of spammer activities in real-time. This approach uses an unsupervised machine learning approach to detect spamming patterns. A Detection system for spammers and Fake user identification on the social network [18], [19], has been performed where the model detects the fake text, spam based on URL, spam in trending topics, and fake users. The detection is done using various techniques such as user features, content features, structure features, and time features. Various machine learning approaches like Naïve Bayes classifier, Neural network, and Support Vector machine have been used in the spamming technique but are not effective to predict the correct accuracy in high-dimensional data. Social

Media Spam Classification is a tough challenge[20], and using malicious links and with the evolution of time, the filtered model could be deceived, causing loss to both users and the whole network.

2 Data Imbalance Aware XGBoost (DIA-XGB) based malicious URL detection model

This work presents a data imbalance and malicious URL identification and classification model using an improved XGBoost Algorithm. The architecture of data imbalance and malicious URL model using XGBoost Algorithm is shown in Figure 1. This paper presents the Data Imbalance Aware XGBoost (DIA-XGB) Algorithm for the detection of different malicious URLs in the Social Network Platform and reduces the spam more efficiently than the existing model considering the data imbalance. It also presents the spam drift extraction and detection model using KL- divergence for obtaining drift time and changes the values in the XGBoost Classification model to check the efficiency of the model.

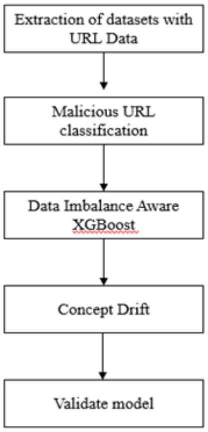
Extraction of datasets with URL Data
Malicious URL classification
Data Imbalance Aware XGBoost
Concept Drift
Validate model

Fig 1. The architecture of Data Imbalance and Concept drift

XGBoost is an enhanced distributed gradient boosting model intended to be exceptionally effective, adaptable, and versatile. It carries out machine learning calculations under the Gradient Boosting structure. XGBoost gives a parallel tree boosting (otherwise called GBDT, GBM) that tackles numerous data science issues quickly and exactly. A similar code runs on major different environments (Hadoop, MPI) and can take care of issues of past billions of models. The idea behind using gradient tree boosting methodologies is to obtain outcomes by cumulating several tree classifiers. Therefore, for classifying the malicious URL, the model trains dataset with o samples with several classifiers as described in below equation

$$\widehat{A}_j = H(Z_j) = \sum_{m=1}^{M} h_m(Z_j), \;\; h_m \in \alpha \tag{1}$$

where Z_j shows the j^{th}data within training data, K depicts tree size used for classifying the malicious URL in the social network dataset, $\widehat{A}_j$ defines the classification outcomes of our multi-label classification model with certain dimensions, k^{th} dimension describes the probability that it will be classified as being belonged to the k^{th} class and α defines set of decision trees as described below

$$\alpha = \left(h(z) = y_{u(z)} \right\} \tag{2}$$

where every tree $h(z)$ agree with respect to leaf weight y and structure parameter u. The objective of the XGBoost classification model is to minimize the loss parameter

$$N(H) = \sum_k n(\widehat{a}_k, a_k) + \sum_m \beta(h_l) \tag{3}$$

where,

$$\beta(h_1) = \delta V + \mu \|y\|^2 \tag{4}$$

The first parameter $n(\widehat{a}_k, a_k)$ in Eq. (3) defines the loss function among actual and classified outcomes. The second parameter $\beta(h_l)$ in Eq. (3) depicts the penalizing term; U depicts leaves size within a tree, δ, and μ depicts the controlling parameter used for controlling computational complexity. In this work weighted loss function is considered for training data z whose ID is described by$n$, the negative log probabilistic loss function is obtained using the following equation

$$n(\widehat{a}_k, a_k) = -\sum_l a(l)\, log\widehat{a}(n) = -log\widehat{a}(n) \tag{5}$$

where $a(l)$ depicts l^{th} dimension of a, $\widehat{a}(n)$ depicts the l^{th} dimension of output $\widehat{a}$. Further, the loss function is optimized iteratively for obtaining minimum loss. Thus, the optimized loss function under certain iteration u can be described using the following equation

$$N^K = \sum_{k=1}^{p} n\left(\widehat{a}_k^{(p-1)} + h_p(z_k), a_k\right) + \beta(h_v) \tag{6}$$

The proposed methodology establishesh_p that can greedily minimize the loss using the following equation

$$N^p \cong \sum_{k=1}^{p} \left(n\left(\widehat{a}_k^{(P-1)} + a_k\right) + i_k h_k(z_k) + \frac{1}{2} j_k h_p^2(z_k)\right] + \beta(h_p) \tag{7}$$

where h_j depicts the first-order gradient of $n\left(\widehat{a}_k^{(P-1)} + a_k\right)$ and j_k depicts the second-order gradient of $n\left(\widehat{a}_k^{(P-1)} + a_k\right)$; thus, the tree h_p can be established by minimizing Eq. (7).

The proposed algorithm Data Imbalance Aware XGBoost helps us to detect the malicious URL efficiently and address the drift problem. The malicious URL change with respect to time. As a result, it affects the performance and efficiency of the existing classifier due to the concept drift problem. Thus, it is mandatory to study the distribution of different types of malicious URLs and consider the different periods. The distribution of these different types and different periods can be described as follows

$$D_{lm}(P(Q) = \sum_j P(j)\, log\frac{P(j)}{Q(j)}. \tag{8}$$

It is used for acquiring a similar analysis of likelihood appropriations. This work maps the information point into dispersion boundary. Consider a multi-set $Y = (y_1, y_2, \ldots, y_o\}$ from a finite set F composed of feature parameter, and represent $O(y(Y)$ the number of appearances of $y \in Y$, thus the relative ratio of each y is represented as follows

$$P_Y = \frac{O(y(Y)}{o}. \tag{9}$$

Thus, the distance among the successive malicious URL, $D_1$and D_2 is computed as follows

$$D(D_1(D_2) = \sum_{y \in F} P_{D_1}(y)\, log\frac{P_{D_1}(y)}{P_{D_2}(y)} \tag{10}$$

This work does the calculation of distance distribution (Eq. (10)) of each element of non-malicious URL and malicious URL taking the calculated value. The higher the value, the more complex the distribution will be and from this the distance distribution estimations, we can acquire the distribution of malicious URL features which is changing quickly according to the time. Further, for non-malicious URL information, this work expects that there are no changes in distribution value. As long as, the distribution of preparing the test values isn't changed/adjusted. Since, the information base that gains from unmodified preparing test isn't altered while being used for characterizing upcoming malicious URL, as a result, the precision of the characterization model will be changed. Resolving issues of gathering altered information to refresh the classifier model is a key factor. By gathering such unlabeled approaching malicious URL and to address this issue, this work presents a model Data Imbalance Aware XGBoost which is made out of two-component, such as to learn from distributed malicious URL and further gain knowledge from the manual data. The calculation to address the malicious problem issue is introduced in Algorithm 1.

The Algorithm takes malicious URL $(\alpha_1, \ldots, \alpha_2\}$ unclassified malicious URLsU_{ulbl}, a binary DIAX classifier model β as input and obtain an output of manually labeled chosen malicious URLu_n. In step 2, we initialize labeled training malicious URL data U_{lbl}. In step 3, using β we construct a classification model C fromU_{lbl}. In step 4, the unlabeled data U_{ulbl} is classified as malicious URL data$U_{S'}$ and non-malicious URL data$U_{S''}$. In step 5, the malicious URL$U_{S'}$ classified by classification model C are grouped into labeled data U_{lbl}. In step 6, utilizing U_R the classification model C is further retrained. In step 7, we establish the freshly coming URLs for the selection process. In steps 8 to 12, we obtain URLs that meet selection condition T. In steps 13 to 17, we manually label or classify each URL datav_j. In this way, the training data is updated. Therefore, the aid in addressing concept drift problems and attaining better accuracy performance is experimentally proven in the below section.

3 Results and Discussion

Here the performance of the Data Imbalance Aware XGBoost (DIA-XGB) model is compared with the other existing models[19],[21]. Different evaluation measures and classification performance for the model have been compared with the required different existing models. The experiment of this model has been performed on Intel Pentium I-7 Class Processor that was composed of 8 GB of RAM and all the Data Imbalance Aware XGBoost (DIA-XGB) and other existing algorithms have been implemented using Python 3 framework. All the existing and proposed classification algorithms have been implemented using the sci-kit-learn package.

- **Dataset Description**

We have used a well-known and widely used social network malicious URL dataset[19],[21]. In the Social network, the URL only is used for evaluation. The Dataset comprises 12 features sets that are described in Table I. A total of 10,000 malicious URLs of data is collected per day and the data similarly has been collected for 10 days. According to the real-world scenarios, only 5% of data is discarded as it is of no use. Table 1 shows the Dataset that has been considered for experimental analysis.

Table 1. Dataset Considered for Experimental Analysis

Feature Number	Feature Named
F1	Account Age
F2	No_follower
F3	No_following
F4	No_userfavourites
F5	No_lists
F6	No_tweets
F7	No_retweets
F8	No_hashtags
F9	No_usermention
F10	No_URLs
F11	No_char
F12	No_digits

- **Model Performance**

Here the performance of the DIAXGB and existing systems such as KNN, RF, XGB, and DIA-XGB is evaluated. The experiments are conducted for comparing imbalance performance and drift detection of the different methods. The performance is evaluated in terms of accuracy, recall, and F-measure. The ROC confusion matrix is shown in Table 2.

Table 2. ROC Confusion Matrix

	Malicious URL	Normal URL
Predicted Malicious URL	True Positive (TP)	False Positive (FP)
Predicted Normal URL	False Negative (FN)	True Negative (TN)

The accuracy performance is calculated as follows

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{11}$$

The recall performance is calculated as follows

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

The F-measure performance is calculated as follows

$$F-measure = \frac{2*Precision*Recall}{Precision*Recall} \tag{13}$$

- **Malicious URL Accuracy Performance**

This section checks the accuracy of the proposed model with the other existing models. The accuracy of the model is calculated using Eq. (11). The accuracy achieved by the DIA-XGB algorithm over the existing system such as KNN, RF, and XGBoost[19]is shown in Figure 2. From the figure, it can be seen that the model DIA-XGB gives a better accuracy rate than the other algorithm used in the existing model. The KNN shows an accuracy of 96.78% while the RF algorithm shows an accuracy of 98.43%. It can be seen that the existing XGBoost model is giving 98.85% while the DIA-XGB provides an accuracy of 99.684%. Thus, the model DIA-XGB gives us more accuracy than the existing models.

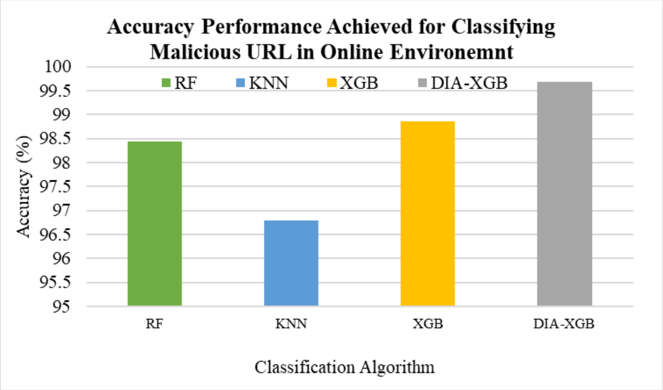

Fig 2. Malicious URL Accuracy Performance

- **Malicious URL Recall Performance**

This section checks the recall performance of the proposed model with the other existing models[19]. The recall performance of the model is calculated using Eq. (13). The recall performance achieved by the DIA-XGB algorithm over the existing system such as KNN, RF, and XGBoost[19] is shown in Figure 3. From the Figure, it can be seen that the RF algorithm gives a recall of 0.69 and the KNN algorithm gives a recall of 0.61 which is comparatively less. The DIA-XGB gives a recall of 0.93 whereas the existing XGB gives a recall of 0.79. The DIA-XGB gives a 0.14 better recall than the existing model of XGBoost. This shows that the DIA-XGB has a good recall performance as compared with the existing system.

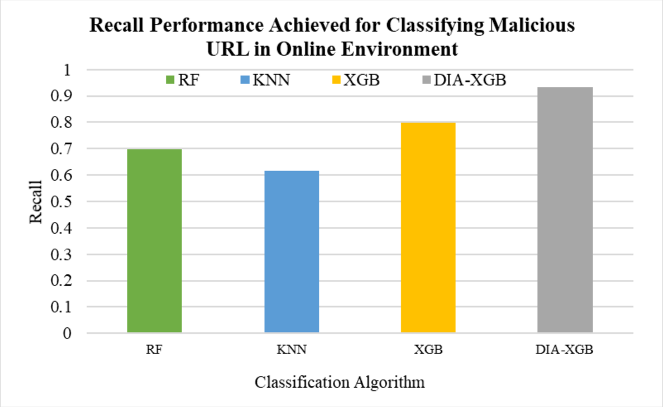

Fig 3. Malicious URL Recall Performance

- **Malicious URL F-Measure Performance**

This section checks the F-measure performance of the proposed model with the other existing models[19]. The F-measure performance of the model is calculated using Eq. (14). The F-measure performance achieved by the DIA-XGB algorithm over the existing system such as KNN, RF, and XGBoost[19] is shown in Figure 4. The RF algorithm shows an F-measure of 0.80 and the KNN show an f-measure which is relatively low, that is of 0.65. The XGBoost shows an f-measure of 0.86 and the DIA-XGB shows an f-measure of 0.96. The DIA-XGB shows a 10% increase to the existing XGBoost.

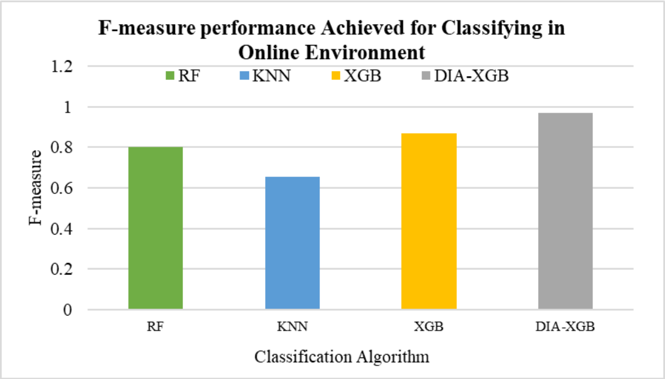

Fig 4. Malicious URL F-measure Performance

4 Result and discussion

Table 3. Comparative analysis of DIA-XGB over existing methodologies

	LFUN-RF[21] 20 16	AKNCN[3], 2019	MDDT,[19], 2 020	DIA-XGB
Address class imbalance	No	No	No	Yes
Address concept drift	Yes	Yes	Yes	Yes
Accuracy	80%	91.95%	98.85%	99.68%
F-measure	84%	89.7%	86.7%	96.67%
Recall	88%	90.05%	79.77%	93.425%

The comparative analysis of the proposed DIA-XGB and the existing model is shown in Table 3. The following existing methodology is chosen as they all worked using the dataset. The Lfun-Random Forest model[21] attained an accuracy, F-measure, and recall performance of 80%, 84%, and 88%, respectively. The AKNCN model[3] attained an accuracy, F-measure, and recall performance of 91.95%, 89.7%, and 90.05%, respectively. The MDDT-XBG[19] accuracy, F-measure, and recall performance of 98.85%, 86.7%, and 79.77%, respectively. The DIA-XBG accuracy, F-measure, and recall performance of 99.68%, 96.67%, and 93.42%, respectively. The existing methodologies[3,19,21], addressed either class imbalance or concept drift individually; thus, poor F-measure and recall performance is achieved. The significant result achieved considering different metrics is attributed to addressing class imbalance and concept drift together using DIA-XGB. Our model can be adapted to detect attacks where high classification accuracy with minimal false detection where the existing model predominantly fails.

This section of results and analysis provides better comparison of the existing system when compared with the DIA-XGBoost model. The DIA-XGBoost can address both the class imbalance and concept drift problems faced by the existing systems using the machine learning algorithm which has been presented in the above section. Further more when compared with the existing methodologies, all the methods have less accuracy, recall and f-measure when compared with the proposed DIA-XGBoost model. The proposed model has attained better results in less time and detects and classifies the incoming attack coming by the URL. Hence our model is more efficient than the existing models.

5 Conclusion

This study first identified the problem of malicious URLs in Social Networking platforms. This paper presents an improved version of the XGBoost Algorithm that is the Data Imbalance Aware XGBoost (DIA-XGB). The proposed model is similar to the XGBoost but has an extended version where the weights of the model can be taken as negative also. This helps to maintain the positive weights along with the negative weights but with the positive values. The experimentation on the datasets has been performed by considering the real-time malicious URL attacks in the Social Networking Platform. The results show that the

DIA-XGB Algorithm achieves a higher accuracy rate than the existing machine learning-based classification algorithms with fewer false negative and false positive values in the confusion matrix. The results show that our DIA-XGB attains higher accuracy performance by 1.254%, URL recall performance by 0.14%, and increased F-measure performance by 10% when compared with the existing ML techniques (RF, KNN, XGB). The DIA-XGB performed well in the recall performance as compared to the existing models where there was an improvement of correctly predicted positive observations to all observations in the actual experimental data. It showed a 0.14 higher than the existing XGBoost model. Similarly, the DIA-XGBoost showed a higher F-measure performance by an increase of 10% to the existing model. The overall result shows that the DIA-XGBoost Algorithm has a higher performance in terms of accuracy, recall, and F-measure in a real-world situation. As the existing models cannot detect the sudden attack our model, DIA-XGBoost will identify the attack and tell the user that he is being attacked on the given URL.

Future work would be to present another algorithm for the multiple URL detection and train the model on complicated malicious URLs and also improve its efficiency in classification and detection and improve the concept drift problems in the detection of malicious URLs in Online Social Networking Platform.

References

1) Mukunthana B, Arunkrishnab M. Detection of Malicious Data in Twitter Using Machine Learning Approaches. *Turkish Journal of Computer and Mathematics Education*. 2021;12(3):4951–4958. Available from: https://doi.org/10.17762/turcomat.v12i3.2008.
2) Sahoo SR, Gupta BB. *Classification of spammer and nonspammer content in online social network using genetic algorithm-based feature selection Enterprise Information Systems*;2020:1–27. Available from: https://doi.org/10.1080/17517575.2020.1712742.
3) Lalitha LA, Hulipalled VR. Adaptive k-Nearest Centroid Neighbor Classifier for Detecting Drifted Twitter Spam. *International Journal of Engineering and Advanced Technology (IJEAT)*. 2019;8:235–243. Available from: https://www.ijeat.org/wp-content/uploads/papers/v8i5S/E10480585S19.pdf.
4) Singhal S, Chawla U, Shorey R. Machine Learning & Concept Drift based Approach for Malicious Website Detection. *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. 2020. doi:10.1109/COMSNETS48256.2020.9027485.
5) Wahab OA. Sustaining the Effectiveness of IoT-Driven Intrusion Detection over Time: Defeating Concept and Data Drifts. *TechRxiv IEEE*. 2021;p. 1–10. doi:10.36227/techrxiv.13669199.
6) Setha S, Singha G. Kuljit Kaur Chahala. Drift-based approach for evolving data stream classification in Intrusion detection system. *Workshop on Computer Networks & Communications*. 2021;p. 23–30. Available from: http://ceur-ws.org/Vol-2889/PAPER_03.pdf.
7) Mehmood H, Kostakos P, Cortes M, Anagnostopoulos T, Pirttikangas S, Gilman E. Concept Drift Adaptation Techniques in Distributed Environment for Real-World Data Streams. *Smart Cities*. 2021;4(1):349–371. Available from: https://dx.doi.org/10.3390/smartcities4010021.
8) Adnan A, Muhammed A, Ghani AAA. Azizol Abdullahand Fahrul Hakim. An Intrusion Detection System System for the Internet of Things Based on. *Machine Learning: Reviews and Challenges Symmetry*. 2021;p. 1–13. Available from: https://doi.org/10.3390/sym13061011.
9) Museba T, Nelwamondo F, Ouahada K, Akinola A. Recurrent Adaptive Classifier Ensemble for Handling Recurring Concept Drifts. *Hindwai Applied Computational Intelligence and Soft Computing*. 2021;p. 1–13. Available from: https://doi.org/10.1155/2021/5533777.
10) Dai Y, Li H, Qian Y, Guo Y, Zheng M. Anticoncept Drift Method for Malware Detector Based on Generative Adversarial Network. *Security and Communication Networks*. 2021;2021:1–12. Available from: https://dx.doi.org/10.1155/2021/6644107.
11) Museba T, Nelwamondo F, Ouahada K. An Adaptive Heterogeneous Online Learning Ensemble Classifier for Nonstationary Environments. *Hindwai Computational Intelligence and Neuroscience*. 2021;p. 1–11. Available from: https://doi.org/10.1155/2021/6669706.
12) Dizaj EA, Kashani MM. Nonlinear structural performance and seismic fragility of corroded reinforced concrete structures: modelling guidelines. *European Journal of Environmental and Civil Engineering*. 2021;p. 1–30. Available from: https://dx.doi.org/10.1080/19648189.2021.1896582.
13) Schwengber BH, Vergutz A, Prates NG, Nogueira M. A Method Aware of Concept Drift for Online Botnet Detection. *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*. 2020;1. doi:10.1109/GLOBECOM42002.2020.9347990.
14) Henke M, Santos E, Souto E, Santin AO. Spam Detection Based on Feature Evolution to Deal with Concept Drift. *JUCS - Journal of Universal Computer Science*. 2021;27(4):364–386. Available from: https://dx.doi.org/10.3897/jucs.66284.
15) Korycki L, Krawczyk B. Concept Drift Detection from Multi-Class Imbalanced Data Streams. *2021 IEEE 37th International Conference on Data Engineering (ICDE)*. 2021;p. 1068–1079. Available from: arXiv:2104.10228.
16) Washha M, Qaroush A, Mezghani M, Sedes F. Unsupervised collective-based framework for dynamic retraining of supervised real-time spam tweets detection model. *Expert Systems with Applications*. 2019;135:129–152. Available from: https://dx.doi.org/10.1016/j.eswa.2019.05.052.
17) Masood F, Ammad G, Almogren A, Abbas A, Khattak HA, Din IU, et al. Spammer Detection and Fake User Identification on Social Networks. *IEEE Access*. 2019;7:68140–68152. Available from: https://dx.doi.org/10.1109/access.2019.2918196.
18) Barushka A, Hajek P. Spam detection on social networks using cost-sensitive feature selection and ensemble-based regularized deep neural networks. *Neural Computing and Applications*. 2020;32(9):4239–4257. Available from: https://dx.doi.org/10.1007/s00521-019-04331-5.
19) Wang X, Kang Q, An J, Zhou M. Drifted Twitter Spam Classification Using Multiscale Detection Test on K-L Divergence. *IEEE Access*. 2019;7:108384–108394. Available from: https://dx.doi.org/10.1109/access.2019.2932018.
20) Alrubaian M, Al-Qurishi M, Alamri A, Al-Rakhami M, Hassan MM, Fortino G. Credibility in Online Social Networks: A Survey. *IEEE Access*. 2019;7:2828–2855. Available from: https://dx.doi.org/10.1109/access.2018.2886314.
21) Chen C, Wang Y, Zhang J, Xiang Y, Zhou W, Min G. Statistical Features-Based Real-Time Detection of Drifted Twitter Spam. *IEEE Transactions on Information Forensics and Security*. 2017;12(4):914–925. Available from: https://dx.doi.org/10.1109/tifs.2016.2621888.