USAGE OF SUPERVISED MACHINE LEARNING TECHNIQUE WITH FEATURESELECTION IN NETWORK INTRUSION DETECTION

Mr.V.Shiva Prasad[1]., Mr.Y.Praveen Konda Reddy[2]
1,2 Assistant Professor, Department Of CSE.,
(✉vadlashivaprasad @gmail.Com, ✉y.kreddy105@gmail.Com)
1,2 Malla Reddy College Of Engineering For Women., Maisammaguda., Medchal., Ts, India

Abstract –

In order to determine whether or not network traffic is malicious or benign, an innovative supervised machine learning method has been built. In order to locate the ideal model taking detection rate into account rate, supervised learning algorithm, and feature selection approach have all been utilized in this process. Based on the findings of this research, it has been determined that the Artificial Neural Network (ANN) strategy of machine learning with wrapper feature selection performs better than the support vector machine (SVM) method when it comes to categorizing network traffic. The NSL-KDD dataset is used to classify network traffic using supervised machine learning methods such as SVM and ANN. This is done so that the performance may be evaluated. According to the findings of the comparative analysis, the suggested model is more effective than other models already on the market in terms of the success rate of intrusion detection.

I. INTRODUCTION

The number of people using the internet and the number of people who have access to online information is both rising, which has led to a rise in the frequency of cybercrime [1-2]. The initial stage in protecting a network is called intrusion detection. Ward off an assault on the security. Studies are paying a lot of attention to various security solutions, such as firewalls, intrusion detection systems (IDS), unified threat modeling (UTM), and intrusion prevention systems (IPS). IDS is able to identify assaults coming from a wide range of computer systems and networks because it gathers information and then examines that data in search of potential vulnerabilities [3]. The examination of the data packets that go around a network is performed in two different methods by the network-based intrusion detection system (IDS). Even up to this day, anomaly-based detection is light years behind detection that operates on the basis of a signature; as a result, anomaly-based detection continues to be a significant focus of study [4-5]. The fact that it must be able to handle new attacks for which there is no previous information to identify the anomaly is one of the difficulties associated with intrusion detection that is based on anomalies. Because of this, the system in some way has to have the intelligence to distinguish between communication that is safe and traffic that is malicious or abnormal, and researchers have been looking into various machine learning approaches in order to accomplish this goal over the last several years [6]. On the other hand, IDS is not a panacea for all issues pertaining to security. IDS, for instance, is unable to compensate for

Insufficient identification and authentication techniques or for deficiencies in network protocols when there is such a deficiency.
1980 saw the beginning of serious research into the topic of intrusion detection, and 1987 saw the publication of the first model of its kind [7].
Over the course of the last several decades, despite enormous commercial despite significant expenditures and research efforts, the technology used for intrusion detection is still in its infancy and, as a result, is ineffective [7]. Anomaly-based network intrusion detection systems have not achieved the same level of commercial success as signature-based network intrusion detection systems, despite the fact that signature-based network intrusion detection systems have been widely adopted by technology-based businesses all over the world. Anomaly-based detection is now a key emphasis area for research and development in the field of intrusion detection systems (IDS), due to the aforementioned rationale [8]. And important problems still need to be resolved before any large-scale implementation of an anomaly-based intrusion detection system [8]. However, there is only a small amount of research available now that compares how well intrusion detection works with supervised machine learning approaches [9]. This is a significant limitation.

The anomaly-based network intrusion detection system (IDS) is a useful technology that may safeguard target systems and networks from hostile actions. In spite of the many different kinds of anomaly-based Anomaly detection features enabled security systems are only starting to surface, and several significant challenges still need to be resolved. Network intrusion detection strategies have been published in the literature in recent years [8]. Linear Regression, Support Vector Machines (SVM), Genetic Algorithm, Gaussian mixture model, nearest neighbor algorithm, Naive Bayes classifier, and Decision Tree are only few of the anomaly-based methods that have been developed [3,5]. SVM is the sort of learning algorithm that is used the most often among those three since it has previously shown its worth on a variety of problems [10]. Although all of these suggested approaches are capable of detecting unique threats, in general they all suffer from a high percentage of false alarms. This is one of the most significant problems with anomaly-based detection. The reason for this is due to the difficulty of developing profiles of typical, practical behavior via the process of learning from data sets used for training [11]. The back propagation technique, which has been available since 1970 and serves as the opposite mode of automated differentiation [12], is one of the most common methods used to train artificial neural networks (ANN) in use today.

The absence of a complete network-based data collection is one of the most difficult obstacles to overcome when conducting an evaluation of the effectiveness of network IDS [13]. The vast majority of the suggested irregularity utilizing the KDD CUP 99 dataset [14], based strategies discovered in the literature were examined. In this article, we used the machine learning methods of support vector machines (SVM) and artificial neural networks (ANN) to the widely used benchmark dataset for network intrusion known as NSLKDD [15].

II. SYSTEM MODEL

The suggested system is made up of the feature selection algorithm and the learning algorithm that is shown in Fig. 1. Feature selection components are in charge of extracting the characteristics that are the most relevant. Or properties that allow a certain class or group to be identified with the instance. The output of the work done by the component responsible for feature selection is used by the learning algorithm component to construct the requisite intelligence or knowledge. The model's intelligence is developed via the process of training, which is carried out with the help of the training dataset. After that, the newly acquired intelligences are put to the testing dataset in order to see how well the model can classify information that it has not previously been shown.

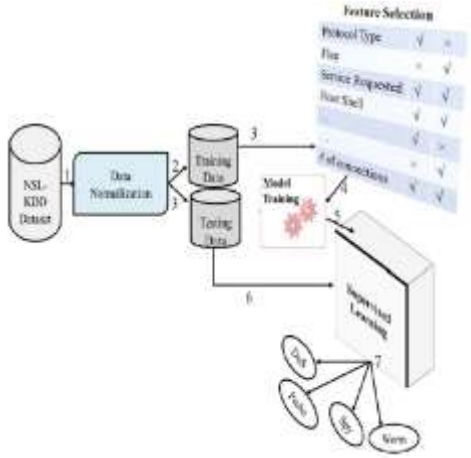


Fig 1: Proposed supervised machine learning classifier System

A. The Choice of Characteristics

The process of selecting features is an essential component of machine learning, as it helps to minimize the extensiveness and dimensionality of the data. Investigation labored over in pursuit of a trustworthy approach for feature selection. Both the filter approach and the wrapper method have been used throughout the feature selection process. When using the filter technique, features are chosen based on how well they performed in a variety of statistical tests that evaluate the significance of features based on the degree to which they correlate with the dependent variable or the outcome variable. The utility of a subset of a feature is evaluated in relation to the dependent variable in order for the Wrapper technique to locate a subset of features. Therefore, filter techniques are

not reliant on any particular machine learning algorithm, while the optimal feature subset that is picked when using a wrapper method is determined by the machine learning approach that was used to train the model.

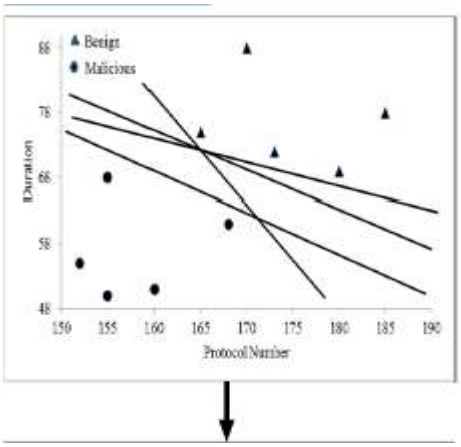
In the wrapper technique, a subset evaluator utilizes all of the potential subsets before turning to a classification algorithm to persuade classifiers based on the characteristics included inside each subset. The classifier takes into account the particular subset of features with which the classification method functions most effectively. The evaluator may use a variety of search strategies, including depth first search, random search, breadth first search, or hybrid search, in order to locate the subset.

1. Choosing Which Attributes to Focus On A.

Because it helps to reduce the extensiveness and dimensionality of the data, the process of picking features is a crucial component of machine learning. Inquiry that took a lot of time and effort in order to find a reliable method for feature selection.

Throughout the whole of the process of selecting features, the filter technique in addition to the wrapper method has both been used. When utilizing the filter technique, features are selected on the basis of how well they performed in a variety of statistical tests that evaluate the significance of features based on the degree to which they correlate with the dependent variable or the outcome variable. For example, if a feature had a strong correlation with the dependent variable, it would be selected for use in the filter technique. In order for the Wrapper approach to identify a subset of features, the utility of a subset of a feature is assessed in respect to the dependent variable. Therefore, filter techniques are not dependent on any specific machine learning algorithm, whereas the optimal feature subset that is chosen when utilizing a wrapper method is determined by the machine learning approach that was used to train the model. This is in contrast to the situation in which a wrapper method is dependent on the machine learning approach that was used to train the model. A subset evaluator will use all of the possible subsets in the wrapper approach before moving on to a classification algorithm in order to convince classifiers based on the features that are included inside each subset. The classifier takes into consideration the specific subset of characteristics with which the classification technique performs at its optimum level. In order to discover the subset, the evaluator may use a number of different search algorithms, such as depth first search, random search, breadth first search, or hybrid search.

The Support Vector Machine, or C. (SVM) when using SVM, the classifier is determined by a separating hyper plane, and this plane's characteristics change based on the nature of the issue and the datasets that are available. If the dataset is one dimension, the hyper plane is a point; if it is two dimensions, it is a dividing line, as illustrated in Figure 2; if it is three dimensions, it is a plane; and if it is higher than three dimensions, it is a hyper plane. The classifier or decision function will take the form for a dataset that can be linearly separated if it is linearly separable.



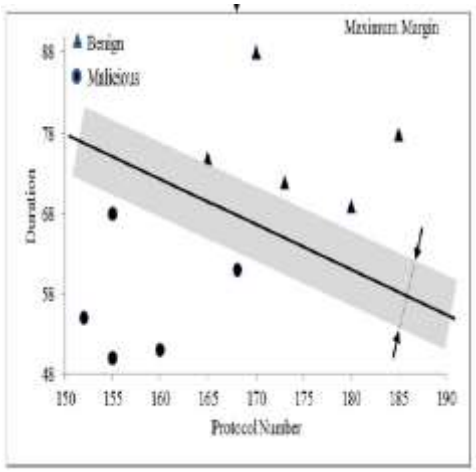


Fig 2: SVM classifier in two dimensional problem spaces

$$ax + by + c = 0 \qquad (1)$$

If axe + by is less than c, the aforementioned judgment function will categories the point as belonging to one class; alternatively, it will categories the point if axe + by is more than c. The equation of a line, which may be written as y = axe + b should read as follows: y + axe + b = 0 something may be represented by using two vectors in the following fashion:

$$\mathbf{w}\begin{pmatrix} -b \\ -a \\ 1 \end{pmatrix} \text{ and } \mathbf{x}\begin{pmatrix} 1 \\ x \\ y \end{pmatrix} \qquad (2)$$

Which says we can write the linear equation of a line using two vectors as below?

$$\mathbf{w}^T\mathbf{x} = (-b)\times(1) + (-a)\times x + 1\times y, \text{ or}$$
$$\mathbf{w}^T\mathbf{x} = y - ax - b \qquad (3)$$

It is simpler to deal in more than two dimensions with this notation, which is why the hyper plane equation wax is used instead of the equation y=ax. Additionally, the vector w will always point in the same direction. Maintain your normalcy relative to the hyper plane. When the hyper plan that provides the greatest margin has been identified, it is possible to utilize this hyper plane to generate predictions [11]. The results of the hypothesis function h are going to be-

$$h(x_i) = \begin{cases} +1; \ if \ \ \mathbf{w}.\mathbf{x}+b \geq 0 \\ -1; \ if \ \ \mathbf{w}.\mathbf{x}+b < 0 \end{cases} \qquad (4)$$

D. Artificial Neural Network (ANN)

Another technology that may be used in machine learning is called an artificial neural network. As its name indicates, artificial neural networks (ANN) are learning systems that are inspired by the human brain system and duplicate human learning. Human cognitive processing system in the majority of instances, it comprises of input and output layers in addition to one or more hidden levels, as illustrated in Fig 3. Back propagation is the method that the ANN employs in order to alter the outcome so that it corresponds with the intended result or class.

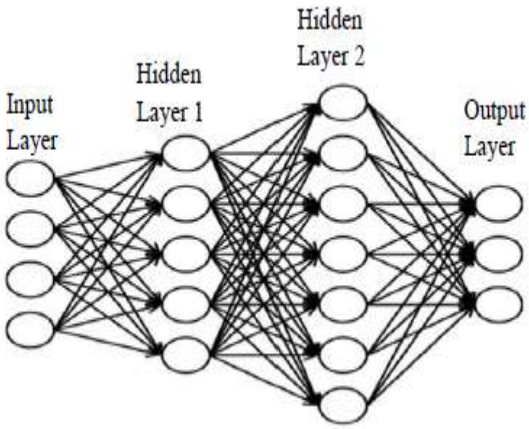


Fig 3: Artificial neural network showing the input, output and hidden layers

III. EXPERIMENTAL ANALYSIS OF THE SYSTEM

A. Feature Selection

The experiment, which is divided into two phases and was carried out with the help of the well-known data mining and machine learning open source software suite known as Weka, in the first section, we were able to get the majority of important characteristics by using a variety of techniques for feature selection (FS). To prevent issues with over fitting and under fitting, the wrapper approach that we developed made use of the SVM classification algorithm in conjunction with cross-validation. During the filter procedure, a ranker algorithm is utilized to determine which result is most appropriate for the classifier that we have provided. The NSL-KDD dataset that we used for training purposes has a total of 25,191 occurrences that have been tagged.

Table I displays the findings obtained from the experiment involving feature selection.

TABLE I

RESULT OF FEATURE SELECTION

FS Technique	FS Type	Input Features	Output Features
Correlation Based	Wrapper	41	17
Chi-Square Based	Filter	41	35

The correlation-based feature selection method identified a total of 17 features from the training dataset's 41 features as being the most important, whereas the Chi-Square approach kept 35 of the original features. Characteristics to be more relevant to the class that was ultimately produced. These 17 and 35 retained characteristics were used to train the model on the training dataset, also known as the dataset that was seen. They were also used to test the model on the dataset that was not seen, also known as the testing dataset.

B. Classification

Using the training dataset, a total of four models are constructed in the Weka software package based on the characteristics that were discovered in the feature selection portion. Before supervised machine learning can be used for classification, the model has to be trained with the help of a training dataset. As training data, we used 20% of the NSL-KDD dataset, which contains 25,191 labeled data instances in total. For the training of the model, we used SVM and ANN learning algorithms for the various kinds of feature selection methods.

As a result, we develop four different learning models, two of which are based on SVM and the other two on ANN. The first model that was constructed for each individual learning method was the One is constructed with the use of 17 features, while the other is constructed with the use of 35 features discovered in the feature selection portion. Following that, these four trained models were assessed making use of 22,542 instances of testing data chosen from the NSL-KDD testing dataset.

The results are shown in Table II with the following summary:

TABLE II
RESULT OF CLASSIFICATION

Learning Type	Number of Features	Detection Accuracy
SVM	17	81.78%
SVM	35	82.34%
ANN	17	94.02%
ANN	35	83.68%

In Table III, we compared our findings to those that had been published not too long ago in the relevant academic literature. When evaluating the effectiveness of the suggested model in comparison to that of other works, we chose certain works. Possessing a notion of common characteristics connected to benchmarking datasets and learning algorithms. But there are also additional considerations, such as the reduction of attributes, the number of instances, the number of layers, and the learning rates that are applied. The success rate of detection achieved by the proposed model is compared with that of other models already in existence in Table III, as shown below-

TABLE III
PERFORMANCE COMPARISON WITH EXISTING MODELS

Learning Type	Our Model Accuracy	Existing Model	Existing Model
SVM	82.34%	92.84% [16]	69.52% [17]
ANN	94.02%	81.2% [18]	77.23% [19]

IV. DISCUSSION ON SYSTEM IMPLEMENTATION

We have used the highly popular open source machine learning software package known as Weka in order to both develop and evaluate the system. In addition to the algorithm for machine learning Weak not only has numerous algorithms and search techniques built, but it also has these things implemented to do feature selection. We performed several tests using the ANN model with a variety of various numbers of hidden layers, and we discovered that the detection success rate is dependent on the number of hidden layers. Following a number of iterations of trial and error, we discovered that the optimal detection rate could be achieved with three hidden layers and a learning rate of 0.1. In addition to using the SVM algorithm as a classifier, we also employed the wrapper feature selection approach. The model that was built in Weak and tested was executed on a computer platform that had a 64-bit Intel core i5 CPU operating at 2.6 GHz, 8 GB of RAM, Windows 7 as the operating system, and restricted instances of network traffic. In order to implement the solution on a big scale network, extra infrastructure with some kind of greater capacity server platform will be required.

V. CONCLUSION

In order to determine the most effective machine learning model, we have provided a variety of machine learning models in this work. Each model makes use of a unique machine learning algorithm and a unique strategy for feature selection. The results of the investigation reveal that the model that was constructed using ANN and wrapper feature selection fared the best out of all the other models when it came to accurately identifying the network traffic. The detection rate for this model was 94.02%. We have every reason to assume that these discoveries will help to future study in the field of developing a detection system that is capable of detecting both well-known and newly discovered threats. The current generation of intrusion detection systems can only identify previously known

assaults. Due to the high false positive rate of the present methods, detecting new attacks or zero day attacks is still something that has to be researched further.

REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.
[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.
[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, 2015, pp. 513–517.
[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.
[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.
[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.
[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.
[8] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.
[9] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016.
[10] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," Neural Computing and Applications, vol. 22, no. 5, pp. 1023–1035, 2013.
[11] F. Gharibian and A. A. Ghorbani, "Comparative study of supervised machine learning techniques for intrusion detection," in Communication Networks and Services Research, 2007. CNSR'07. Fifth Annual Conference on, 2007, pp. 350–358.
[12] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.
[13] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in Military Communications and Information Systems Conference (MilCIS), 2015, 2015, pp. 1–6.
[14] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," in Industrial Electronics (ISIE), 2017 IEEE 26th International Symposium on, 2017, pp. 1881–1886.
[15] L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446–452, 2015.