Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Banking Intrusion Detection Systems based on customers behavior using Machine Learning algorithms: Comprehensive study

Wissam Salih Mahdi (Wissamas20@gmail.com) **Dr. Abeer Tariq Maolood(110032@uotechnology.edu.iq)**

Computer Science Department, University of Technology, Baghdad-Iraq

ARTICLE INFO

Article history:
Received: 12/08/2020
Rrevised form: //
Accepted : 27/08/2020
Available online: 19/11/2020

Keywords:

Network Intrusion Detection,
Machine Learning algorithms,
Anomaly Detection.

ABSTRACT

In recent years, The computer networks has tremendous growth and the Internet usage became essential in many fields in real life, combined with the huge amount of data transmitted over networks, generated an exponential increase in the amount of malicious and ambiguous threats to computer networks. By implementation of Machine Learning (ML) algorithms to protect computer networks and to overcome network security breaches. Many approaches appeared to the surface to achieve that purpose, one of them is the Network Intrusion Detection System (NIDS). This research aims to present a comprehensive study about employing machine learning algorithms because of what they have of effective and productive characteristics and capabilities when used in the area of tracking user behavior embedded to construct a framework simulates a banking system and examines the deviation of customer's normal utilization of the system, that is called intrusion. The experimental results obtained by this research that involves combining four algorithms in one framework showed high accuracy and low false alarm detection rate.

MSC : 30C45 , 30C50

DOI : https://doi.org/10.29304/jqcm.2020.12.4.711

1. Introduction

Online banking networks had several vulnerabilities related to network interactions in the case of certain intruders in a network or related to network overload. Such anomalous events influence the usual operation of the Bank's services. Malicious network behavior may be defined by various parameters, such as the type of network data to be evaluated, the level of network traffic and the nature of actions of the customer accessing the network. IDS is the mechanism of traffic inspection over the network and analyzing them to determine whether there is intrusion or not and looking for possible event that can be considered as a suspicions event which breach the security policies of computer system and network. Intrusion refers to an intruder's unauthorized use targeted at breaching the confidentiality, integrity and availability of network elements and attempting to disrupt the network protection policy.

*Corresponding author : ***Wissam Salih Mahdi ,Dr. Abeer Tariq Maolood***

Email addresses: ***110032@uotechnology.edu.iq***

Communicated by : ***Alaa Hussein Hamadi***

IDS has three main components [1]: Firstly, data preprocessor, Second part is the analyzer, and lastly comes the response engine.
a. Data Preprocessor is for making a decision through the process of delivering data by supplying and collecting it in a specific format and then converting it to a new format this format must be understandable by the next part (analyzer).
b.Analyzer (Intrusion detector): which is considered as the core component of IDS, its mission is to accomplish the analyzing of the audit data in order to detect the attacks.
c.Response Engine : has the job of determining the action taken when the analyzer find the event is considered as an intrusion and sets the mechanism for reaction.

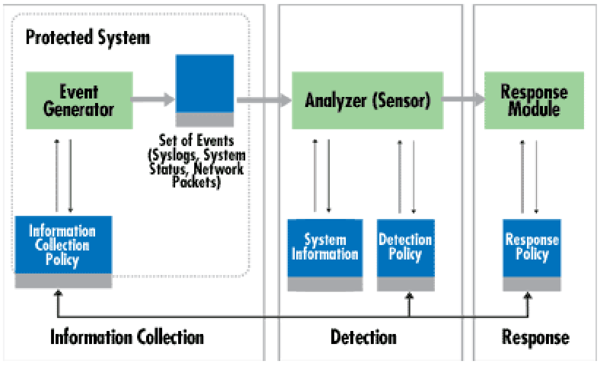


Figure (1): Intrusion detection system components

The Classification of Intrusion Detection system range into two categories: They are signature-based detection and anomaly detection, Signature based misuse detection technique depends on the known unauthorized users only. [2]
Anomaly detection techniques attempts to define normal activity profiles when intrusions are detected in the case deviation of actual system behavior from the authorized user profile. [3]

Machine learning the special branch of artificial intelligence work by predicting and acquiring knowledge derived from the process of training the data using bases of known facts. Machine learning based IDSs using the ability to develop a system that can learn about data and knowledge and make the decision on unknown data. Machine learning algorithms can be either a Supervised or Unsupervised. [4]

Supervised machine learning algorithms work by minimizing some relevant error measurement. With the goal to have the ability to train the model to be used on other sets of data with unknown correct answers. Two main categories with the supervised learning they are: Regression, where when you try to find the best convenient curve between a set of points thus you have a continuous output, and Classification, which divide data points into different classes and thus have a discrete output.

Unsupervised machine learning algorithms works with no labeled training data. Unsupervised learning algorithms depends on similarity between objects to group them together in clusters. The no need for labeled data is a big advantage for unsupervised learning, but it has the disadvantage of making evaluations of the model more difficult than supervised learning, since there is no absolute error measurement and there is no good approximate error measurement for the given application.

This paper is organized as follows. Section 2 will present the most related works to our research, Section 3 provides an overview of machine learning algorithms used in this system: Decision tree, Naïve Bayes, K-Nearest Neighbor, and Artificial Neural Network, Section 4 describes the proposed method that employs the ML algorithms mentioned above, section 5 is about performance measurement used for the experiment, and section 6 considers the dataset used for experiments. Section 6 includes a comparison between the given methods, and section 7 will contains the Conclusion and discussion for future research.

2. Literature Review

In [5], Guojun Z. et al. a cooperative IDS is presented by their work. The system consists of four parts: data flow tracking and analysis, capturing packets and rules matching, disaster recovery, and blocking. The technique of cooperative ID is introduced into the system for realizing the coordination control among parts. The system has a perfect detection rating. In [6], K.Ho Law, and L.For Kwok, they used KNN classifier as a filtering method to differentiate the normal from abnormal points in their example for IDS alarms that significantly reduce false alarm load. Ghosh and others [7] employed artificial neural network techniques to learn program behavior profiles with system call sequences for the 1998 DARPA BSM data. More than 150 program profiles were established. For each program, a neural network was trained and used for anomaly detection. Their Elman recurrent neural networks were able to detect 77.3% of all intrusions with no false positives, and 100% of all attacks with about 10% mis-classified normal sessions. In [8], S. Andropov, A. Guirik, M. Budko, and M. Budko, proposed a system composed of two capacities: offline and online traffic analysis, with the offline a model of normal behavior for the network can be created. It also provides information about different anomalies. Implementation of offline analysis requires a dataset of normal network behavior for a certain period. And with the online analysis which is the primary mode for the system. Data from the Netflow protocol is received, processed and stored by the system; an artificial neural network in real time analyzes different aspects of it. If an anomaly is detected, a warning will be issued along with a report containing information about the incident. An overview a bout User Behavior based Intrusion Detection System is presented by [9], Z. S. Malek, and B. Trivedi, they presented a rich information belonging to the intrusion detection system and its classification with the related advantages and disadvantages of each class with the architecture of the IDS system and a table of its types of alerts.

3. Machine Learning Algorithms:

3.1. Decision Tree algorithm has the task of building a classifier based on many already defined instances to predict the value of a target class for an unknown test instance. The classification of a sample by means of a sequence of decisions using the DT, in which the current decision helps to make the next decision. That sequence of decisions is shown in a tree structure.

The main components of decision tree are nodes, leaves, and edges. Each node is labeled as having the attribute to partition the data. Every node has many edges. The edge already connect either two nodes or a node and a leaf. The sample classification advances from the root node to an appropriate end leaf node, where the end leaf node indicates the classification group. The attributes of each sample are assigned to each node, the value of each branch corresponds to the attributes and the decision values for categorizing data are labeled with Leaves. Decision tree can be extended into two categories: (i) classification tree, with a list of symbolic class labels and (ii) regression tree, with a set of class labels measured numerically. [10]

Two decision tree algorithms are commonly used they are ID3 and C4.5; They are simply built top-down. First, the tree is null, and the algorithm starts to create it from the root node by inserting internal decision nodes that hold a check loop with a particular attribute or leaf nodes that is a terminal node containing the class name. The result of that is the decision tree.

The steps bellow are done recursively [11]:

1. For each attribute, compute the information gain.

2. The highest information gain the selected attribute that represent a split attribute.
3. When the chosen attribute is distinct (categorical), all possible values are branched to the node. If the attribute is constant, a cut-point will pick the maximum information gain.
4. After splitting, that those new nodes where leaves (their data belong to the same type); otherwise, the root of the sub-trees is new nodes.
5. Repeat all the steps above until all new nodes are out.To accomplish the steps of algorithm we will use the following equations:

```
Algorithm NB
Input: A set D of training examples
Output: Naive Bayes for D

Begin
        for each value c of C
             Compute P(c) from D.
        for each attributes $A_i$
        for each assignment $a_i$ and c to $A_i$ and C
             Compute P($a_i$|c) from D
End
```

3.2 Naïve Bayes Classifier

Naïve Bayes classifier is a statistical way for classification and one of machine learning algorithms that depends on statistical. NB classifier have several properties that make it widely used due to its usefulness and accuracy. NB classifier is based on applying Bayes theorem with strong independence assumption that makes it a simple probabilistic classifier. In other words, It is assumed that the appearance (or absence) of a given attribute of a class is not linked to the existence (or absence) of certain features. Depending on the clear nature of the probability paradigm, the naive Bayes classifiers can be trained effectively in a supervised area of learning, based on the simple structure of the probability paradigm and the significant consequent computational advantages of the naive Bayes classifier are its short computational training time. [12] [13]

Two types of variables used in NB classifier: C is the class variable and X = $\{X1; X2;\ldots; Xn\}$, which are the set of features applied on dataset D which is consisting of set of instances $\{E1,E2,..,Et\}$, can be defined as : [14]

$$P(C|X) = \frac{P(X|C)\,.\,P(C)}{P(X)}$$

Where , P(C|X) corresponds to the posterior probability of class (target) given predictor (attribute) , P(C) is the probability of class, P(X|C) is the likelihood which is the probability of predictor given class , and P(X) is the prior probability of predictor . for P(C|X)=P(X1|C).P(X2|C)...P(Xn|C).P(C)

The algorithm of NB classifier is shown as in bellow. [14]

- Entropy function: (used for ID3 and C4.5)

$$Info(D) = -\sum_{i=1}^{m} P_i \log_2 P_i \qquad (1)$$

- Information gain: (First method)

$$\text{Gain (A)} = \text{Info (D)} - \text{Info}_A \text{ (D)} \qquad (2)$$

$$Info_A(D) = -\sum_{i=1}^{v} \frac{|D_j|}{|D|} \times Info\,(D_j) \qquad (3)$$

3.3 K-Nearest Neighbor Classifier:

The KNN (K-Nearest Neighbor Classifier) is a classification method used to classify objects based on the closest training examples in the feature space. KNN is a type of learning based on instance. The KNN is one of the simplest of machine learning algorithms. An object can be classified by a majority vote of its neighbors, the object being assigned to the most similar class of its k closest neighbors, where, k is a positive integer, typically small. The object is simply assigned to its closest neighbor's class in event that when the value of k was equal to one. [14]

Evenly it has some drawbacks as it takes more computational time, KNN classification algorithm can be considered as one of the efficient algorithms of machine learning techniques used in IDS. [15]

KNN classifier measures uses the Euclidean distance to measure the distance between two data points P and Q. Indeed the distance indicates the similarity between them. The shorter distance between them is the similar looks. The Euclidean distance used is shown as:

$$distance(P,Q) = \sqrt{\sum_{i=0}^{N}(p_i - q_i)^2}$$

pi and *qi* indicates the values of the i^{th} feature of points *P* and *Q*. The final similarity value of a classified data point is the sum of its Euclidean distances from the nearest normal k points. [16]

In Figure (2), suppose the point "?" namely as P is the point, for any label needs to predict. The First step is to find the one closest point to P and then assign the label of the nearest point to P.

Second step, identify the k closest point to P and then classify the k-neighbors by majority vote. The item votes for its class and the prediction is the class with the most votes. Calculate the distance between points using the Euclidean distance to find nearest similar points. [6]

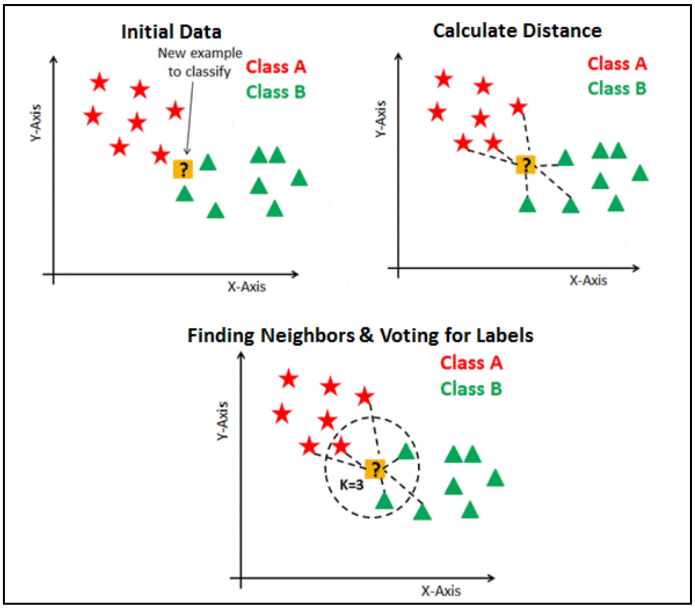


Figure (2): KNN algorithm implementation [6]

3.4 Artificial Neural Network:

Artificial Neural Network algorithm solve problems in the same way the biological neurons system perform the task of solving problems, like the brain which is the information-processing unit in human beings. The network comprises a significant number of integrated processing units known as neurons working together to solve specific problems. [17]

Classification of units in the ANN classified into three categories: input units, receiving processing information; output units, which are responsible for delivering the processing result, and the units between them are called the hidden units. Signals allowed to travel on one way from input to output that function is accomplished with Feed-forward artificial neural networks. [18]

Artificial Neural Network is a network with several specific processors, called neurons. These processors receive data either from outside the network or from inside the network (other neurons). The data goes through communication channels, which are called weights. The ANN needs to be trained (learned) in order to give the more accurate output. Learning is a process of adjusting the parameters of the ANN with a continuing process until it reaches desired values. The nature of the learning is determined by the way the parameter changes. Generally, the learning procedure is classified in to tow categorizes: supervised or unsupervised. With supervised learning, the desired output values are provided and weights through the network must be adapted to reduce the difference between the actual network output and the desired output. When the network produce adaptive output that is closely compatible with the desired output the network is said about it is trained and there is no need to be trained again.

When using unsupervised learning, the network does not have desired output values with the training set. The network is attempting to find the underlying configuration of the input data and the network needs to organize itself when there are existing specialty between the elements that form the entire sample set. [19]

Three main components of Artificial Neural Network they are: node character, network topology, and learning rules. The Node character defines how the node handles the signals, such as number of node-linked inputs and outputs, weight of each input and output node, and activation function. The network topology defines how nodes are ordered and linked. The learning rules determine whether weights are configured and only modified if there is a difference between the current and desired output. [20]

In Figure (3), a simple neural network model with one single node used in this research, where each node receives several inputs from other connections linked to the weight. When the weighted sum of inputs crosses the node-related given threshold, the signal is activated and transferred via a transfer function and transmitted to neighboring nodes. The following equation shows the weighted sum formula: [20]

$$y = f\left(\sum_{i=0}^{n} w_i x_i\right)$$

Where, y represents the node's output, f is the transfer function used, and w_i is the related weight of input x_i . The activation function is the step function: [20]

$$y = \begin{cases} 1 & if \sum_{i=0}^{n} w_i x_i >= 0 \\ -1 & if \sum_{i=0}^{n} w_i x_i < 0 \end{cases}$$

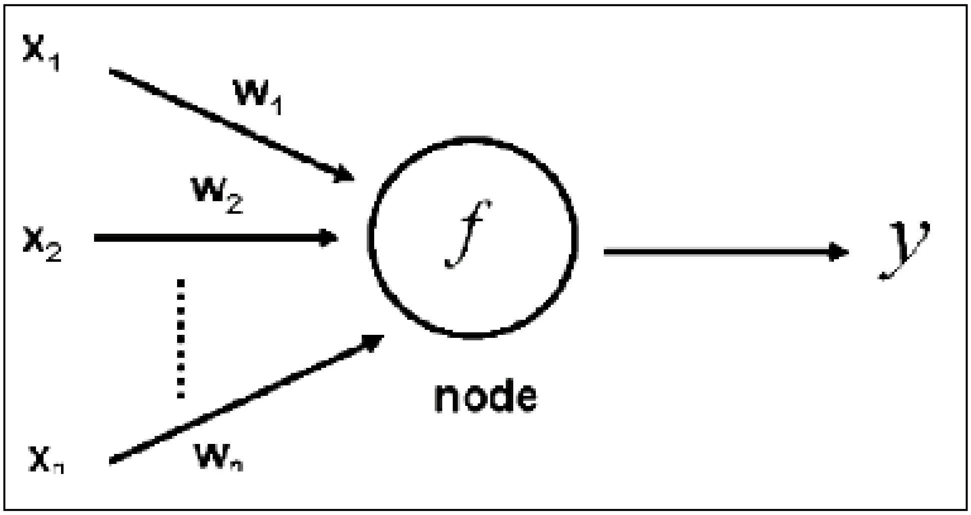


Figure (3): simple neural network model with one single node: {x_1,x_2,… , x_i} are the inputs , {w_1,w_2,… , w_i} are the weight related to the inputs, f is the transfer function, and y = output [20]

4. proposed system:

This paper talking about combined system composed of four machine learning algorithms working together to find out anomalous behavior derived from a particular dataset, this system applied these algorithms in a subsequent way where each algorithm perform a particular task, this task is considered as a stages of checking for a specific aspect that complete a comprehensive checking framework (system).

The system is composed of four main steps, each one of the first three steps perform a specific check on the dataset and the result of each check is taken as input data to one of the three nodes of the forth algorithm (ANN), Which is in turn perform its job to get accurate result of the whole checking platform (system). Figure (4) shows an overview of the proposed system.

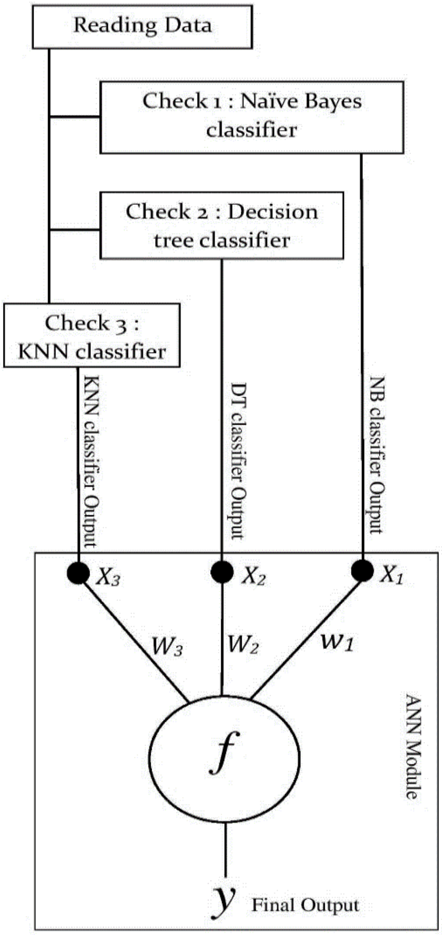


Figure (4): overview of the proposed system

5. Dataset

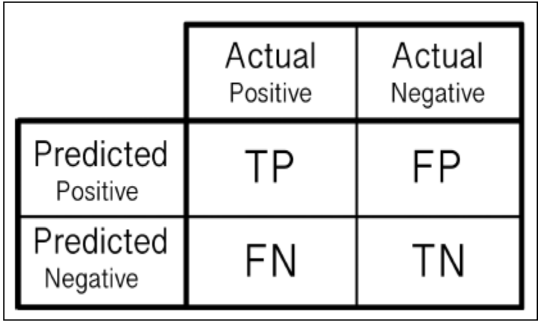
The Dataset used in this system is virtual, hand built and imaginary contains a banking system data for customers that they are getting two types of services (Deposit and withdraw), the deployment of these services in the dataset in a regular distribution over the cells of each row in the dataset, The columns are divided into the services mentioned along 12 months and each column represent a month has two fields, First one contains the name of the service and the other contains the amount of money deployed in that month.

The dataset Consist of 2324 row, each row represent a single user, The row shows the services done by the intended user and the money amount deposited or drawled in over one year, The system has the job of tracking the behavior of that user along the months to discover the consequence of that behavior and employ machine learning algorithms on that exposed data and state whether that user has anomalous or normal behavior.

6. Performance measures and results:

Evaluation of performance of an implementation has many of aspects. Time is one of important aspects; such as the time it takes to train the model, the time it takes to answer a query and so on. Another aspect of performance is how good the predictions of the model are. Many measures are suitable for use that are depending on the style of learning that the algorithm follows and the main purpose of it. The confusion matrix, which is often called the Error matrix, lists a set of instances where a classification process correctly or incorrectly predicts, the matrix is N×N, and N refers to the number of categories. The Confusion Matrix is shown in table (1). [21]

Table (1): Confusion matrix for binary classification



	Actual Positive	Actual Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

The tabulated terms in the confusion matrix are explained in the following terminologies.

1. True Positive (TP), corresponds to the accurate results that the classification model accurately predicts.
2. False negative (FN), corresponds to the positive results incorrectly predicted by the classification model as negative.
3. False positive (FP), corresponds to the negative examples incorrectly predicted by the classification model as positive.
4. True negative (TN), associated with negative instances where the classification model correctly predicts.

The **accuracy** can be measured by using data from the confusion matrix by dividing the number of values that the algorithm got correct by all of the predicted values:

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

Recall is the percentage of true positive results to the all results in the actual class,

$$\textbf{Recall} = \frac{\textbf{TP}}{\textbf{TP + FN}}$$

Precision may be defined as the ratio of correctly predicted positives to the overall predicted positive observations. i.e.,

$$\textbf{Precision} = \frac{\textbf{TP}}{\textbf{TP + FP}}$$

Table (2): results of Naïve Bayes

	accuracy	precision	recall
Normal	0.84	0.88	0.85
Anomaly	0.90	0.79	0.92
Total	0.87	0.83	0.88

Table (3): results of Decision Tree

	accuracy	precision	recall
Normal	0.89	0.92	0.90
Anomaly	0.93	0.90	0.86
total	0.96	0.91	0.88

Table (4): results of KNN

	accuracy	precision	recall
Normal	0.86	0.83	0.91
Anomaly	0.94	0.79	0.79
total	0.91	0.80	0.85

Table (5): results of ANN

	Accuracy	precision	recall
Normal	0.96	0.89	0.87
Anomaly	0.91	0.94	0.96
total	0.93	0.91	0.91

8. Conclusion and Discussion

Machine-learning algorithms has been extensively used in the field of network anomaly detection, in this work, we proposed multi-stage checking system and applied machine-learning algorithms to obtain the anomaly based network intrusion detection. These stages are applied consecutively on our dataset to check the anomalous behavior of the user. This method we proposed and the experimental results showed the efficiency of the system in the mission of detection and distinguish attack and define normal behavior with certainty guaranteed high detection average (98%) and low false alarm.

References:

[1] Kumar A., Maurya H. C. and Misra R., "A Research Paper on Hybrid Intrusion Detection System", International Journal of Engineering and Advanced Technology (IJEAT) Vol. 2, Issue-4,

[2] Chih-Fong Tsai, Yu-Feng Hsu , Chia-Ying Lin , Wei-Yang Lin, "Intrusion detection by machine learning: A review," Expert Systems with Applications 36 ,2009.

[3] 2013Zakiyabanu S. Malek, and Bhushan Trivedi "User Behaviour based Intrusion Detection System Overview", International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2018.

[4] Nutan F. H. , Musharrat R. , Abdur R. O. , Faisal M. S. , Md. Avishek K. H. , and Dewan Md. F. "Application of Machine Learning Approaches in Intrusion Detection System: A Survey", International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No.3, 2015.

[5] Guojun Z., Liping C. and Weitao H., "The Design of Cooperative Intrusion Detection System", IEEE Computer Society, Seventh International Conference on Computational Intelligence and Security, 2011.

[6] K.Ho Law, and L.For Kwok, "IDS False Alarm Filtering Using KNN Classifier", Springer-Verlag Berlin Heidelberg, 2004.

[7] A. K. Ghosh, A. Schwartzbard and A. M. Shatz, "Learning Program Behavior Profiles for Intrusion Detection", Proceedings of 1st USENIX Workshop on Intrusion Detection and Network Monitoring, Santa Clara, 1999.

[8] S. Andropov, A. Guirik, M. Budko, and M. Budko, "Machine learning: a review of classification and combining techniques" Springer Science+Business Media B.V., 2007.

[9] Zakiyabanu S. Malek, and Bhushan Trivedi "User Behaviour based Intrusion Detection System Overview", International Journal for Research in Applied Science & Engineering Technology, 2018.

[10] S. B. Kotsiantis , I. D. Zaharakis , P. E. Pintelas "Machine learning: a review of classification and combining techniques" Springer Science+Business Media B.V., 2007.

[11] Dewan M., Farid N., H. Emna B., Mohammad Z. R., Chowdhury M. R., "Attacks Classification in Adaptive Intrusion Detection using Decision Tree", World Academy of Science, Engineering and Technology , 2010 .

[12] Dewan M., Farid N., H. Emna B., Mohammad Z. R., Chowdhury M. R., "Attacks Classification in Adaptive Intrusion Detection using Decision Tree", World Academy of Science, Engineering and Technology , 2010 .

[13] Hafsa A. M., " Modified Hidden Naïve Bayes Classifier to Enhance Intrusion Detection system," thesis submitted to Department of Computer Sciences of University of Technology in partial fulfilment of requirements for the degree of Master of Science in Computer Science, 2016.

[14] S.V. Lakshmi, and T.E Prabakaran, "Application of k-Nearest Neighbour Classification Method for Intrusion Detection in Network Data", International Journal of Computer Applications (0975 – 8887), 2014.

[15] B.Basaveswara , and K.Swathi, "Fast kNN Classifiers for Network Intrusion Detection System", Indian Journal of Science and Technology, Vol 10(14), 2017.

[16] Wenchao Li, Ping Yi, Yue Wu, Li Pan, and Jianhua Li, "A New Intrusion Detection System Based on KNN Classification Algorithm in Wireless Sensor Network", Journal of Electrical and Computer Engineering , 2014.

[17] I.N. da Silva et al., "Artificial Neural Networks", Springer International Publishing, 2017.

[18] Sufyan T. Faraj Al-Janabi and Hadeel Amjed Saeed," A Neural Network Based Anomaly Intrusion Detection System", IEEE, 2011.

[19] S. B. Kotsiantis , I. D. Zaharakis , and P. E. Pintelas," Machine learning: a review of classification and combining techniques", Springer, 2007.

[20] J Zou , Y Han , and Sung-Sau So," Artificial Neural Networks Methods and Applications", Springer, 2008.

[21] Samuel J. and Karol W., "Machine learning algorithms in a distributed context", Bachelor thesis submitted to the Department of Computer and Information Science of Linköping University, 2018.