*2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*
*Pune, India. Aug 26-28, 2022*

# Network Data Feature Selection in Detecting Network Intrusion using Supervised Machine Learning Techniques

Arjonel M. Mendoza
*Batangas State University*
Batangas, Philippines
mendozaarjonel@gmail.com

Rowell M. Hernandez
*Digital Transformation Center*
*Batangas State University*
Batangas, Philippines
rowell.hernandez@g.batstate-u.edu.ph

Ryndel V. Amorado
*Batangas State University*
Batangas, Philippines
ryndel.amorado@g.batstate-u.edu.ph

Myrna A. Coliat
*Batangas State University*
Batangas, Philippines
myrna.coliat@g.batstate-u.edu.ph

Poul Isaac C. De Chavez
*Batangas State University*
Batangas, Philippines
poulisaac.dechavez@g.batstate-u.edu.ph

2022 2nd Asian Conference on Innovation in Technology (ASIANCON) | 978-1-6654-6851-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASIANCON55314.2022.9909208

***Abstract*— Network attacks have become necessary in today's time due to increased network traffic. To determine whether network traffic is normal or anomalous a supervised machine learning system is developed. A network intrusion detection system (IDS) is a must-have piece of a security system. This proposed study aims to discover new patterns automatically from substantial quantities of network data, reducing time manually compiling intrusion and normal behavior patterns. The best model in terms of detection success rate was discovered using a supervised learning algorithm and feature selection method. AdaBoost outperforms Neural Network, kNN, and Naive Bayes in supervised machine learning with feature selection in this study, with a detection accuracy of 100.00%, 99.30%, 91.60%, and 99.70%, respectively. The Network Intrusion Detection dataset is used to classify network intrusions to evaluate the study and it has also been used in past studies. On the other hand, the proposed model proved to be more effective than other studies in terms of intrusion detection. The proposed approach can be used in various fields, including finance, health, and transportation. Furthermore, additional parameter tuning could be added, and different feature selection techniques could be used to improve the performance of the classifiers.**

***Keywords—intrusion detection, machine learning classifications, neural network, dataset, feature selection.***

## I. INTRODUCTION

Data security is becoming increasingly important as the computer network develops at a rapid pace. The term "security" refers to the level of protection provided to a network or system. Availability, confidentiality, and integrity of data are the three fundamental security purposes [1]. Intrusion is the term used to describe network attacks. It refers to any set of opposed behaviors to jeopardize the information's security goals. One of the massive data sets is intrusion detection. It aids the system in protecting itself against external threats [2].

According to the detection approaches, IDS are classified into two groups: Anomaly Detection and Misuse Detection.

Misuse detection works by creating a malicious behavior pattern and identifying patterns based on those intrusions, i.e., it looks for activity that corresponds to recognized intrusion techniques. The critical benefit of misuse detection is that it is more accurate in detecting all known attacks. This method, on the other hand, has the disadvantage of only detecting intrusions that follow predetermined patterns. As Muda et al. mention, anomaly detection anticipates the network's or profile's expected behavior. Any significant variations from this established typical behavior are flagged as potential attacks.

However, not all of these are attacks. This method's primary advantage is its ability to investigate unknown and complex intrusions. This method, nevertheless, has a high false alarm rate and a low detection rate [3].

As a result, to increase network security, a dynamic approach known as Intrusion Detection System (IDS) is proposed. The IDS collect online information from the monitors, networks and analyzes it, and categorizes it as usual or anomaly before transmitting the findings to the system administrator [4].

This paper introduces machine learning algorithms, which examine massive amounts of computer system or network data logs to quickly and accurately identify intrusion activity. Data analysis is complex due to the high volume of network traffic. This necessitates using intrusion detection systems (IDS) in conjunction with various data mining approaches for intrusion detection. Machine learning is defined as the process of retrieving knowledge from a database using Knowledge Discovery from Database (KDD).

A supervised machine learning-based distributed network IDS system concept has been developed for a higher accuracy rate, and data cleaning is completed at the preprocessing stage. After preprocessing, machine learning techniques are used to predict the data set's accuracy rate, the algorithm was presented, and relevant experimental results were examined.

## II. RELATED LITERATURE

Numerous studies have been conducted to detect an intrusion into a network using various machine learning techniques. Researchers focused on developing efficient methods for detecting network intrusion and achieving high accuracy. The different machine learning algorithms, tools, and features provided for network intrusion detection in previous years are evaluated in this literature.

With data preprocessing and cleaned data, a machine-learning algorithm was used for classification techniques.

A classification technique comparison was performed to compare the data more precisely. The Random Tree, J48, Random Forest, and Naive Bayes approaches are used for classification. The classification technique, random forest, outperforms other algorithms in terms of accuracy, with 99.9 percent for normal and 99.7 percent for anomaly [5]. To improve the efficiency of the detection process, (Sharma, 2018) used data mining in conjunction with fuzzy and GA. The proposed algorithm is compared to the single Naïve Bayes classifier using the KDD Cup '99 dataset. The outcomes display that the K Decision Tree-Based classification strategy improved accuracy and detection rates by accurately detecting novel incursions while lowering false alarms. The performance of the Naive Bayes classifier was improved by using Decision Tree-Based classification. On the other hand, Decision Tree-Based Categorization has limitations in detecting similar attacks like U2R and R2L [6].

Bhosale et al. presented a hybrid approach determination for Intrusion Detection systems in their study. The researchers utilized the feature selection approach and the Naive Bayes classifier. The proposed method's performance demonstrates its efficiency when compared to other methods. The accuracy was 92 percent higher than the previously proposed technique, 90 percent [7].

Gupta et al. used Data Mining methods in their study and demonstrated various levels of accuracy. The NSL-KDD dataset was preprocessed using the mean normalization method. Surprisingly, linear regression was found to be extremely effective at detecting network attacks, with an accuracy rate of 80%. With a 67.5% accuracy, K-Means Clustering, a semi-supervised technique, produced satisfactory results [8].

Similarly, Wankhade et al. used a fine IDS with a low false alarm rate and a high accuracy and detection rate. The primary research method is clustering analysis, aiming to improve detection rates while lowering false alarm rates. None of the known clustering algorithms, according to the findings, has a high detection rate and a low false alarm rate. As a result, the intrusion detection system will use a mixed data mining approach [9]. Correspondingly, the study Limiao et al. relied on a correlation analysis method that included the data mining algorithm in the IDS. The proposed strategy improves the detection efficiency of intrusion detection systems [10].

Xue et al. reviewed current intrusion detection and data mining technologies in their study. Anomaly and Misuse Detection: A Review of Data Mining Algorithms was used. Based on their findings, the researchers identify the issues that the existing data mining technique faces in intrusion detection applications [11]. Kumari et al. conducted a study that presented a research problem and accomplishments in extensive data security using anomaly-based IDS. According to the findings, data mining applications can help humanity and save lives. This technology does have some drawbacks. Intruders may steal data from a variety of databases and files. People now recognize that to combat terrorism, new ideas and rules must be developed. The main problem is the difficulty in securing databases [12]. Interestingly, Fuertes et al. sought to develop and implement a system based on data mining model approaches and its application in detecting network data as an information security innovation strategy in a public-sector organization dedicated to social security. To create IDS/IPS, the J48 and REPTree decision algorithms and free software processed in the WEKA tool were used. The seriousness of the threats discovered in the organization's assets prompted the development of strategic safeguards to reduce risks, such as data protection, service protection, communications protection, and facility protection [13].

The work of Haq et al. on applying several classifier techniques in intrusion detection systems is a new study in machine learning and artificial intelligence. For a long time, it has piqued the interest of researchers. This article discovered 49 research papers related to various classifiers for intrusion detection. This survey paper does not claim to be an in-depth examination of those studies, but it does provide a reasonable viewpoint and a valid comparison of works in this subject over those years [14].

In the studies presented, the random forest classification technique outperforms other algorithms in terms of accuracy, with 99.9% for normal and 99.7% for anomaly [5]. Surprisingly, with an accuracy of 80%, linear regression proved to be extremely effective at detecting network intrusions. K-Means Clustering, a semi-supervised approach, produced satisfactory results with a 67.5% accuracy [8]. Furthermore, the experimental results show that the K Decision Tree-Based classification technique improved accuracy and detection rates by recognizing unique invasions while decreasing false alarms [6]. Bhosale et al. proposed a hybrid determination approach for IDS in their study. The accuracy was 92% higher than the previously proposed method, 90% [7].

Besides that, Wankhade et al. used an effective IDS that demanded a high detection rate, a low false alarm rate, and a high accuracy. According to the data, none of the current clustering methods increased the height detection rate and a low false alarm rate. As a result, the intrusion detection system will use a mixed data mining strategy [9]. In their investigation, Limiao et al. used a correlation analysis method that integrated data mining technologies into the intrusion detection system. The proposed method enhances the detection efficiency of intrusion detection systems [10]. In their study, Xue et al. evaluated existing intrusion detection and data mining tools. The Review of Data Mining Algorithms was used in Anomaly and Misuse Detection. The researchers identify the existing data mining technique's challenges in intrusion detection applications based on their findings.

## III. MATERIALS AND METHODS

The audited dataset was presented, which included a variety of simulated intrusions in a military network setting. It established a setting for obtaining raw IP/TCP dump data for a network. It is surrounded by a barrage of attacks as if it were an open area. Using a well-defined protocol TCP packet is a set of connections that start and stop at regular intervals, during which data is sent and received from a source IP address to a target IP address. Each connection is classified as either normal or an attack with a single attack type. Each connection record length is approximately 100 bytes. Each TCP/IP connection has forty-one quantitative and qualitative features extracted from it (38 quantitative and 3 qualitative characteristics). There are 25,192 instances and 42 attributes

in the dataset. Two categories are included for class variables: normal and anomalous [21].

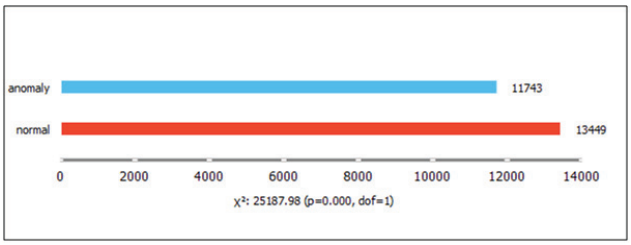


Fig. 1. Class Variables

Figure 1 depicts the data distribution in terms of the class variables: normal and anomalous. Anomaly or with an attack obtained 11,743. On the other hand, the normal class has 13,449 instances—the total instances, both normal and with an anomaly, is 25,192.

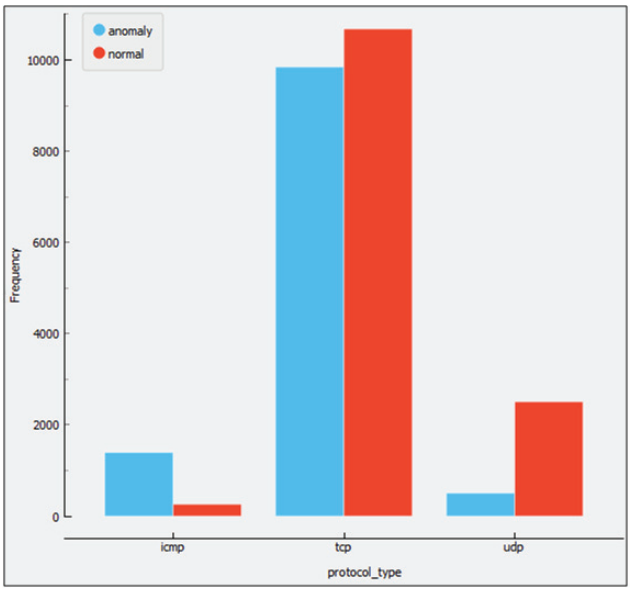


Fig. 2. Protocol Type

Figure 2 shows a normal and anomaly attack for each protocol type. ICMP has 1,655, but only 261 of them are normal, and the rest are affected by an anomaly attack. TCP normal attack yielded 10,681, and anomaly produced 9,845 with 20,526 instances. UDP has 3,011 instances, 504 of which are anomalous, and the rest are normal.

The proposed work uses Network Intrusion Detection datasets because modern networks are heavily hampered by anomalies. To address this issue, the authors use an IDS that completes its task in two phases: preprocessing and feature selection, and it is algorithm-based.

Figure 3 depicts the entire investigation process. As an early model creation stage, the authors employed the Network Intrusion Detection dataset from Kaggle with 41 attributes, which exposed the distribution of data methods to provide training and testing data. Data distribution is a crucial element that can affect each model's performance, whether it employs the default parameters or not, and even the status of feature selection. The dataset features were reduced to 20 by selecting the best features using scoring methods such as ANOVA, information gain, and ratio.

Neural Networks. kNN, Naive Bayes, and AdaBoost are the machine learning models used in this study. The confusion matrix is also utilized to check the proportions of instances between the predicted and actual classes.

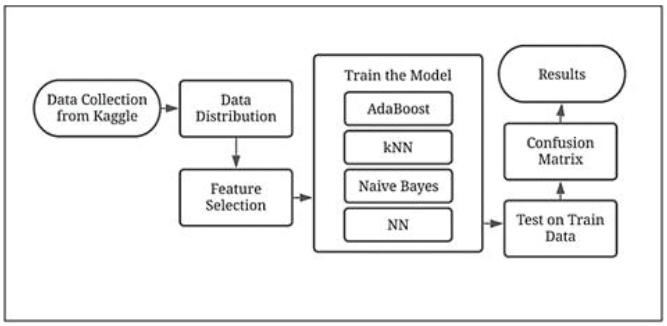


Fig. 3. Process Flow of the Proposed Model

Following the initial training, the model's initial training was evaluated using the K (10)-fold rule. The statistical evaluation performed during the initial training duration produced results higher than most studies mentioned above but lower than the most recent works [22].

Nonetheless, when the authors used the cross-validation technique with test-on-train data, the evaluation metrics produced a higher rating, making the phase the most notable contribution and critical process in this work.

### *A. Feature Selection*

A critical aspect of machine learning is data dimensionality reduction, and extensive research has been conducted to find a reliable feature selection method. The ranking approach is utilized for feature selection. The filter technique chooses its attributes based on its scores in various statistical tests that determine their significance based on their relationship to the dependent or outcome variables. The best features were identified using ANOVA, Information Gain, and Information Gain Ratio. The ranker technique ranks all of the features in the dataset using an attribute evaluator. According to the study (Taher et al. 2019), a combination of a supervised learning algorithm and feature selection method must be used to find the best model in terms of detection success rate. According to their findings, when classifying network traffic, with wrapper feature selection, support vector machine (SVM) techniques are outperformed by ANN-based machine learning [23] [24] [25].

### *B. Evaluation*

According to (Hu et al. 2008), most learning algorithms are prone to overfitting, but the AdaBoost algorithm corrects weak classifier misclassifications and is less prone to overfitting. The recognition results of the AdaBoost-based classifiers are generally encouraging. Intrusion detection data sets are a mix of categorical and continuous data types. The various feature types in such datasets make finding relationships difficult. With no forced conversions between continuous and categorical features, the relationships between these two features are combined by the weak classifiers for constant features and the weak classifiers for absolute attributes into a robust classifier. The AdaBoost algorithm is high-speed [15] [26]. The KNN algorithm was cited by Chen et al. as a non-parametric approach used for classification or regression in pattern recognition. It determines the degree of similarity between all training samples and unlabeled samples. Each selection is labeled in the training samples, which are vectors in a multidimensional feature space. K is a user-defined constant value in the

classification phase. An unknown vector (a query or test point) is distinguished by assigning the most k training samples most frequent label closest to the query point [16]. According to the study of Panda and Patra, with Naive Bayes, there's no need for complicated iterative parameter estimation schemes. As a result, it's simple to use with large data sets. Even though it may not be the best classifier in any given application, it is robust, easy to interpret, and often performs surprisingly well [17]. Debar et al. mention that a group of simple units known as neurons, are extremely unified and alter a set of inputs into a set of anticipated outputs [18]. (Sani et al. 2009) mention that neural network is used in anomaly detection to aid in the isolation of new threats or to continuously adapt and learn what is and is not typical for a given system [19].

## IV. RESULTS AND DISCUSSION

The experiment is divided into two parts and is carried out using the Orange open-source software suite, widely used in data mining and machine learning. The authors extracted the most relevant features in the first section using various feature selection scoring methods. To find the best result suitable for the proposed algorithm a ranker algorithm is used in the filter method. The training data is drawn from the Network Intrusion Detection dataset, containing 25,192 labeled instances. Using feature selection, it was discovered that 20 features out of 41 features present in the datasets are the most relevant.

Using the training dataset and feature selection, the Orange software suite employs four models. The model must first be trained using a training dataset to classify using supervised machine learning. The authors used 20% of the Network Intrusion Detection dataset, which has 25,192 labeled data instances, as training data. We used the AdaBoost, kNN, Nave Bayes, and NN learning algorithms to train the model for each feature selection method.

Table 1 shows the measurements of all classifiers using Orange software and the outcomes of the classifier in the confusion matrix table over-tested network data. Figure 6 depicts the accurate measurements of classification techniques, also known as the efficient method. Table 2 shows the four classifiers' accuracy based on different classification attributes.

The detection accuracy in AdaBoost was 100.00%, with 13,448 instances predicted as normal and 11,744 instances predicted as an anomaly. With 13,431 and 11,761 predicted normal and anomaly classes, respectively, kNN achieves 99.30 % detection accuracy. On the other hand, Naïve Bayes predicted 15,073 for the normal class and 10,119 for the anomaly class with 91.60% detection accuracy. Finally, Neural Networks achieved a detection accuracy of 99.70%, with 13,456 and 11,736 predicted normal and anomaly classes, respectively.

More so, this type of study entails the acceptance of the end-users which includes the technical and non-technical people who intend to receive the technology. A similar approach by previous studies can be incorporated to assess the mentioned concerned [27] [28].

A supervised learning algorithm and feature selection method were used by the authors to find the best model in terms of detection success rate. AdaBoost supervised machine learning with feature selection outperforms kNN, Naïve Bayes, and Neural Network in this study, with a detection accuracy of 100.00%, 99.30%, 91.60%, and 99.70%, respectively.

TABLE I. THE EFFICIENCY OF THE MODEL

| Criteria | Supervised Machine Learning Model | | | |
|---|---|---|---|---|
| | ***AdaBoost*** | ***kNN*** | ***Naïve Bayes*** | ***Neural Networks*** |
| Detection Accuracy | 100.00 | 99.30 | 91.60 | 99.70 |
| Correctly Classified Instances | 25,191 | 25,008 | 23,064 | 25,125 |
| Incorrectly Classified Instances | 1 | 184 | 2,128 | 67 |

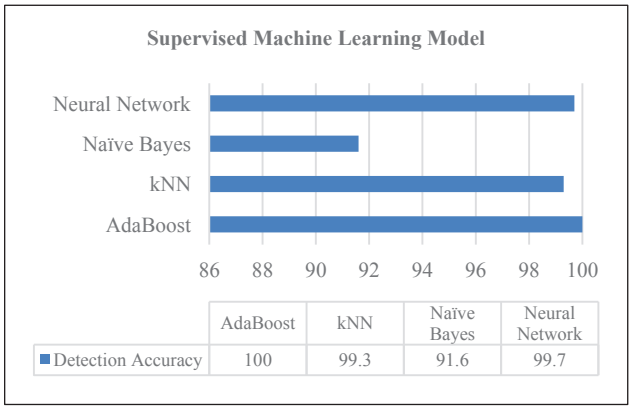


Fig. 4. Accuracy between algorithms

(Taher et al. 2019) used a supervised learning algorithm in conjunction with a feature selection method. When classifying network traffic, ANN-based machine learning with wrapper feature selection outperforms support vector machine (SVM) techniques by 94.02 percent and 82.34 percent, respectively [20].

On the other hand, (Mohan et al. 2020) classified the IDS in their study using a data mining classification system. J48, Random Tree, Random Forest, and Naïve Bayes approaches were used with a machine-learning algorithm that included cleaned data and data preprocessing. In terms of accuracy, the random forest classification technique outperforms other algorithms with 99.78 %, 90.17%, 99.43%, and 99.58% [5].

As a result, the proposed model proved to be more effective in intrusion detection than other studies.

Table II compares the detection efficiency of various algorithms using classifier performance measures such as the area under the curve, F1, precision, and recall. The table also shows the normal and anomaly classifications for the network intrusion detection dataset. The AdaBoost supervised machine learning algorithms consistently produce a score of 1.000 for both normal and anomaly attacks. For AUC, F1, precision, and recall separately, the kNN model yielded 1.000, 0.993, 0.994, and 0.992 for normal attacks and 1.000, 0.992, 0.991, and 0.993 for anomaly attacks. Furthermore, Nave Bayes achieved 0.982 AUC for both normal and anomaly attacks, 0.925 and 0.903, F1 for normal and anomaly attacks, and 0.876 and 0.975 for precision for normal and anomaly attacks, respectively. In terms of recall, it was 0.981 and 0.840 for

normal and anomaly attacks, respectively. Finally, Neural Networks produced nearly identical results for both normal and anomaly attacks.

TABLE II. DETECTION EFFICIENCY OF DIFFERENT ALGORITHMS

| Model | Supervised Machine Learning Model | | | | |
|---|---|---|---|---|---|
| | ***Class*** | ***AUC*** | ***F1*** | ***Precision*** | ***Recall*** |
| AdaBoost | Normal | 1.000 | 1.000 | 1.000 | 1.000 |
| | Anomaly | 1.000 | 1.000 | 1.000 | 1.000 |
| kNN | Normal | 1.000 | 0.993 | 0.994 | 0.992 |
| | Anomaly | 1.000 | 0.992 | 0.991 | 0.993 |
| Naïve Bayes | Normal | 0.982 | 0.925 | 0.876 | 0.981 |
| | Anomaly | 0.982 | 0.903 | 0.975 | 0.840 |
| Neural Network | Normal | 1.000 | 0.998 | 0.997 | 0.998 |
| | Anomaly | 1.000 | 0.997 | 0.997 | 0.997 |

## V. CONCLUSIONS

The authors presented numerous supervised machine learning algorithms and feature selection methods for finding the top model in this paper. The scrutiny of the results demonstrates that the model built with AdaBoost outperformed all other models correctly classifying network intrusion, with a detection rate of 100%. The Neural Network and kNN came next, with 99.70% and 99.30% detection accuracy, respectively. On the other hand, Naïve Bayes has the lowest detection rate of 91.60%. As a result, the proposed model proved to be more effective in intrusion detection than other studies. The authors believe that their findings will help further research to develop a detection system that can detect both known and novel attacks. Today's intrusion detection systems can only detect known attacks. Due to the existing system's high false-positive rate, detecting new attacks or zero-day attacks remains a research topic.

The proposed approach can be considered in various fields, including finance, health, and transportation. Additionally, different feature selection techniques could be used, and additional parameter tuning could be added in the future to improve the classifiers' performance.

## REFERENCES


[1] Denatious, D. K., & John, A. (2012, January). Survey on data mining techniques to enhance intrusion detection. In 2012 International Conference on Computer Communication and Informatics (pp. 1-5). IEEE.

[2] Chaudhari, R. R., & Patil, S. P. (2017). Intrusion detection system: classification, techniques and datasets to implement. International Research Journal of Engineering and Technology (IRJET), 4(2), 1860-1866.

[3] Z. Muda, W. Yassin, M.N. Sulaiman, N.I. Udzir, "Intrusion Detection based on K-Means Clustering and OneR Classification", In Proceedings of 7th International Conference on Information Assurance and Security (IAS), IEEE, 2011, pp.192-197.

[4] Rung-Ching Chen, Kai-Fan Cheng and Chia-Fen Hsieh, "Using Rough Set And Support Vector Machine For Network Intrusion Detection", International Journal of Network Security & Its Applications (IJNSA), Vol 1, No 1, April 2009

[5] L. Mohan, S. Jain, P. Suyal and A. Kumar, "Data mining Classification Techniques for Intrusion Detection System," 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), 2020, pp. 351-355, doi: 10.1109/CICN49253.2020.9242642.

[6] Sharma, U. (2018, October). Association rule mining and Genetic Algorithm (GA) for Data Mining based Intrusion Detection System: A Review Approach. In 2018 3rd International Conference on Contemporary Computing and Informatics (IC3I) (pp. 249-251). IEEE.

[7] Bhosale, K. S., Nenova, M., & Iliev, G. (2018, December). DataMining Based Advanced Algorithm for Intrusion Detections in Communication Networks. In 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS) (pp. 297-300). IEEE.

[8] Gupta, D., Singhal, S., Malik, S., & Singh, A. (2016, May). Network intrusion detection system using various data mining techniques. In 2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS) (pp. 1-6). IEEE.

[9] Wankhade, K., Patka, S., & Thool, R. (2013, August). An efficient approach for intrusion detection using data mining methods. In 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1615-1618). IEEE.

[10] Limiao, Z., Hua, H., & Hao, Z. (2012, November). Research on Intrusion Detection System Model Based on Data Mining. In 2012 Fourth International Conference on Multimedia Information Networking and Security (pp. 113-116). IEEE.

[11] Xue, M., & Zhu, C. (2009, April). Applied research on data mining algorithm in network intrusion detection. In 2009 International Joint Conference on Artificial Intelligence (pp. 275-277). IEEE.

[12] Kumari, U., & Soni, U. (2017, October). A review of intrusion detection using anomaly-based detection. In 2017 2nd International Conference on Communication and Electronics Systems (ICCES) (pp. 824-826). IEEE.

[13] Macas, M., Lagla, L., Fuertes, W., Guerrero, G., & Toulkeridis, T. (2017, April). Data Mining model in the discovery of trends and patterns of intruder attacks on the data network as a public-sector innovation. In 2017 Fourth International Conference on eDemocracy & eGovernment (ICEDEG) (pp. 55-62). IEEE.

[14] Haq, N. F., Onik, A. R., Hridoy, M. A. K., Rafni, M., Shah, F. M., & Farid, D. M. (2015). Application of machine learning approaches in intrusion detection system: a survey. IJARAI-International Journal of Advanced Research in Artificial Intelligence, 4(3), 9-18.

[15] Hu, W., Hu, W., & Maybank, S. (2008). Adaboost-based algorithm for network intrusion detection. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 38(2), 577-583.

[16] Chen, F., Ye, Z., Wang, C., Yan, L., & Wang, R. (2018, September). A feature selection approach for network intrusion detection based on tree-seed algorithm and K-nearest neighbor. In 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS) (pp. 68-72). IEEE.

[17] M. Panda and M. R. Patra, "Network intrusion detection using Naive Bayes", Int. Journal of Computer Science and Network Security, vol.7, no.12, 2007, pp.258-263

[18] Debar H, Becker M, Les Ulis,"A Neural Network Component for an Intrusion Detection System", Proceedings IEEE Computer Society Symposium, 1992.

[19] Sani, Y., Mohamedou, A., Ali, K., Farjamfar, A., Azman, M., & Shamsuddin, S. (2009, November). An overview of neural networks use in anomaly intrusion detection systems. In 2009 IEEE Student Conference on Research and Development (SCOReD) (pp. 89-92). IEEE.

[20] Taher, K. A., Jisan, B. M. Y., & Rahman, M. M. (2019, January). Network intrusion detection using supervised machine learning technique with feature selection. In 2019 International conference on robotics, electrical and signal processing techniques (ICREST) (pp. 643-646). IEEE.

[21] Kaggle. (2018, October 9). Kaggle. Retrieved December 1, 2021, from https://www.kaggle.com/sampadab17/network-intrusion-detection.

[22] Livara, A., & Hernandez, R. (2022, January). An Empirical Analysis of Machine Learning Techniques in Phishing E-mail detection. In 2022 International Conference for Advancement in Technology (ICONAT) (pp. 1-6). IEEE.

[23] Macatangay, L. H., & Hernandez, R. M. (2020, August). A Deep Learning-Based Prediction and Simulator of Harmful Air Pollutants: A Case from the Philippines. In 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC) (pp. 381-386). IEEE.

[24] Hernandez, R. M., & Hernandez, A. A. (2019, October). Classification of Nile Tilapia using convolutional neural network. In 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) (pp. 126-131). IEEE.

[25] Lomboy, K. E. M. R., & Hernandez, R. M. (2021). A comparative performance of breast cancer classification using hyper-parameterized machine learning models. International Journal of Advanced Technology and Engineering Exploration, 8(82), 1080.

[26] Mendoza, A. M., & Hernandez, R. M. (2021, October). Application of Data Mining Techniques in Diagnosing Various Thyroid Ailments: A Review. In 2021 13th International Conference on Information & Communication Technology and System (ICTS) (pp. 207-212). IEEE.

[27] Hernandez, R. M., & Hernandez, A. A. (2020, February). Acceptance Analysis of Mobile Application for Nile Tilapia Classification using Unified Theory of Acceptance and use of Technology. In 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) (pp. 266-271). IEEE.

[28] Hernandez, R. M. (2021, June). Employing Technology Acceptance Model (TAM): An analysis on students' reception on online learning platforms during covid-19 pandemic. In 2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS) (pp. 58-63). IEEE.