Jestr

Journal of Engineering Science and Technology Review 16 (5) (2023) 13 - 18

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

GridBoost: A classifier with Increased Accuracy to Detect Anomaly in Social Media Networks

Sonali Lunawat*, Jyoti Rao and Pramod Patil

Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India

Received 4 June 2023; Accepted 6 October 2023

Abstract

Social media networks are now essential and play a significant role in society. According to data, the number of active users on various social media platforms including Facebook, WhatsApp, Instagram, and many more is growing rapidly. As a result, there is an increase in risky actions, making the area more unsafe. Personal information security is now seriously threatened. The search for anomalous users is a field that is constantly being researched, but because of the threat that it poses, it is also a field that will never end and will face numerous obstacles, including accuracy. Different Machine Learning and Deep Learning models have been proposed and created by numerous researchers. But, many of these models have scope for improvements, in terms of accuracy and reducing false positives, reducing false negatives. To achieve these enhancements, we have compared different models and using our hybrid model, with attempts for increasing accuracy. In this research we have implement an accuracy-based model named GridBoost which uses hyperparameter parameter tuning fusion with XGBoost. We used a variety of popular classifier models, including Linear Regression (LR), Naive Bayes (NB), KNN (K-Nearest Neighbor), Support Vector Machine (SVM), and GridBoost, which were developed for anomaly identification using four different standard datasets. The performance study shows increased accuracy with our proposed hybrid technique up to 98% when compared to other assessment metrics like precision, recall, and F1-score.

Keywords: Machine Learning, Deep Learning, Anomaly Detection, Social Media Networks, Hybrid Model

1. Introduction

1.1 Social Media Networks: Anomaly Detection

As the internet has grown tremendously in popularity, it has become essential for businesses and individuals to communicate with one another and share information. Social networking sites like Twitter, Facebook, and many others have gained new features as a result of the need, which is growing tremendously and has become a requirement for daily life. The social networking websites are nothing more than a platform in cyberspace where individuals and organizations may connect and form networks to engage in social activities [1]. Users can create a network to share their opinions, stay connected, and experience life outside of their comfort zone while experiencing a real-life encounter. Due to this increase, a vast amount of data is collected, and one can find useful information of an individual or group based on interconnections. Massive amounts of data are gathered, which presents a variety of issues in handling and protecting this data from nefarious uses. Because of this, an attacker can access this vast quantity of data by engaging in nefarious actions including making fake profiles, installing malware, running scripts, probing URLs, using DDOS, creating fake accounts and stock market news, among other things, or by stealing users' private information. For all demographics, including kids, teens, and adults, social media has turned into a deadly environment. Due to these websites, we now face several problems with teen violence, cyberbullying, and cybercrime. Numerous hazards are becoming more prevalent every day, providing academics new ways to consider how to keep people secure. "Something that is not expected or outlier" is what the term "anomaly" refers to. The study of unanticipated structures that need to be discovered has increased due to social media networks. For the same reason, developing an intrusion detection system to find abnormalities has grown in importance and requires extensive research in machine learning or other fields [2]. Finding anomalies in the network involves spotting deviations from "normal" patterns [3]. The performance of machine learning models will be improved by separating anomalies from a large sample of typical cases. This will allow for both detection and alerts of malicious behaviors. Numerous studies on anomaly detection are being conducted to identify fake news producers, cyberbullying, fake profiles or accounts, spammers, malicious intrusions, and many other things. Finding outliers that deviate from the majority or group data's typical trend is the aim of any anomaly detection technique [4]. Due to the high dimensional data structure, detection rate, precision, and processing overhead, anomaly patterns are exceedingly challenging to find. As a result of the numerous obstacles faced by researchers, many developed models are unable to identify anomalies.

However, real users and anomalous users in the social networks are distinguished with respect to the dynamic characteristics of features. So, the classification algorithms for anomaly detection in the social media network are associated with different challenges including False alarm rates, Reliability, Accuracy, Computational Overhead, High dimensional data, Limited datasets, Agility in anomalies behavior in groups, Optimization. Below figure 1 is motivation to work in this field.

*E-mail address: sonali.lunawat@gmail.com
ISSN: 1791-2377
doi:10.25103/jestr.165.02

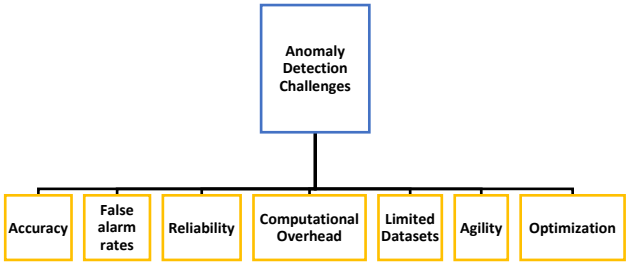


Fig. 1 Challenges in Anomaly Detection

1.2 Type of Anomalies

As shown in Figure 1, there are numerous categories into which anomalies might be divided. Depending on the type of abnormalities are divided into four groups [5]:

A. Point anomalies: This term describes a data instance that differs noticeably as a result of abnormal behaviour in a group. As Friends and their money, for instance.
B. Contextual anomalies: This is referred to as a confined anomaly because the data instance is deemed anomalous as a result of particular restrictions. For instance, climate change [5].
C. Collective/Group anomalies: An assemblage of data instances or clusters that are abnormal in relation to the entire instance is referred to as a collective anomaly. For instance, a group of pupils quitting a class.
D. Horizontal anomalies: A user whose actions are discernible from their participation on various social media platforms, communities, or sources. For instance, a user with connections to many social networks and his past actions.

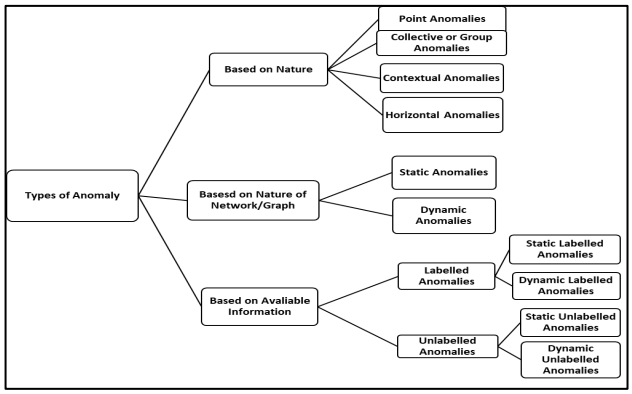


Fig. 2. Classifications of Anomalies

1.3 Type of Learning Techniques in Anomaly Detection

1. Supervised Learning Techniques: Supervised learning techniques use pre-labeled data as normal and abnormal. [5].

2. Unsupervised Learning Techniques: An unsupervised anomaly detection technique not uses pre-labeled data as normal and abnormal. These methods work well with clustering techniques. [5].

3. Semi-supervised Learning Techniques: In semi-supervised techniques dataset uses or defines the labeled information as normal while training model it creates itself an abnormal class. [5].

Machine learning models are effectively applied for anomaly detection in Social Media Networks to achieve accuracy. Many of these models uses classification models to detect anomaly based on input data into labelled classes. To improve the accuracy of these models is challenging. As well as working on one dataset will not be considered to prove the accuracy of any model.

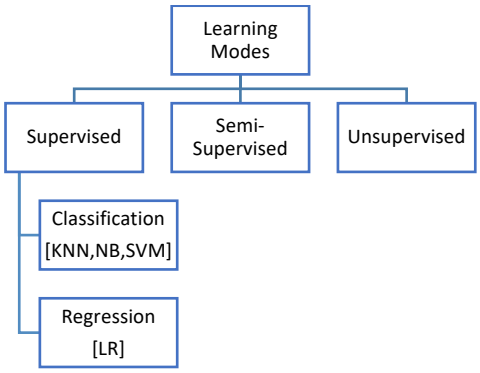


Fig. 3 Classification of Anomaly Detection Learning Modes

With this intent and previous model investigation it is realized that fusion of classifier GridBoost will give better accuracy. The algorithms like Linear Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Naïve Bayes (NB) have shown better accuracy but this classifier have test on one standard dataset or two. In this paper, comparison of all the different classifiers with fusion classifier on all four standard datasets is carried with concluding obtained accuracy till 98% with compared to existing models.

2. Related Work

Rahman M S, et.al. [1] have proposed a Hybrid algorithm combining Decision Tree, Support Vector Machine and Naïve Bayesian classifier (DT- Machine Learning models SVMNB) Model worked well to detects anomaly patterns by producing results need by real life application. But it fails to explore user interest in agile user activities. Garg S, et.al. have implemented Software-Defined Networking (SDN)- based anomaly detection [6] which efficiently worked on large-scale for detecting anomalous events by giving low detection rate. Wanda P, et.al., developed DeepOSN method efficiently to solve scalability problems which increased computational overhead and due to which loss was increased [7]. Yazdi H S, et.al. which does not require retraining if social media behavior changes due to reliability not achieved [8]. Zhong M, et.al., have implemented Security Log Analysis Scheme by achieving high detection capability for all types of attacks but failed to evaluate accuracy on new datasets and algorithms [9]. Gao, et al. [10] has proposed a system for having ensemble learning to collect the advantages of different algorithms and for excessive classification scenarios for selecting best features to increase the accuracy of machine learning models. Machine learning models and evaluated on different parameters like Recall, Precision, Error Rate, Accuracy. Alqahtani, et al. have used the Extreme gradient boosting (XGBoost) algorithm used to detect multi class attacks in wireless sensor networks [11]. Chebrolu et al. has developed an Intrusion Detection system which is life-threatening for dynamic intrusions in real life. [12]. Latah et. al., 2018 [13] investigate the performance of the well-known anomaly-based intrusion detection approaches in terms of accuracy, false alarm rate, precision, recall, f1-measure, area under ROC curve, execution time. Divakar et. al., study an intelligent intrusion detection scheme powered by boosting algorithm. A machine learning based Intrusion Detection

Scheme (IDS) is being proposed [14]. Many such systems exist but they have critical issues of performance like accuracy and efficiency.

The research on different dataset and different models are analyzed in Table1.

Table 1. Analysis of Literature Review

Paper ID	Dataset	Number of Features	Learning Models Used	Limitations
[15]	NSL-KDD	42	Naïve Bayes, Bayes Net, Random Forest, Linear Regression, J48, Bagging. OneR, PART, ZERO	Model gives accuracy for smaller dataset
[16]	NSL-KDD	42	Support Vector Machine, Gaussian mixture Model, Random Forest, Linear Regression	Less number of datasets with less features
[17]	KDD"99	42	k-Means, Random Forest, Naïve Bayes, Support Vector Machine	Issues are still open to develop a classifier that will increase efficiency
[18]	CICIDS2017	72	**Adaboost**	Need to check the classifier with other or real-time dataset
[19]	CICIDS2017	72	Artificial Neural Network, Random Forest	Classifier to be checked on different dataset as if suffers from overall size and it is often too bulky to use.
[20]	UNSW-NB15	43	Support Vector Machine, J48, Random Forest, ZERO	Proving efficiency of other classifiers and on other standard datasets
[21]	UNSW-NB15	43	Deep Neural Networks, Random Forest	Improving the performance of intrusion detection new approaches should be developed
[22]	UNSW-NB15, CICIDS-2017, ICS Cyber-attack Dataset	43,72	Linear Regression, Gaussian Naive Bayes, K-Nearest Neighbor, Decision Tree, AdaB, RF, Convolutional Neural Network, CNN, short-term memory, Gated Recurrent Unit, SimpleRNN, Deep Neural Networks	Selection of classifiers will depend on the datasets used
[23]	Phishing dataset	48	Linear Regression, Decision Tress, Random Forest, AdaB, K-Nearest Neighbor, Neural Network, Support Vector Machine, Gradient, Boosting, XGBoost Decision Tress, Naïve Bayes, Support Vector Machine, XGBoost, K-Nearest Neighbor, Linear Regression, and Random Forest	Compare with more features to the dataset to improve the performance of these models
[24]	MIDFIELD	-		Analyzing different classifiers for optimizing Hyper Parameters
[25]	Collected customer feedback	-	Random Forest	Parameter tuning has proved to improve accuracy. Random Forest classifier take more execution time when the number of trees in the forest is increased.

Summary of related work

The existing popular classifiers gives better accuracy for smaller datasets and issues are still open to develop a classifier that will increase efficiency. Improving the performance of intrusion detection, one should develop new approaches with parameter tunning to improve accuracy. Comparing classifier with more features dataset to improve the performance of the models.

3. Proposed Workflow

In our proposed model as in Figure 4 we have undergone below step to achieve improved results.

STEP 1: Data Collection: We investigated and gathered as the existing systems four standard datasets, such as the Wireless Sensor Network Detection System Dataset (WSN-DS), the KDD Dataset, the CICIDS2017 Dataset, and the Phishing Dataset.

STEP 2: Labeling

One-hot encoding is used in machine learning to convert categorical data as model can be fit using numerical data. The benefits of one hot encoding are as follows:

- It converts categorical value to a numeric value.
- It is used improve model performance
- It can maintain a proper order while conversion.

STEP 3: Splitting Dataset

To avoid overfitting, problem data is split into a set of Folds. In which at single step execution of all dataset is done giving the best performance accuracy. A k-fold cross-validation is used because of its generalized process.

K-Fold Cross-Validation:

In K-fold Cross Validation K folds of equal size in which K-1 groups as Training Dataset and remaining K is used as Testing Dataset in which validation is carried K times. Estimation of output is based on the K tests. In this paper the dataset is categorized as 80% Training data set and 20% testing data set.

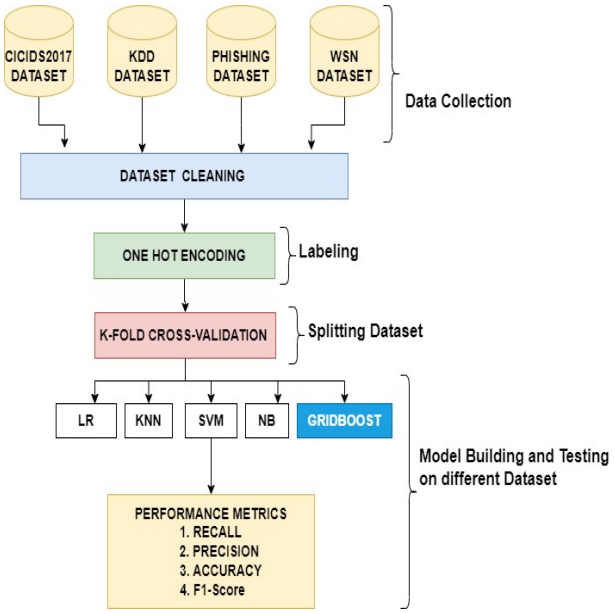


Fig. 4. Proposed Methodology

STEP: 4 Classifiers

Naive Bayes (NB) is mainly used for classification problems and works on principles of Bayes theorem, by finding the probability of a hypothesis of evidence. The "naive" in the algorithm are features that are conditionally independent of each other for given the class label. It is a probabilistic model from the training data and estimating the probabilities of the different features given each class label [5].

Random Forest (RF) is a combination of regression and classification in machine learning algorithms. This algorithm helps to reduce overfitting and increase model diversity. Random Forests can handle high dimensional data and nonlinear relationships [5].

K-nearest neighbors (KNN) used mainly for classification and regression problems. The selection of k is an important hyper parameter of the KNN algorithm, which defines the size of the neighborhood used for prediction. KNN does not make any assumptions for distribution of data, also called a non-parametric algorithm [5].

Support Vector Machine (SVM) is applicable for classification, regression, and outlier detection related use cases in machine learning algorithms. The main clue behind using an SVM is all about deciding the hyperplane that correctly separates the data into binary classes. In the case of a binary classification problem, the hyper plane decides the margin between the classes by finding distance between the hyper plane and the nearest data points from each class.

Grid search is a hyperparameter optimization technique [25][26].

A: Set of hyperparameters
Val: Set of possible values for each hyperparameter
M (A_i, Val_j): Performance metric (e.g., accuracy, loss) for a specific combination (A_i, Val_j)

The goal of grid search is to find the optimal hyperparameter values that maximize (or minimize) the performance metric:

Optimal Hyperparameters = argmax(argmin) [M (A_i, Val for all (A_i, Val_j) in a Grid.

XGBOOST classifier [23]: It combines software and hardware optimization techniques to yield higher accuracy using fewer computing resources less amount of time.

Regularized boosting: To reduce overfitting due noise in the training data could negatively impacts the performance model on new data.

Handle missing values automatically: We don't need to care about the missing values as it is handled automatically.

Cross-validation at each iteration by dividing data into two set one acts train a model and the other to validate the model.

Tree pruning: Pruning is used to remove parts of the tree that does not provide value to classification.

Objective function= Loss (T_i, $T_{pred.}$) + Regularization

Where: Loss (T_i, $T_{pred.}$) is the loss function that for target value as T_i and predicted values $T_{pred.}$ Regularization refers to a penalty term that discourages complexity in the model.

DATASET

1. **CICIDS2017 Dataset:** This dataset contains a total 79 features. The Dataset was constructed using the NetFlow Meter Network Traffic Flow analyzer [30]. The dataset up-to-date data and common attacks, which look like the true real-world data.

2. **KDD Dataset:** It is a DARPA 98 Intrusion Detection Evaluation version created by Lincoln laboratory at MIT containing 43 features labeled as attack or normal [28]. The dataset contains the security attacks as Denial of Service in which illegal users causing resource constraint, User to Root in which attackers acts as normal user in a group and uses root credentials, Remote-to-local in which enemy tries to gain access of owner system by sending vulnerable packets to exploit inside the network, and PROBING in which send vulnerabilities arises by changing network configuration and identifying loop holes and Normal which is not a threat.

3. **Phishing Dataset:** This data set consists of two classes of Normal and malicious (Phishing). Each sample in the data set contains 75 features [27]. Contributed by Yazdi and his team and accessed on March 2022.

4. **WSN-DS Dataset:** The dataset is used to detect intrusions in WSN. The WSN-DS dataset was composed to detect and categorize types of denial-of- service (DoS) attacks. The WSN-DS dataset has 23 features [29]. It has five main groups from which four are of type DoS attack labeled as attacks including Blackhole, Gray hole, Flooding, and Scheduling attacks and Normal.

4. Results and Discussions

Evaluation Parameters [4][9][31]:

True Positives (TP): It is predicted as anomaly by us and the real output was also anomaly.

True Negatives (TN): It is predicted as normal by us and the real output was normal.

False Positives (FP): It is predicted as anomaly by us and but it was normal. False Negatives (FN): It is predicted as normal by us and but it was anomaly.

Precision: Number of correct positive results

Precision = TP/ (TP + FP)

Recall: Correct positive results

Recall = TP / (TP + FN)

F1 Score: It is used to measure the test's accuracy

F1 Score=2*[(Precision*Recall) / (Precision+ Recall)]

Accuracy: Number of correct Predications / Total number of Predications.

Table 2. Comparison of Accuracy achieved by different models and different datasets

	LR	NB	KNN	SVM	GridBoost
KDD Dataset	80.2	77	92	57	96
CICIDS Dataset	98.1	55.6	98.2	94	98.5
WSN Dataset	98	94.2	99	95	94
PHISHING Dataset	96	91	97	87	98

Table 3. Comparison of Precision achieved by different models and different datasets

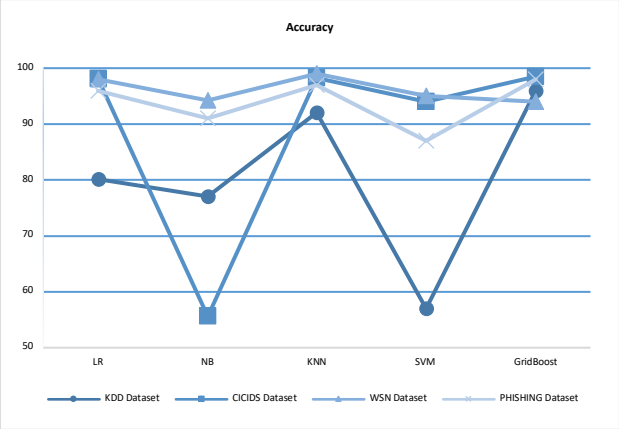
	LR	NB	KNN	SVM	GridBoost
KDD Dataset	80.1	53	92	76	96
CICIDS Dataset	95.6	40.7	98.1	94	98.5
WSN Dataset	94.1	73	94	73	94.3
PHISHING Dataset	96	92	97	89	97.5

Table 3. Comparison of Recall achieved by different models and different datasets

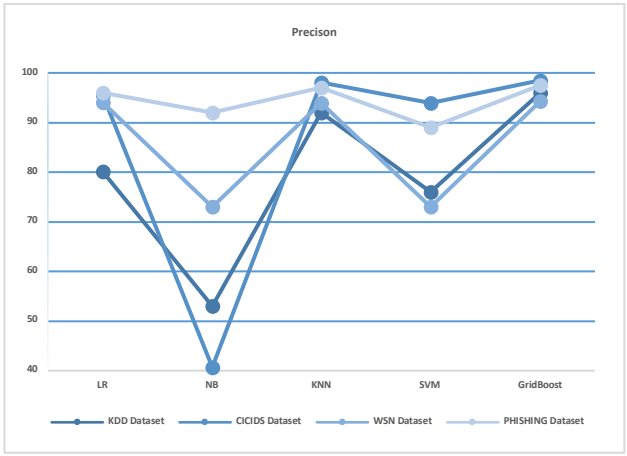
	LR	NB	KNN	SVM	GridBoost
KDD Dataset	81	52	92	54	91
CICIDS Dataset	98.5	98.5	98.2	91	98.5
WSN Dataset	81	79	92	79	94
PHISHING Dataset	95	91	97	86	98

Table 4. Comparison of F1score achieved by different models and different datasets

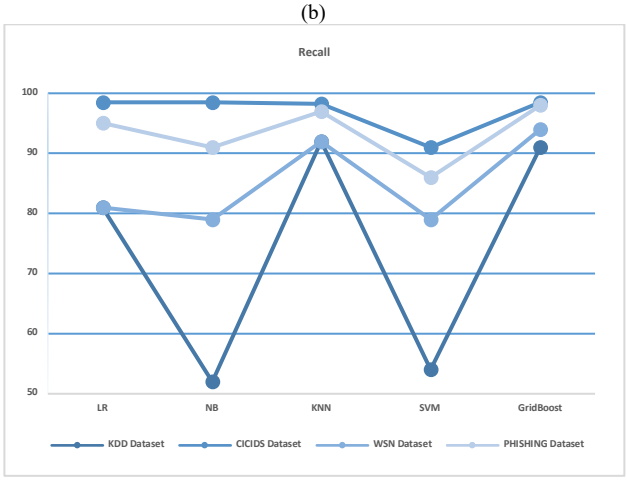
	LR	NB	KNN	SVM	GridBoost
KDD Dataset	81	52	92	54	91
CICIDS Dataset	98.5	98.5	98.2	91	98.5
WSN Dataset	81	79	92	79	94
PHISHING Dataset	95	91	97	86	98



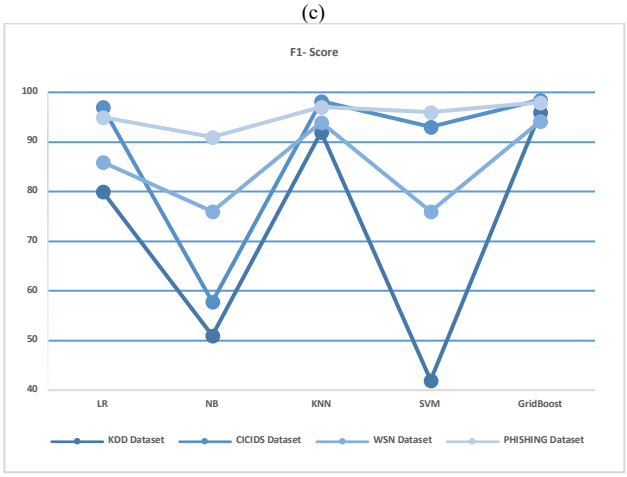
(a)



(b)



(c)



(d)

Fig. 5. Experimental Results for different evaluation parameters (A, B, C, D)

5. Conclusion

In this work, four separate standard datasets were used to conduct an evaluation study based on various metrics and machine learning classification techniques. The study separated data into two groups using a two-fold technique, labelled the data using a single hot encoding, and compared the performance of assessment metrics using standard machine learning models like RF, KNN, LR, SVM, NB, and GridBoost. The datasets WSN-DS, KDD, CICIDS2017, and Phishing were utilized to classify the anomaly across these four datasets. Experimental results are clearly shown utilizing the suggested approaches, and it is discovered that

GRIDBOOST has the highest accuracy and detection rate when compared to other models and datasets, with an accuracy range of 94%. to 97%. In the near future study, we intend to reduce features by employing an alternative ensemble strategy for deeply identifying attacks kinds and tracking group user activity and discover algorithm in the field of optimalization using metaheuristic algorithms.

References

[1] Md. S. Rahman, S. Halder, Md. A. Uddin, and U. K. Acharjee, "An Efficient Hybrid System for Anomaly Detection in Social Networks," *Cybersecurity*, vol. 4, no. 1, Mar. 2021, doi: 10.1186/s42400-021-00074-w.

[2] T. Balaji, C. S. R. Annavarapu, and A. Bablani, "Machine Learning Algorithms for Social Media Analysis: A Survey," *Comput. Sci. Rev.*, vol. 40, p. 100395, May 2021, doi: 10.1016/j.cosrev.2021.100395.

[3] N. Moustafa, J. Hu, and J. Slay, "A holistic Review of Network Anomaly Detection Systems: A Comprehensive Survey," *J. of Netw. and Comput. Appl.*, vol. 128, pp. 33–55, Feb. 2019, doi: 10.1016/j.jnca.2018.12.006.

[4] M. S. Sudha and K. Valarmathi, "An Optimized Deep Belief Network to Detect Anomalous Behavior in Social Media," *J. Ambient Intell. Humanized Comput.*, Jan. 2021, doi: 10.1007/s12652-020-02708-2.

[5] Naznin Sultana, Sellappan Palaniappan, "A Survey on Online Social Network Anomaly Detection," *Int. J. Innovative Sci Res. Technol.*, vol. 3, no. 3, Mar. 2018, ISSN no: -2456-2165.

[6] S. Garg, K. Kaur, N. Kumar, and J. J. P. C. Rodrigues, "Hybrid Deep-Learning-Based Anomaly Detection Scheme for Suspicious Flow Detection in SDN: A Social Multimedia Perspective," *IEEE Trans. on Multimedia*, vol. 21, no. 3, pp. 566–578, Mar. 2019, doi: 10.1109/tmm.2019.2893549.

[7] P. Wanda, M. E. Hiswati, and J. Huang, "DeepOSN: Bringing Deep Learning as Malicious Detection Scheme in Online Social Network," *IAES Int. J. Artif. Intell.*, vol. 9, no. 1, p. 146, Mar. 2020, doi: 10.11591/ijai.v9.i1.pp146-154.

[8] E. Mahmodi, H. S. Yazdi, and A. G. Bafghi, "A Drift Aware Adaptive Method Based on Minimum Uncertainty for Anomaly Detection in Social Networking," *Exp. Sys. Applic.*, vol. 162, p. 113881, Dec. 2020, doi: 10.1016/j.eswa.2020.113881.

[9] M. Zhong, Y. Zhou, and G. Chen, "A Security Log Analysis Scheme using Deep Learning Algorithm for IDSS in Social Network," *Secur. Commun. Netw.*.vol. 2021, pp. 1–13, Mar. 2021, doi: 10.1155/2021/5542543.

[10] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An Adaptive Ensemble Machine Learning Model for Intrusion Detection," *IEEE Access*, vol. 7, pp. 82512–82521, Jan. 2019, doi: 10.1109/access.2019.2923640.

[11] M. Alqahtani, A. Gumaei, H. Mathkour, and M. M. B. Ismail, "A Genetic-Based Extreme Gradient Boosting Model for Detecting Intrusions in Wireless Sensor Networks," *Sensors*, vol. 19, no. 20, p. 4383, Oct. 2019, doi: 10.3390/s19204383.

[12] S. Chebrolu, A. Abraham, and J. Thomas, "Feature Deduction and Ensemble Design of Intrusion Detection Systems," *Comput. & Secur.,* vol. 24, no. 4, pp. 295–307, Jun. 2005, doi: 10.1016/j.cose.2004.09.008.

[13] M. Latah and L. Toker, "Towards An Efficient Anomaly-Based Intrusion Detection for Software-Defined Networks," *IET Netw.*, vol. 7, no. 6, pp. 453–459, Nov. 2018, doi: 10.1049/iet-net.2018.5080.

[14] S. Divakar, R. Priyadarshini, R. K. Barik and D. Sinha Roy, "An Intelligent Intrusion Detection Scheme Powered by Boosting Algorithm," in *11th Int. Conf. on Cloud Comput., Data Sci. & Engineering (Confluence), Noida, India*, 2021, pp. 205-209, doi:10.1109/Confluence51648.2021.9377076.

[15] H. L. Malhotra and P. Sharma, "Intrusion Detection using Machine Learning and Feature Selection," *Int. J. Comput. Netw. Inf. Secur.*, vol. 11, no. 4, pp. 43–52, Apr. 2019, doi: 10.5815/ijcnis.2019.04.06.

[16] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," *Procedia Comput. Sci.*, vol. 89, pp. 117–123, Jan. 2016, doi: 10.1016/j.procs.2016.06.016.

[17] Y. E. Mourabit, A. Toumanari, A. Bouirden, and N. E. Moussaid, "Intrusion Detection Techniques in Wireless Sensor Network using Data Mining Algorithms: Comparative Evaluation Based on Attacks Detection," *Int. J. Adv. Comput. Sci. and Appl.*, vol. 6, no. 9, Jan. 2015, doi: 10.14569/ijacsa.2015.060922.

[18] Yulianto, Arif and Sukarno, Parman & Suwastika, Novian. "Improving AdaBoost-based Intrusion Detection System (IDS) Performance on CICIDS 2017 Dataset," *J. Phys.: Conf. Series*, doi: 10.1088/1742-6596/1192/1/012018.

[19] Zachariah Pelletier and Munther Abualkibash, "Evaluating the CICIDS-2017 Dataset Using Machine Learning Methods and Creating Multiple Predictive Models in the Statistical Computing Language R," *Int. Res. J. Adv. Eng. Sci.*, vol. 5, no. 2, pp. 187-191, 2020.

[20] M. Hammad, W. Elmedany, and Y. Ismail, "Intrusion Detection System using Feature Selection With Clustering and Classification Machine Learning Algorithms on the UNSW-NB15 dataset," in *Proc. Int. Conf. Innov. Intell. for Inform., Comput., and Technol. (3ICT)*, Dec. 2020, doi: 10.1109/3ict51146.2020.9312002.

[21] O. Faker and E. Doğdu, "Intrusion Detection Using Big Data and Deep Learning Techniques," in *ACM SE '19: Proc. of the 2019 ACM Southeast Conf.*, Apr. 2019, doi: 10.1145/3299815.3314439.

[22] N. Elmrabit, F. Zhou, F. Li, and H. Zhou, "Evaluation of Machine Learning Algorithms for Anomaly Detection," in *Proc. Int. Conf. Cyber Secur. Protection of Digital Services (Cyber Security)*, Jun. 2020, doi: 10.1109/cybersecurity49315.2020.9138871.

[23] V. Shahrivari, M. M. Darabi, and M. Izadi, "Phishing Detection using Machine Learning Techniques," *Cryptography Secur.* Sep 2020, doi: 10.48550/arXiv.2009.11116.

[24] L. Zahedi, F. G. Mohammadi, S. Rezapour, M. W. Ohland, and M. H. Amini, "Search algorithms for automated Hyper-Parameter tuning," in *Proc. Int. Conf. on Data Sci. (ICDATA'21),* Apr. 2021.

[25] S. G. C. G and B. Sumathi, "Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 9, Jan. 2020, doi: 10.14569/ijacsa.2020.0110920.

[26] L. Yang and A. Shami, "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020, doi: 10.1016/j.neucom.2020.07.061.

[27] Sadoghi Yazdi, Hadi; Mahmodi, Emad; Ghaemi Bafghi, Abbas (2020), "Data for: An Online Minimal Uncertainty Drift-Aware Method for Anomaly Detection in Social Networking," Mendeley Data. [Online]. Available: https://data.mendeley.com/datasets/zw7knrxpy5/1

[28] Kaggle Online, "KDD99 dataset Intrusion detection dataset," Available: https://www.kaggle.com/datasets/toobajamal/kdd99-dataset

[29] Iman Almomani,Bassam Al-Kasasbeh and Mousa AL-Akhras, "WSN-DS: A dataset for intrusion detection systems in wireless sensor networks," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/bassamkasasbeh1/wsnds

[30] Iman Sharafaldin, Arash Habibi Lashkari, and Ali A. Ghorbani, "CICIDS2017 Intrusion Detection Evaluation Dataset." Kaggle. [Online]. Available: https://www.kaggle.com/datasets/cicdataset/cicids2017

[31] N. Kerthana,V. Vinod, and Sudhakar Sengan, "A Novel Method for Multi-Dimensional Cluster to Identify The Malicious Users on Online Social Networks*," J. Eng. Sci. Technol.*, vol. 15, no. 6, Sep 2020.