MACHINE LEARNING TECHNIQUES FOR DESIGN OF INTRUSION DETECTION SYSTEM FOR BIG DATA NETWORKS

Abstract

As digital technology advances, gigabytes and terabytes of data are now generated every second. Businesses in a variety of industries are finding that using the internet to manage their resources and transactions is useful. Given the value of data and the need to safeguard its security and privacy, securing big data remains a major challenge for all solutions. Due to the exponential expansion of network data, intrusion detection is becoming increasingly important, and manual analysis would be either impossible or take the same amount of time as analysing it. As a result, there is an urgent need for an automated system capable of extracting relevant information from enormous amounts of hitherto untapped data when it comes to network intrusion detection. Data mining can perform a variety of tasks, such as clustering, prediction, classification, and the extraction of association rules between data pieces. This paper discusses machine learning techniques for designing intrusion detection systems for big data networks. In this approach, the NSL KDD data set is used as input. First, the CFS-correlation feature selection approach is used to pick only relevant features from the NSL KDD data set. The NSL KDD data collection contains 41 features. The number of characteristics was reduced to 16 after applying the CFS algorithm. The 16 attributes are then used by machine learning techniques to classify and predict malware data in the NSL KDD data set.

Keywords: Intrusion Detection, Big Data, machine Learning, LS SVM, CFS Feature Selection, Accuracy

Authors

Dr. Sammy. F
Assistant Professor
Department of CSE
Koneru Lakshmaiah Education Foundation
Vaddeswaram, Andra Pradesh, India.

Dr. Meenakshi
Assistant Professor
Apeejay Stya University
Sohna, Haryana, India.

Dr. Umesh Kumar Singh
Director
Institute of Computer Sciences
Vikram University
Ujjain, India.

Dr. Malik Jawarneh
Faculty of Computing Sciences
Gulf College
Oman.

Dr. Abhishek Raghuvanshi
Professor
Department of computer Science & Engineering
Mahakal Institute of Technology
Ujjain, India.

I. INTRODUCTION

Because of advancements in digital technology, gigabytes and terabytes of data are now being generated every second. Utilizing the internet to manage their resources and transactions is proving to be beneficial for businesses across a wide range of industries. Given the importance of data and the need of protecting its security and privacy, securing big data continues to be a key challenge for all solutions. Undoubtedly, one of the most complicated security challenges connected to network traffic data is the ability to detect and avoid network assaults while maintaining high accuracy and minimal prediction time. These incursions jeopardise the security, integrity, and availability of big data resources and services, among other things [1].

The firewall is the first line of defence, but it can only sniff packets that are coming from outside the network (from intruders), not packets that are arriving from within the network. Because of the complexity of intruders, a standard firewall is unable to identify a large number of them. However, the traditional firewall is not a reliable technique of protecting against all types of attacks. In order to provide an effective solution for these difficulties, a high-speed intrusion detection system that can work in a big data environment and handle a significant quantity of network traffic at the same time is required [2].

According to CERT, a computer emergency response team that monitors network incursions, the amount of data on the Internet is growing at an exponential rate. When an intruder breaches network services, it results in not only a financial loss, but also a violation of the security requirements of both organisations and individuals. Malware, access control systems (such as firewalls), email security, intrusion prevention systems (IPS), and intrusion detection systems (IDS) concepts do not provide adequate data security and protection [3] [4].

Due to the exponential growth of network data, intrusion detection is becoming increasingly crucial, and manual analysis would be practically impossible or would take the same amount of time as it would take to analyze it. As a result, when it comes to network intrusion detection, there is a pressing need for an automated system that can extract useful information from a vast amount of hitherto untapped data. Data mining is capable of a wide range of tasks, including clustering, prediction, classification, and the extraction of association rules between data elements.

As a result, intrusion detection systems based on data mining are the most effective encapsulation of the technical foundation for combating the wide spectrum of network traffic incursions that exist. All of these attacks have significant repercussions for the environment in which they occur. To prevent these assaults from occurring in the first place, it is best to categorise them using the most widely used categorization methodologies from the outset, so that the attacker can be prevented. As a result, an effective intrusion detection system, which can detect intrusions before an attack can take place and inform users that an attack may take place, can help to achieve this goal [5].

This article presents a machine learning techniques for design of intrusion detection system for big data networks. NSL KDD data set is used as input in this framework. First of all, CFS- correlation feature selection method is applied on the NSL KDD data set to select

only relevant features. NSL KDD data set has 41 features. On applying CFS algorithm, numbers of features were reduced to 16. Then machine learning algorithms are applied on these 16 features to classify and predict malware data in NSL KDD data set.

II. LITERATURE SURVEY

The methods of intrusion detection used in a variety of industries are covered in this chapter. Detecting cyberattacks in large volumes of data is a problem that many academics have attempted to solve in the academic literature over the years. When it comes to dealing with computer security risks, the scientific community has offered a diverse spectrum of intrusion detection techniques.

Data pieces, which may be characterized as either categorical or numerical in nature, contain a diverse range of features that can be accessed. A nominal, ordinal, interval, or ratio can be further subdivided into other categories. The multi-class issue decaying technique is being used to address the difficulty in separating the training dataset. This technique, which is based on tuple or feature samples and space, divides the primary problem into multiple subproblems. Data A crucial element in the data mining process, pre-processing increases data quality while also allowing it to be prepared and changed from the original dataset [6].

In this study, ZReddy et al. investigated two prominent dimensionality reduction approaches to determine how they affected the accuracy of various machine learning computations. The linear discrimination analysis techniques LDA and PCA are two examples of linear discrimination analysis (LDA). Decision Tree, Navie Bayes, Random Forest, and Support Vector Machine are the machine learning techniques that were employed in these investigations. All of these algorithms were trained using the Cardiotocography dataset from the University of California, Irvine, which was made publicly available in 2011. (UCI). The data reveal that PCA classifiers appear to be more prominent in the display than LDA classifiers, which is consistent with previous research. [7] The classifiers based on trees and random forests both outperformed the other two approaches.

To improve accuracy, Zhang and Zhao et al. proposed a new strategy for selecting a subset of offitting capacities in an effort to improve precision. PCA has been shown to be a successful pre-preparation stage in machine preparation for increasing handling force and consistency while reducing costs. This technique assists clinicians in making clinical decisions by allowing them to analyseCTGreadings more quickly and successfully [8].

AbdiHervé and colleagues published a study in which they discussed the PCA in great detail. When performing statistical analysis, a PCA makes use of orthogonal transformations as part of the process. Using PCA, it is feasible to convert a set of correlated variables into an array of uncorrelated variables. The exploratory data can be analysed using principal component analysis (PCA). It is feasible to use PCA to investigate the associations between a group of variables in a given situation. It is possible to lower the dimensionality in this manner [9].

IremErsoez Kaya and colleagues researched and analysed the clustering performance of five alternative PCA-based approaches using images of brain tumours as well as the images of the tumours. In this model, MRI images of varied sizes and clusters are first

subjected to the PCA technique, which employs K-means and Fuzzy C-Means to classify the images (FCM). By combining PCA and K-means, it is possible to reach higher output levels [10].

Qiang Liu and colleagues undertook a thorough analysis of the present state of machine learning safety from two perspectives: the training and testing processes. Their findings were published in the journal Machine Learning Safety. Five major research issues were posed by authors in the field of study. These polls will be extremely beneficial in the areas of machine learning and security. Although five major trends' detailed investigations on machine learning's defence approaches and security threats are mentioned in the study, they are not included in the report [11].

Shuai Zhao and colleagues have suggested a new framework for real-time network traffic anomalies identification using a machine learning method. The system's architecture is capable of supporting both batch and real-time processing. Massive data processing frameworks such as Apache Storm, Apache Kafka, and Apache Hadoop are examples of the kind of frameworks that are used in conjunction with Machine Learning methodologies and tools in the curriculum. The use of supervised learning methodologies allows decision trees, support vector machines, and naive bayesians to attain great precision and efficiency while maintaining low cost (NB). Anomaly detection in networks employing machine learning and a prediction model with a 90.3 percent success rate is achieved on average. Future research priorities include in-depth analysis of anomaly detection for a real-time network management platform; improving machine learning and optimising algorithms for real-time computation; and introducing visualisation tools to provide an inclusive understanding of complex networks' dynamic behaviours [12].

Researchers Tamer F. Ghanem and colleagues suggested a hybrid approach for detecting abnormalities in huge data sets that included evolutionary algorithms and the meta-heuristic multi-start procedure. The "NSLKDD dataset" was utilised to evaluate the effectiveness of this strategy. The accuracy rate of this method is 96.1 percent when compared to the accuracy rates of other Machine Learning algorithms. A comparison was made between the performance of their approach and that of six other algorithms, which included Bayes Network, Bayesian Logistic Regression, Naive Bayes, Radial Basis Function Network, Multilayer Feedback Neural Network, and Decision Trees, among others (J48). Machine learning algorithms were tested using Weka 3.6, and the researchers discovered that their methodology was 96.1 percent accurate, which was far better than the accuracy of other approaches [13].

When data is processed or moved, it becomes more vulnerable to attack than it was before. According to ShikhaAgrawal et al., they investigated various data mining anomaly detection methodologies. However, while there are several apps and methods for safeguarding data, there are also numerous security flaws and vulnerabilities. As a result, there has been a higher emphasis placed on data analysis, as well as the identification of various data mining assault strategies. Anomaly detection makes use of these data mining techniques to find hidden unusual activity in the data that enhances the likelihood of an intrusion or assault taking place. In an effort to better detect known and unknown attacks, a number of hybrid approaches were also used in conjunction with one another. The authors

investigated anomaly detection strategies in data mining in order to acquire a better grasp of current methods that may be employed in the future [14].

Muhammad A. et al. developed the FMIFS technique, which stands for Flexible Mutual Information Feature Selection, in order to eliminate incompatible and redundant data sets from a dataset. The hybrid feature selection (HFSA) technique was applied to both the indirect and direct datasets, and it was used to handle distinct types of information. It has been a long-standing difficulty in network traffic classification to deal with data properties that are both irrelevant and redundant. It is constructed using the approach for picking characteristics discussed above, as well as the LSSVM-IDS (Least Square Support Vector Machine dependant IDS). To conduct intrusion detection tests, three datasets were used: Kyoto 2006 +, NSL-KDD, and KDD Cup 99.

To conduct intrusion detection tests, three datasets were used: Kyoto 2006 +, NSL-KDD, and KDD Cup 99. In compared to prior approaches [15], we were able to minimise computing costs while simultaneously enhancing accuracy as a result of the enhancements made to our selection algorithm.

Andres Robles-Durazno and his colleagues presented controlled energy monitoring as a way to discover anomalies in a clean water distribution network using machine learning. The data from the testbed is utilised to train machine learning algorithms, such as Random Forest, KNN, and SVM, that are used to execute classification tasks on the testbed itself. The F-measure and accuracy are used to compare the performance of algorithms. The results demonstrate that Random Forest beats SVM and KNN by 5 percent when dealing with small data sets and by 4 percent when dealing with large data sets when dealing with small data sets, respectively. When it comes to efficiency, KNN is the best option because it requires the least amount of time to construct. The disparity between RF and SVM, on the other hand, is more pronounced. Researchers hope to widen their research in the future by using a more realistic model of client demand as a foundation. It was also an objective of the study's authors, who hoped to undertake a variety of attacks with the help of machine learning [16].
Several approaches, including fuzzy clustering, an Artificial Bee Colony (ABC), and a Multi-Layer Perceptron (MLP) network, were employed by Hajimirzaei and Navimipour et al. to develop new IDS. For the preparation of MLPs, the ABC method was utilised, and abnormal and normal traffic packets were distinguished by minimising biases and weight linkage values for the ABC algorithm, respectively. This technique has been proved to function with both the NSL-KDD data set and the Cloud Sim simulation tool [17], demonstrating that it is effective.

A random forest classifier was used to develop an intrusion detection model for intrusion detection by Nabila Farnaaz and her colleagues, who collected data using the NSL-KDD protocol. While traditional classifiers perform better when it comes to classifying attacks, Random Forest (RF) performs better when it comes to classifying attacks. We conducted NSL-KDD data set experiments to determine the effectiveness of our model. Scientists have demonstrated that the proposed FAR/DR paradigm with low false alarm rates is effective in tests they have carried out. To classify their data, the authors employed a total of ten cross validations. A comparison was made between random forest modelling and the j48 classifier [18] in terms of accuracy, precision, DR, FAR, and MCC.

A machine learning technique was used by Kamaldeep Singh et al. to identify Peer-to-Peer Botnet attacks, resulting in a scalable implementation of an intrusion detection system. The researchers used open source tools such as Mahout, Hive, and Hadoop to accomplish this. It was possible to detect Peer-to-Peer Botnets because of the parallel processing capacity of Mahout, which was used by the researchers. To train the categorization modules, captured files from bot assaults such as Conficker, Storm, and Waledac were utilised in conjunction with other captured files. These data sets were stored in PCAP (libpcap) files, which were created by the author. The 84,030 instances of mixed traffic were utilised as a data set, with 10% of the instances being used for testing and 90% being used for trainingPerformance testing of systems is available using a variety of network bandwidths ranging from 20Mbps to 300Mbps [19], depending on the network configuration.

SanjaiVeetil and colleagues set out to build an intrusion detection system that was trustworthy, fault tolerant, and scalable, as well as distributed, using a Naive Bayes algorithm on Hadoop as the foundation. For real-time intrusion detection, they make use of the HStreaming and Apache Hadoop APIs, as well as classifiers. For the aim of training, homogenous and heterogeneous clusters were both considered. When developing their models, they used 10% of the KDD intrusion detection data set to train their algorithms on. The homogeneous cluster, on the other hand, was 67 percent faster [20] when compared to the more sophisticated clusters and the Naive Bayes method alone.

It was proposed by Ren et al. to combine K-means clustering with 1V1-SVM classifiers (one to one support vector machines) to improve the accuracy of classification. This IDS model is used for a multi-classification system built on Hadoop that leverages fusion of the classifications from each of the classes. It was discovered by the classification centre that, after employing a mapper to construct a key-value pair from KDD CUP99 data sets, the fused classifier outperformed a basic classifier in tests using KDD CUP99 data sets. It was decided to employ only 50 thousand records for training purposes, whereas 5 million records would be used for testing purposes. Approximately 65 percent of the normal data and 19 percent of the intrusion data are present in the testing data, with the remainder of the data being unknown [21].

III. METHODOLOGY

This section presents a machine learning techniques for design of intrusion detection system (Figure 1) for big data networks. NSL KDD data set is used as input in this framework. First of all, CFS- correlation feature selection method is applied on the NSL KDD data set to select only relevant features. NSL KDD data set has 41 features. On applying CFS algorithm, numbers of features were reduced to 16. Then machine learning algorithms are applied on these 16 features to classify and predict malware data in NSL KDD data set.

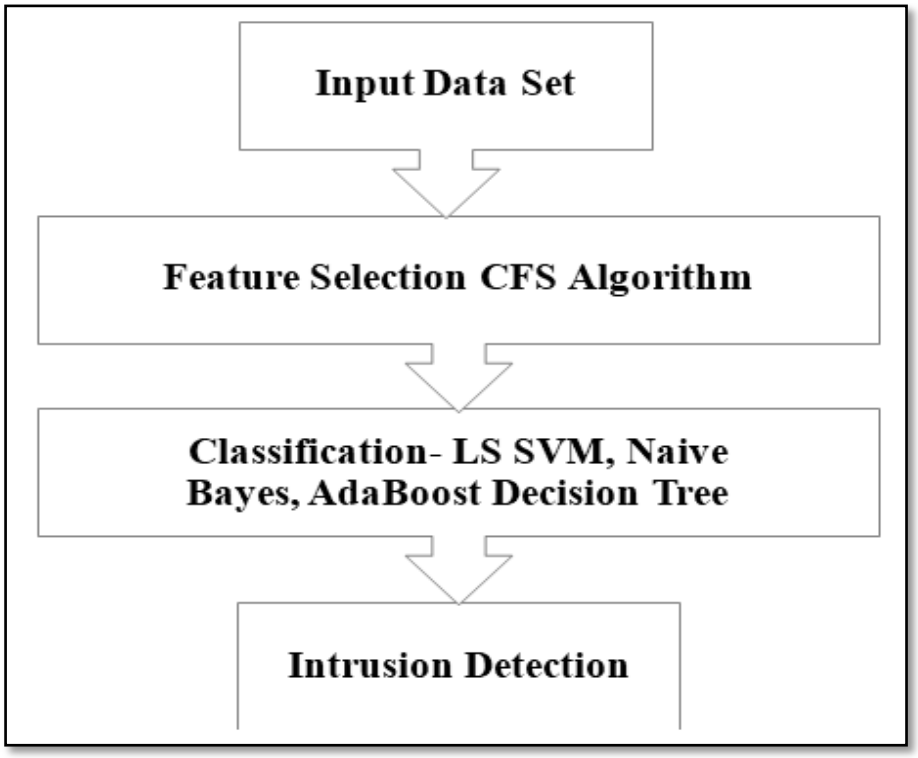

Figure 1: Machine learning techniques for design of intrusion detection system for big data networks

CFS is one of the key features selection evaluations, which evaluates the worth of a subset of attributes by taking into account each feature's particular ability as well as the degree of redundancy between them. Subsets of characteristics with high correlation with the class but low intercorrelation with the other attributes are selected [22].

This is for problems involving categorization and regression. SVM categorises data by identifying a hyperplane (line) that divides learning data into classes. The discovery of the hyperplane, which optimises the distance in the middle of classes, increases the likelihood of generalising hidden data. SVM provides the best classification performance, i.e. the training set's accuracy. It does not cause the data to overflow.

LS SVM (Least Square Support Vector Machine) makes no assumptions about the data. Demonstrate more efficiency in future data classification. SVMs are divided into two types: linear and non-linear. A line, i.e. hyperplane, is used to represent training data in a linear way [23].

The Bayesian classification is based on the theorem of Bayes. When applied to a large database, these Nave Bayesian Classification techniques characterise simple bases akin to the classification of end trees and selected networks. A subset of dependent attributes can be represented using Naive Bayes classification. The posterior probability P(x|c) of each class is derived using this procedure. In this study, the class with the highest probability predicts the outcome.

$$P(c|x) = \frac{P(x|c)\,P(c)}{P(x)} \qquad (1)$$

Where the $P(c|x)$posterior probability of each class given diabetes x attribute
$P(x|c)$ Is the likelihood value
$P(c)$ Prior Probability of diabetes class
P(x) is the prior probability of predictor
Each attribute conditionally forgives the subset class
A similar strategy is used by Naive Bayes to predict multiple sorts of probabilities based on various attributes [24].

A separate basis classification has a weighted dataset in Adaboost Decision tree [25] if the weight of a single instance in the dataset depends on the previous base classifier outcomes for each of these instances. If they misclassify an instance, the weight of that instance will be increased in future models; if the classification is correct, the weight will remain constant.

The ultimate conclusion is reached through weighted voting of the fundamental classification, which is decided by the model's weight, which is determined by the misjudgment rate.

$$w_n(x) = sign\left(\sum_{i=1}^{n} \alpha_n \, w_n(x)\right) \qquad (2)$$

If the model has higher classification accuracy, it gets low weight. If it has poor classification accuracy, it gets the highest weight. $w_n(x)$ Refers to output classifier

IV. RESULTS AND DISCUSSION

NSL-KDD is a new and improved version of the cup dataset (KDD99) [26]. Among the most essential aspects of the (NSL-KDD) dataset is that it does not contain duplicate entries in the testing data and redundant instances in training data. Because of this, the classifier is more precise. A publicly accessible version of the (NSL-KDD) is available for use by researchers. It has a total of 41 components. There are 25192 records of testing data in the NSLKDD dataset, and the rest of the data (80%) is training data (100781 records). First CFS is applied for feature selection. Then 16 features are selected. These 16 features are as follows:

Table 1: Features selected using CFS Algorithm

CFS – 16 Features
protocol_type
Service
Flag
src_bytes
dst_bytes
Land
wrong_fragment
lroot_shell
Count
srv_count
same_srv_rate

dst_host_srv_count
dst_host_same_src_port_rate
dst_host_srv_diff_host_rate
dst_host_rerror_rate
dst_host_srv_rerror_rate

For performance comparison, three parameters, accuracy, sensitivity and specificity are used.
Accuracy= (TP + TN) / (TP + TN + FP + FN)
Sensitivity = TP/ (TP + FN)
Specificity = TN/ (TN + FP)
Where
TP= True Positive
TN= True Negative
FP= False Positive
FN= False Negative
Results of different machine learning predictors are shown below in figure 2, figure 3 and figure 4. Accuracy of LS SVM is better than naïve bayes and adaboost algorithm

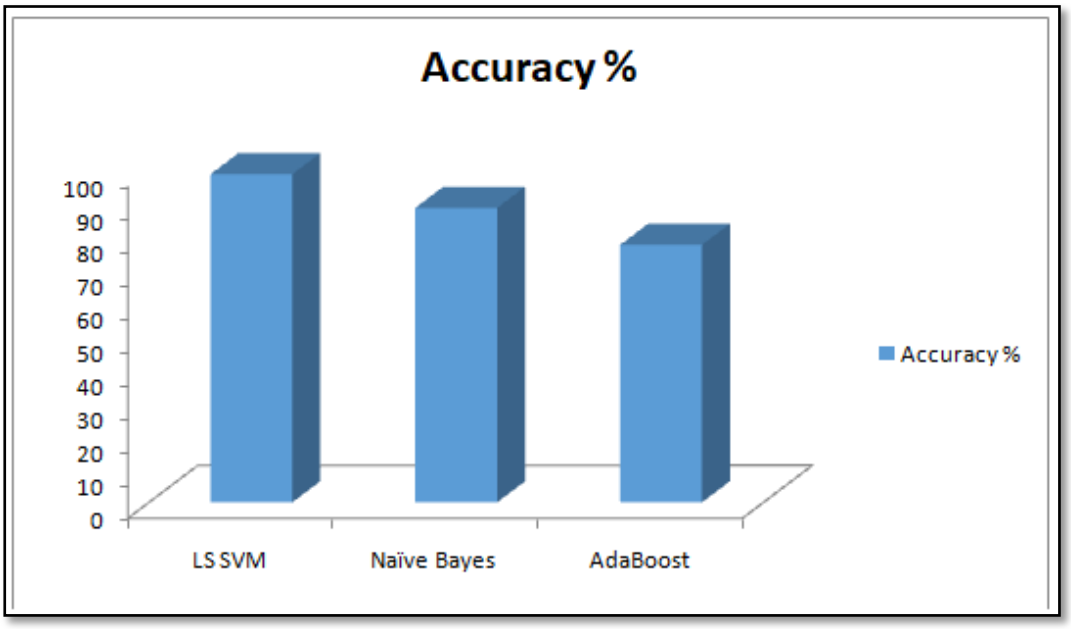

Figure 2: Accuracy of Machine Learning Techniques for Classification and Prediction of NSL KDD Data Set

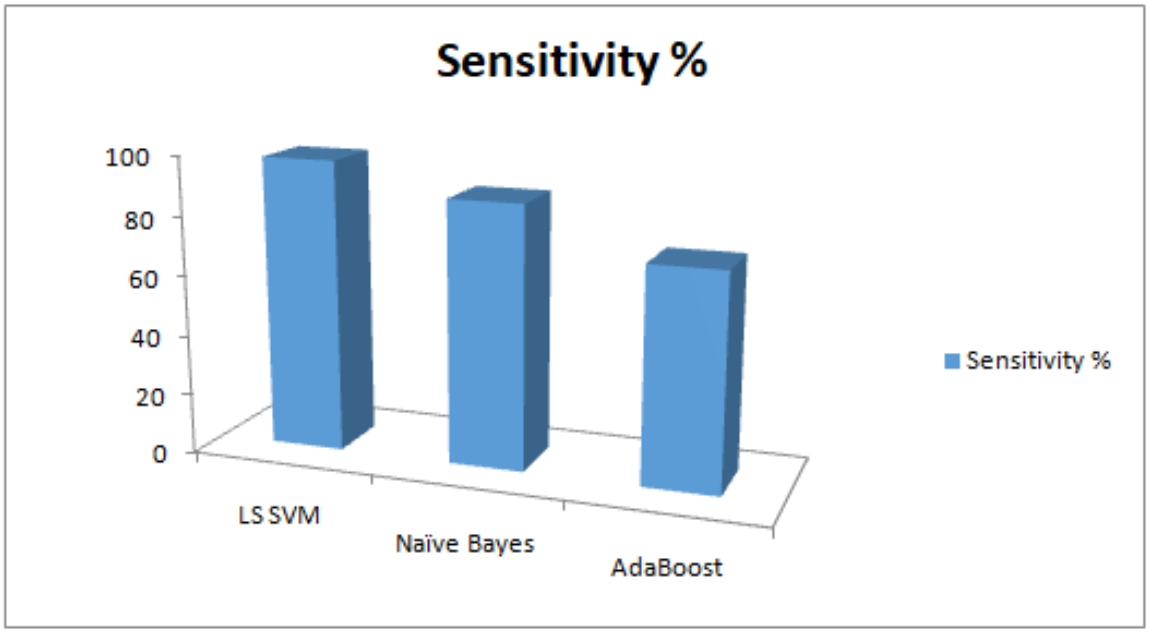

Figure 3: Sensitivity of Machine Learning Techniques for Classification and Prediction of NSL KDD Data Set

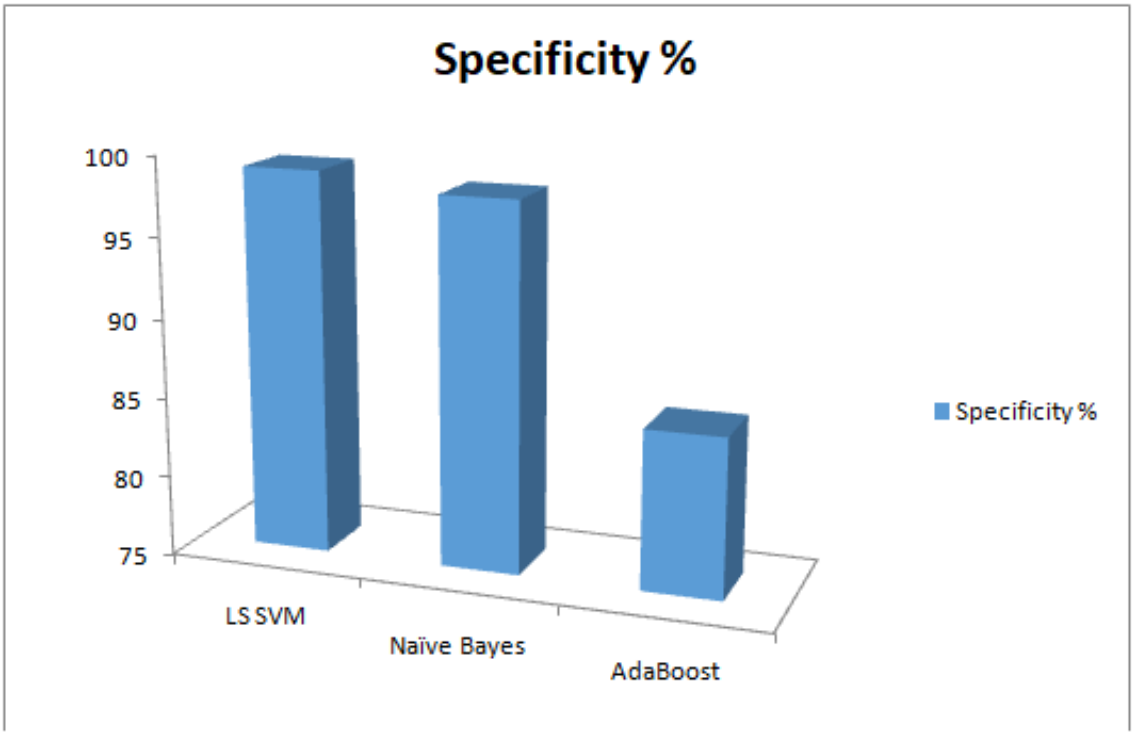

Figure 4: Specificity of Machine Learning Techniques for Classification and Prediction of NSL KDD Data Set

V. CONCLUSION

Intrusion detection is becoming increasingly crucial as network data continues to grow at an exponential rate, and manual analysis would either be impossible or take the same amount of time as manually analyzing the data would take. Because of this, when it comes to network intrusion detection, there is an urgent need for an automated system that is capable of extracting important information from large amounts of data that has hitherto been untouched. Data mining can be used to perform a wide range of tasks, including clustering, prediction, classification, and the extraction of association rules between data pieces, among other things. The purpose of this study is to examine machine learning techniques for constructing intrusion detection systems for massively parallel networks (Big Data). Specifically, the NSL KDD data set is used as input in this approach. First, the CFS-correlation feature selection approach is used to select only the most relevant features from the NSL KDD data set, which is then utilized to refine the selection. The NSL KDD data collection consists of 41 distinct characteristics. Following the use of the CFS algorithm, the number of attributes was decreased to 16. Machine learning techniques are then utilized to categorize and predict malware data in the NSL KDD data set based on the 16 attributes that were identified. Accuracy of LS SVM is better than naïve bayes and adaboost algorithm

REFERENCES

[1] M. Tang, M. Alazab, and Y. Luo, "Big data for cybersecurity: vulnerability disclosure trends and dependencies," Institute of Electrical and Electronics Engineers Transactions on Big Data, vol. 5, no. 3, pp. 317–329, 2019.

[2] Raghuvanshi, A., Singh, U., Sajja, G., Pallathadka, H., Asenso, E., & Kamal, M. et al. (2022). Intrusion Detection Using Machine Learning for Risk Mitigation in IoT-Enabled Smart Irrigation in Smart Farming. Journal Of Food Quality, 2022, 1-8. doi: 10.1155/2022/3955514

[3] D. Vasan, M. Alazab, S. Venkatraman, J. Akram, and Z. Qin, "MTHAEL: cross-architecture IoT malware detection based on neural network advanced ensemble learning," Institute of Electrical and Electronics Engineers Transactions on Computers, vol. 69, no. 11, pp. 1654–1667, 2020.

[4] OmarAlmoman., 2020. "A Feature Selection Model for Network Intrusion Detection System Based on PSO, GWO, FFA and GA Algorithms", Symmetry 2, 1046; doi:10.3390/sym12061046.

[5] Raghuvanshi, A., Singh, U., & Joshi, C. (2022). A Review of Various Security and Privacy Innovations for IoT Applications in Healthcare.Advanced Healthcare Systems, 43-58.doi: 10.1002/9781119769293.ch4

[6] J. J. Davis and A. J. Clark, "Data preprocessing for anomaly based network intrusion detection: A review," *Comput. Secur.*, vol. 30, no. 6–7, pp. 353–375, 2011.

[7] [7] G. T. Reddy *et al.*, "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020.

[8] Y. Zhang and Z. Zhao, "Fetal state assessment based on cardiotocography parameters using PCA and AdaBoost," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2017, pp. 1–6.

[9] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010.

[10] I. E. Kaya, A. Ç. Pehlivanli, E. G. Sekizkardecs, and T. Ibrikci, "PCA based clustering for brain tumor segmentation of T1w MRI images," *Comput. Methods Programs Biomed.*, vol. 140, pp. 19–28, 2017.

[11] Q. Liu, P. Li, W. Zhao, W. Cai, S. Yu, and V. C. M. Leung, "A survey on security threats and defensive techniques of machine learning: A data driven view," *IEEE access*, vol. 6, pp. 12103–12117, 2018.

[12] S. Zhao, M. Chandrashekar, Y. Lee, and D. Medhi, "Real-time network anomaly detection system using machine learning," in *2015 11th International Conference on the Design of Reliable Communication Networks (DRCN)*, 2015, pp. 267–270.

[13] T. F. Ghanem, W. S. Elkilani, and H. M. Abdul-Kader, "A hybrid approach for efficientanomaly detection using metaheuristic methods," *J. Adv. Res.*, vol. 6, no. 4, pp. 609–619,2015.

[14] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *ProcediaComput.Sci.*, vol. 60, pp. 708–713, 2015.
[15] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE Trans. Comput.*, vol. 65, no. 10, pp. 2986–2998, 2016.
[16] A. Robles-Durazno, N. Moradpoor, J. McWhinnie, and G. Russell, "A supervised energy monitoring-based machine learning approach for anomaly detection in a clean water supply system," in *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 2018, pp. 1–8.
[17] B. Hajimirzaei and N. J. Navimipour, "Intrusion detection for cloud computing using neural networks and artificial bee colony optimization algorithm," *ICT Express*, vol. 5, no. 1, pp. 56–59, 2019.
[18] N. Farnaaz and M. A. Jabbar, "Random forest modeling for network intrusion detection system," *ProcediaComput.Sci.*, vol. 89, no. 1, pp. 213–217, 2016.
[19] K. Singh, S. C. Guntuku, A. Thakur, and C. Hota, "Big data analytics framework for peerto-peer botnet detection using random forests," *Inf. Sci. (Ny).*, vol. 278, pp. 488–497, 2014.
[20] S. VeetiL and Q. Gao, "A real-time intrusion detection system by integrating hadoop and naive bayes classification," 2013.
[21] X.-Y. Ren and Y.-Z.Qi, "Hadoop-based multi-classification fusion for intrusiondetection," *JApSc*, vol. 13, no. 12, pp. 2178–2181, 2013.
[22] Aggarwal, Megha. "Performance analysis of different feature selection methods in intrusion detection." International Journal of Scientific & Technology Research 2.6 (2013): 225-231.
[23] M. E. Kabir and J. Hu, "A statistical framework for intrusion detection system," 2014 11th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2014, pp. 941-946, doi: 10.1109/FSKD.2014.6980966.
[24] Z. Shen, Y. Zhang and W. Chen, "A Bayesian Classification Intrusion Detection Method Based on the Fusion of PCA and LDA", Security and Communication Networks, vol. 2019, pp. 1-11, 2019. Available: 10.1155/2019/6346708
[25] Shahraki, A., Abbasi, M., & Haugen, Ø. (2020). Boosting algorithms for network intrusion detection: A comparative evaluation of Real AdaBoost, Gentle AdaBoost and Modest AdaBoost. Engineering ApplicationsOf Artificial Intelligence, 94, 103770. doi: 10.1016/j.engappai.2020.103770
[26] R e v a t hi, S., D. A. M a l a t hi. A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection. – International Journal of Engineering Research & Technology (IJERT), Vol. 2, December 2013, Issue 12, pp. 1848-1853

