Global Journal of Computer and Engineering Technology (GJCET)

Vol. 1, No. 2, July, 2021 ISSN: 2767-1933

EFFECTIVE CLASSIFICATION MODEL FOR INTRUSION DATASETS CONSISTING OF REAL-TIME ATTACKS

Reshamlal Pradhan, Dr. C. V. Raman University, India (reshamlalpradhan6602@gmail.com)
S.R.Tandan, Dr. C. V. Raman University, India (srtandan26@gmail.com)
Pushkar Dubey, Pt. Sundarlal Sharma(Open) University, India (drdubeypkag@gmail.com)

ABSTRACT

Today data analytics and data mining are the key attraction for researchers in machine learning. Serious network security issues and intrusions are found in individual and organizational Computers and on the web. In computer and web usage, vulnerable and suspicious events and activities are increased. With the rapid growth in computer technology and networking services, a considerable amount of data is getting generated on the web. Traditional data processing applications are not sufficient and inadequate to deal with these vast data. It consists of not only traditional threats but also many real-time threats known as attacks. NSL KDD dataset consists of 42 features and 37 attack types, CICIDS 2017 dataset consists of around 80 features and real-time attacks of Brute force attack, Botnet attack, Heartbleed, Infiltration, Web attacks. To efficiently deal with this vast amount of data, consisting of real-time attacks, there is a need to advance data mining techniques. There are challenges of performance improvement, accuracy, security, and privacy with the existing traditional data mining techniques. Machine-learning techniques and applications of prime concern are Feature reduction, Decision tree techniques, Neural Networks, Genetic algorithm, Ensemble Techniques, Statistical techniques. Here a model for the effective classification of intrusion datasets consisting of real-time attacks is presented to explore data processing, data analysis and security challenges issues. Result analysis on NSL KDD Dataset is also presented with specific decision tree and ensemble data mining techniques, which shows the performance of models with CFS Feature selection techniques. It is observed that models are maintaining good results with fewer features of the dataset. With 18 features, the accuracy of Random forest is 88.1%, the accuracy of REPTree is 84.8%, the accuracy of Bagging is 87.09%, and the accuracy of boosting is 87.09%, which are comparatively good with the accuracy of models with all 42 feature. A literature study of related work in the field is also presented to depict sever machine learning techniques applied on datasets of KDD cup99, NSL KDD, CICIDS 2017.

Keywords: Datasets, Data mining, Machine learning technique, Classification, Feature reduction, Genetic algorithm, Neural network, Ensemble Technique.

1. INTRODUCTION

Information and network security is a critical concern in the era of computer and network services. With rapid development in computer technology and networking services, individuals, scientific organizations, and business organizations need to maintain information security. Malicious activity and intrusions are increasing in the computer network and web with technological advancement. Data mining has become of prime concern for scientific and business organizations to deal with vast amounts of structured and unstructured data sets. Machine learning techniques are used in Data mining, which is the process of discovering knowledge and exciting patterns from a huge amount of data (Nejad et al., 2008).

Many research institutions and organizations deal with a massive amount of structured and unstructured data in today's technological era. Challenges for these organizations and researchers are to detect intrusions (or attacks) more accurately and efficiently. Machine learning techniques play an essential role in dealing with these types of data. Security challenges are the prime concern. Classification, Clustering, and Regression are used to identify and ensure security measures with machine learning. A variety of computer security threats necessitate intrusion detection systems and intrusion prevention systems to be highly effective.

Intrusion detection systems (IDS) monitor the behavior of a network system or Computer System to detect potential security problems. Security violations can be defined as different events or activities like integrity, confidentiality, availability, or bypassing the security wall of the network system. Data mining techniques are used in Intrusion Detection systems. Data mining is often referred to as Knowledge discovery of data (Xu et al., 2014). To extract knowledge from data following steps are performed in KDD: Data preprocessing, Data transformation, Data mining, and Pattern evaluation and presentation.

For the experimental and research purpose, different datasets are available, such as DARPA98, KDD99, NSLKDD, ISC2012, CAIDA and ADFA13, which researchers use to evaluate performance of their proposed Intrusion Detection and Intrusion Prevention models. For the practical evaluation of the performance of the Intrusion Detection model, it is desired to know the sophisticated structure of these datasets, which consist of diverse nature of attacks and feature sets. The intrusion detection system is used as a network defence system with the aim of security administration to forecast malicious events such as intrusions. Thus research on IDS domains motivated researchers to propose better IDS models over the years.

In the current computer technology era, machine learning is the centre of attraction for data mining researchers. Organizational and individuals information in the Computers and Web is going through severe issues of threats and intrusions.

The goal of the research is to help Intrusion Detection Systems better detect and mitigate vulnerabilities with the help of a variety of machine learning algorithms. The general objectives of the research are:

- This paper provides a taxonomy of work in ensemble classifiers applied to the intrusion detection system.
- Informative and promising feature selection using Feature selection techniques. As NSL KDD dataset consists of 42 sets of features, while CICIDS2017 consists of more than 80 network traffic features.
- Design a suitable framework using ensemble methods based on a decision tree, neural networks, and evolutionary algorithms like genetic algorithms for intrusion detection with enhanced performance.
- In order to deal with the real-time attacks, performance of different models should be tested with datasets that represent the network structure of different scenarios.NSL KDD dataset consists of four categories of attacks which are DOS, R2L, U2R and Probing (Sun et al., 2010), while cicids2017 consists of seven categories of attacks which are Brute force attacks, Heartbleed attacks, Botnet attacks, DoS, DDoS, Web attacks, and Infiltration attack (Ring et al., 2019). The objective of the research is to test the accuracy and performance of the models with these real-time attacks.

Further, in the second section of the paper, different machine learning techniques are explained. The third section defines machine learning techniques. In the fourth section, some of the recent work done by researchers is in tabular form. The methodology is explained in the fifth section, while the result and performance analysis is in the sixth section. Finally, the paper concluded in the seventh section by mentioning findings of the work and future scope.

2. MACHINE LEARNING TECHNIQUES

Data mining is the process of the discovery of knowledge and valuable patterns (Nejad et al., 2008). Machine learning techniques are used in the process of data mining. There are different types of machine learning techniques that perceive and generate knowledge for fraud detection. These techniques are used for classification, Regression and Clustering in order to extract information and detect intrusions. Different types are decision trees, artificial neural networks, genetic algorithms, statistical techniques, ensemble techniques etc.

Techniques, which are a rule-based tree structure to classify data, are used to define decision sets. Classification is performed through a set of decisions. The decisions in the tree structure can be viewed as representing this information. The data samples travel from the root to the leaf node through multiple branches (Shrivas et al., 2013). Types of decision tree techniques are CART, REPTree, Random Forest etc. (Tsai et al., 2009). REPTree is considered a fast decision tree learner. It uses information gain as the division criterion to build decision trees, and it uses to reduce error pruning methods to prune them. The error reduction results in a more accurate classification tree, even with vast test and training data (Belouch et al., 2017). Random forest is a scheme to construct a set of predictors, among which decision trees grow in a randomly selected data subspace. Random Forest is a type of

ensemble classification algorithm. Construct numerous decision trees, each of which employs a subset of attributes, is the notion behind creating a classifier model. Thus Random forest is a combination of tree predictors so that each tree depends on the value of an independently sampled random vector (Biau, 2012).

Artificial Neural Network is a powerful predictive and processing technique. It is a non-linear predictive model, which can be learnt trained through data. Data are processed in many layers, either in a supervised or unsupervised manner (Hota et al., 2013). Statistical models have gained popularity in the field of data mining. These can deal in a systematic approach with noisy labels and minimal information (Chen, 2010). Statistical techniques are N.B., Bayesian net, SVM etc (Hota et al., 2013). An evolutionary search technique is called a genetic algorithm. A genetic algorithm is a way of computing results for an algorithm based on the data of many individuals. These individuals are called "populations," and they are data representations for the problem input data. The optimal solution is discovered through a chromosome fitness function (Azar et al., 2014). The fitness function selects the best solution and defines selection criteria to evaluate the individuals who can take part in the process (Aziz et al., 2013).

In the Ensemble classifier, several machine learning techniques are combined to gain significantly improved performance. An ensemble model is a combination of two or more IDS techniques to detect the attack, overcome the limitations of individual models, and achieve high performance (Abawajy et al., 2014). Types of Ensemble techniques are Bagging, Boosting and Stacking. Bagging can be used with many classification methods to reduce the variance associated with predictions, thereby improving the prediction process. The idea is to take a lot of bootstrap samples from the available data, apply some prediction methods to each bootstrap sample, and then combine the results by voting for classification. Boosting is a committee-based strategy for increasing the accuracy of classification. Boosting takes a weighted average of findings from various samples after using a prediction method. Furthermore, when using boosting, the samples utilized at each step are not all selected from the same population. As a result, boosting is an iterative process that incorporates weight.

3. FEATURE SELECTION TECHNIQUES

In the current era, Feature selection is becoming an essential part of intrusion detection in order to gain high performance. Datasets in the field are composed of numerous features. Features differ in relevancy. For better performance, it is essential to select the relevant features (Hota, 2015). However, it may be unnecessary or redundant to include any of these features since the information is already included elsewhere. Computation time is affected, which results in slower IDS performance. Feature selection is used to find the best relevant features for the classification of training and testing data. (Bolon-Canedo et al., 2011).

Different feature selection techniques are:

- Correlation Feature Selection (CFS),
- Information Gain,
- Gain Ratio.

Feature selection using correlation functions is known as correlation feature selection (CFS). CFS is a basic filter technique that ranks feature subsets using a heuristic evaluation function based on correlation. The evaluation function is biased in favor of subsets with attributes that are substantially correlated with the class but uncorrelated with one another. Irrelevant features should be discarded because their correlation with the class will be less. Because they will be substantially associated with one or more of the remaining features, redundant features should be screened out (Shahbaz et al., 2016). The method of feature evaluation called I.G.(Information Gain) defines the information provided by the attribute items as the amount of information provided by the text category. The gain Ratio Feature selection method is used to rank the attributes of high dimensional datasets.

4. RELATED WORK IN THE FIELD

Research work in the field of machine learning and data mining has been in continuation since 1980. Every year new security challenges are encountered with the technological advancement in computer science and networking services. Researchers are working to identify security challenges and to design appropriate models and techniques for the solution. In today's scenario, data mining is one of the prime interests of research areas for researchers in computational science. There are numerous data mining techniques used for knowledge discovery. Mining

techniques of critical attention are ensemble techniques, Genetic algorithms, Fuzzy logic and big data analytics to explore data processing and security challenges.

Datasets in the current era consisting of numerous real-time attacks, to which traditional data processing systems are not sufficient to deal. There is a need for advancement in data processing techniques and services. Many authors trained their model with KDD cup99 dataset and NSL KDD dataset. NSL KDD is the extended version of the KDD cup dataset, which consists of 41 features. Some recent works are with CICIDS 2017 datasets composed of around 80 features and numerous real-time attacks, which are Brute force attack, Botnet attack, Heartbleed, Infiltration, and Web attacks. A literature study of work in the field is presented in Table 1.

Table 1: Comparison of results achieved by different authors using various methods

Reference	Dataset and Techniques	Observation & Outcome
Bansal et al. (2019)	NSL KDD, CICIDS 2017, SVM, Neural Network, Ensemble classifier	DDR scheme provides better results with accuracy and computational time.
Alrowaily et al. (2019)	CIDS 2017, Adaboost, MLP,DT,NB,KNNQDA,RF	Techniques result in high accuracy, precision and recall.
Tao et al. (2018)	Genetic algorithm (G.A.) and support vector machine (SVM)	Increased intrusion detection rate with accuracy and true positive rate; reduces the SVM training time and decreases the false positive rate.
Nema et al. (2016)	NSL KDD, SVM, GA	Accuracy of 98.30% achieved.
Ever et al. (2019)	NSL KDD, BPNN, SVM, Decision Tree	It delivers excellent results when utilised with training ratios that are considered, while preventing a significant decrease in accuracy.
Katkar et al. (2013)	KDD cup99, Naive Bayesian, J48, Ensemble classifier	Accuracy of Naive Bayesian classifier is improved with Feature Selection. provides maximum accuracy of 99.89785%
Abawajy et al. (2014)	Large iterative multitier ensemble classifiers(LIME)	The outcome with AUC 0.998
Hedar et al. (2015)	AGAAR algorithm	Reducing the dimensionality of the dataset results in improved classifier accuracy. With this reduction, both memory and CPU time are improved. The classification increased from 75.98% to 81.44%.
Roy et al. (2014)	KDDcup99, Ensemble-based Stacking Approach.	The accuracy rate is 82.7206%.
Bolón-Canedo et al. (2011)	KDD Cup 99, Combination of discretization, filtering and Naive Bayes, C 4.55,	While other classifiers are computationally slower and less efficient, machine learning algorithms have the benefit of being faster and more economical in use of resources.
Abadeh et al. (2011)	Genetic fuzzy system	In the Michigan approach, GFS, genetic operators (crossover and mutation) are used to guarantee valid individuals and results in a good performance.

The works depicted above clearly describes that different author's use different standalone machine learning techniques and ensemble Techniques. They trained their model with different datasets like KDD cup99, NSL KDD, CICIDS 2017. One of the everyday observations from all mentioned works is to improve performance in terms of computational time, accuracy, memory and CPU time, or feature reduction and efficiency.

5. METHODOLOGY

In this research, the proposed work is explained in different categories: feature selection, model designing and validation, Performance evaluation of models, and comparative performance study. In this section, Informative feature selection, model designing and validation is explored.

5.1 Informative feature selection

The selection of features selects the relevant features and discards irrelevant features and provides a subset of features that correctly define the given problem with high performance. Additional functions may lead to high calculation time and have an impact on IDS performance. Feature selection improves classification by looking for a sub-set of features which best classifies the IDS model through training and test data (Hota et al., 2018). NSL KDD dataset consists of 42 sets of features, while CICIDS2017 consists of more than 80 network traffic features. An evolutionary algorithm (gene selection based on informative feature selection) will be developed to select the most promising and best feature.

Here correlation feature selection (CFS) technique is used to select promising features in the process of intrusion detection. The feature selection process is depicted in figure 1given below.

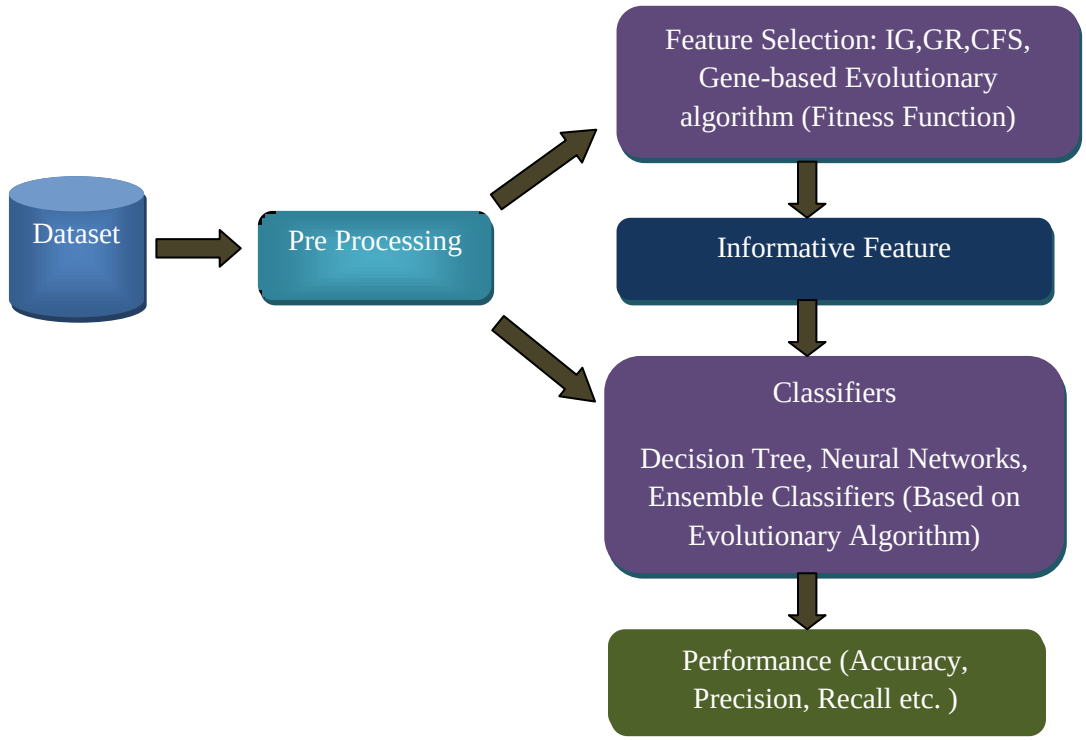

Figure 1: Feature Selection

5.2 Model Building and Validation

A Framework is designed for the research, which consists of model building and validation phase. In the model-building phase, standalone machine learning techniques like decision trees and ensemble techniques based on evolutionary techniques like genetic algorithms and neural network fuzzy logic should be used. A training set of data should be applied for the model building, while a model validation test set should be used.

A model can be developed with the training set of data. A training set of data consists of labelled data, where data definitions, whether it is average data or attack-type of data, are clearly defined. Thus these data are known training data and used in the model building phase.

Once the model builds with the training set of data, then models are validated with the test set of data. The testing set of data consists of unlabeled data. A trained model classifies this unlabeled test set of data. With the classification outcome model validated. With good results, validated models are considered for real-time implementations. This framework of model training and testing is depicted in figure 2.

In the process of model building and validation, different machine learning techniques are used. Here REPTree, Random tree, Bagging and Boosting techniques are used in the evaluation process. Performance of the model is measured in terms of accuracy, precision and recall.

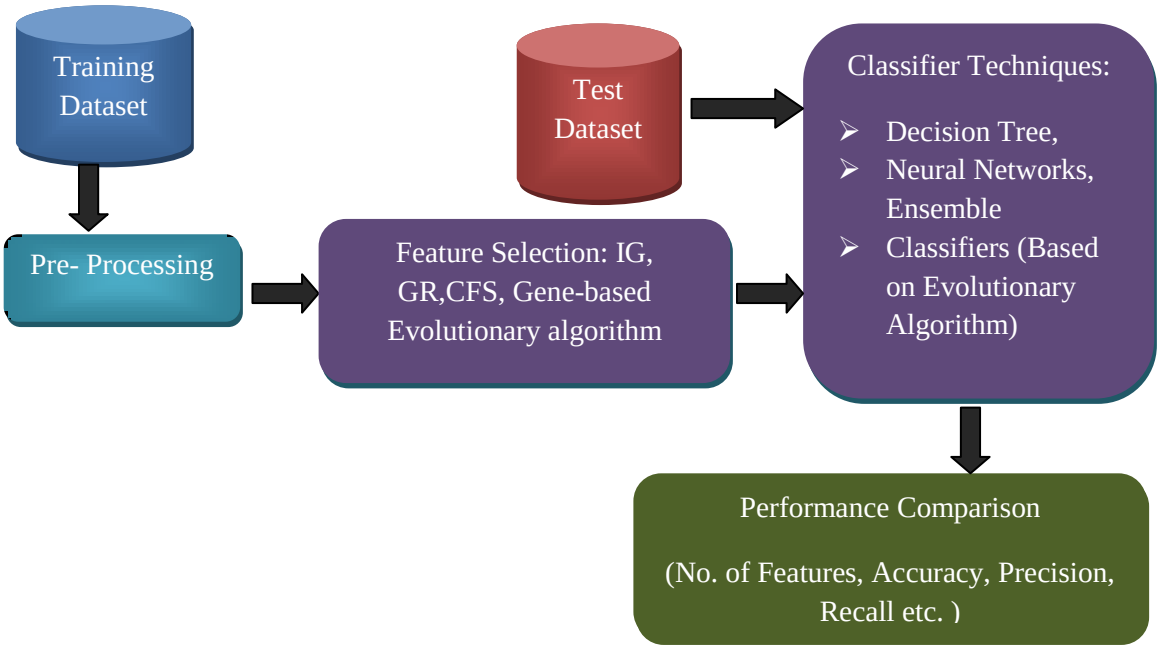

Figure 2: Data Mining model

6. PERFORMANCE AND RESULT ANALYSIS

Classifier result is evaluated using confusion matrix with accuracy, precision and recall (Mukherjee et al., 2012), where

- Accuracy is defined the proportion of the total number of correct forecasts.
- Precision is the proportion of the positive cases which had been predicted to be correct.
- The recall is the proportion of positive cases identified correctly.

This paper's result analysis on NSL KDD Dataset is presented with specific decision trees and ensemble data mining techniques. CFS Feature selection technique applied on the dataset to trace best features to classify the dataset with high accuracy. Techniques used are REPTree, Random tree, Bagging and Boosting. Initially, models are tested with all 42 features of the dataset.

All 42 features of the NSL KDD Dataset are(Shahbaz et al., 2016) are depicted in the table 2 depicted below.

Table 2: All 42 feature of NSL KDD Dataset

S. No.	Attribute Name	S. No.	Attribute Name
1	Protocol_type,	22	num_outbound_cmds,
2	Service,	23	is_guest_login,
3	Duration,	24	count,
4	flag,	25	serror_rate,
5	src_bytes,	26	rerror_rate,
6	dst_bytes,	27	srv_serror_rate,
7	wrong_fragment,	28	same_srv_rate,
8	urgent,	29	srv_rerror_rate,
9	hot,	30	diff_srv_rate,
10	num_compromised,	31	srv_diff_host_rate,
11	land,	32	dst_host_count,
12	num_failed_logins,	33	dst_srv_host_count,
13	logged_in,	34	dst_host_same_srv_rate,
14	root_shell,	35	dst_host_same_src_port_rate,
15	num_shells,	36	dst_host_diff_srv_rate,
16	su_attempted,	37	dst_host_srv_serror_rate,
17	num_root,	38	dst_host_serror_rate,
18	is_host_login,	39	dst_host_srv_diff_host_rate,
19	num_file_creations,	40	dst_host_rerror_rate,
20	num_access_files,	41	dst_host_srv_rerror_rate,
21	srv_count,	42	class.

Through the Correlation Feature selection technique best 18 features are selected for the classification of the dataset. Selected features are depicted in table 3.

Table 3: Selected Attributes (18) by CFS Feature Selection

S. No.	Attribute Name	S. No.	Attribute Name
1	protocol_type	10	serror_rate
2	service	11	same_srv_rate
3	flag	12	diff_srv_rate
4	src_bytes	13	dst_host_diff_srv_rate
5	dst_bytes	14	dst_host_same_src_port_rate
6	wrong_fragment	15	dst_host_srv_diff_host_rate
7	hot	16	dst_host_serror_rate
8	logged_in	17	dst_host_rerror_rate
9	count	18	class

The performance of models with all 42 features is depicted in table 4 given below. A graph is also plotted for the classification performance with all 42 features in figure 3.

Table 4: Classification with all features of the NSL KDD dataset			
Techniques	**Accuracy(%)**	**Precision(%)**	**Recall(%)**
Random Forest	86.02	88.9	92.3
REPTree	88.1	87.5	94.2
Bagging	87.09	85.7	92.3
Boosting	88.17	87.7	96..2

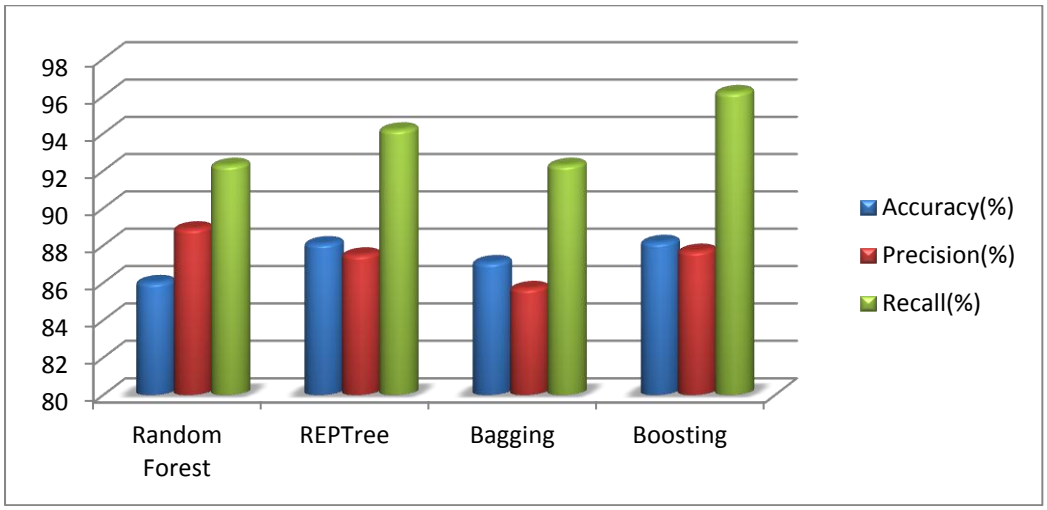

Figure 3: Classification with all 42 features

The performance of models with feature selection (18 features) is depicted in table 5 given below. A graph is also plotted for the classification with selected 18 feature set in figure 4.

Table 5: Classification with 18 features of NSL KDD dataset			
Techniques	**Accuracy(%)**	**Precision(%)**	**Recall(%)**
Random Forest	88.17	87.5	94.2
REPTree	84.9	85.2	88.5
Bagging	87.09	85.7	92.3
Boosting	87.09	90.6	92.3

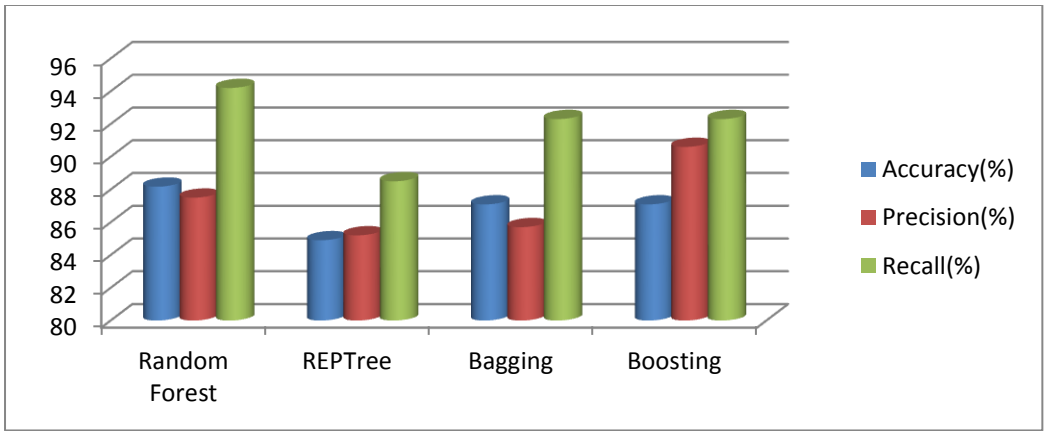

Figure 4: Classification with selected 18 features

The above results show that with feature selection, models are performing well. Decision tree techniques and ensemble techniques are maintaining good results with the dataset. Standalone techniques Random forest maintains the accuracy of 86.2%, and REPTree maintains the accuracy of 88.1 with all 42 features of the dataset, while with feature reduction maintaining the accuracy of 88.1 % and 84.9 %, respectively. With all 42 features, the accuracy of Bagging is 87.09% and boosting is 88.17%, while with feature selection (18 features), Bagging is maintaining the accuracy of 87.09% and Boosting is 87.09%. Precision and recall is also recorded for the models to depict the performance.

7. CONCLUSION

In this paper, a model for effective classification of intrusion datasets consisting of real-time attacks is presented to explore data processing, data analysis and security challenges issues. The performance of different models is tested on the NSL KDD dataset. CFS feature selection technique is applied on the models; It is observed that fewer features of the dataset also models are maintaining good results. With 18 features, random forest accuracy is 88.1%, REPTree accuracy is 84.8%, Bagging accuracy is 87.09%, and boosting accuracy is 87.09%, which are comparatively good with the accuracy of models with all 42 features. A literature study of work in the field is also presented. Result analysis of the NSL KDD dataset and Literature study depicts that standalone techniques and Ensemble techniques perform well. Still, there is a need for advancement in machine learning techniques to achieve high accuracy. NSL KDD dataset consists of 42 features, CICIDS 2017 dataset consisting of around 80 features with severe attacks, which it needs to be explored to identify real-time attacks. Other Evolutionary algorithm needs to be tested for the informative feature selection to achieve high performance, while Standalone decision tree techniques, neural network, ensemble techniques, Genetic evolutionary algorithm can be tested for model evaluation and performance analysis on different datasets.

REFERENCES

Abadeh, M.S., Mohamadi, H., & Habibi, J. (2011). Design and analysis of genetic fuzzy systems for intrusion detection in computer networks. *Expert Systems with Applications*, 38 (2011) 7067–7075, Elsevier.

Abawajy, J. H., Kelarev, A., & Chowdhury, M. (2014). Large iterative Multitier Ensemble Classifiers for security of Big data. *IEEE Transactions on Emerging Topics In Computing*, Volume 2, No. 3, September 2014.

Alrowaily, M., Alenezi, F., & Lu, Z. (2019). Effectiveness of machine learning based Intrusion detection systems. SpaCCS 2019, lncs 11611,pp. 277-288,2019, Springer.

Ambusaidi, M.A., & Nanda, P. (2016). Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm. *IEEE Transactions On Computers*, VOL. 65, NO. 10, OCTOBER 2016.

Anand, A., & Patel, B. (2012). An Overview on Intrusion Detection System and Types of Attacks It Can Detect Considering Different Protocol. *International Journal of Advanced Research in Computer Science and Software Engineering*. Volume 2, Issue 8, August 2012.

Anderson, J.P. (1980). Computer security threat monitoring and surveillance. Technical report, James P Anderson company , fort Washington,Pennsylvania.1980.

Azar, A.T., Elshazly H.I., Hassanien, A.E., & Elkorany, A.M. (2014). A random forest classifier for lymph diseases. Computer Methods and Programs in Biomedicine, 2014.

Aziz, A.S.A., Hassanien, A.E., Hanaf, S.E. & Tolba, M.F. (2013). Multi-layer hybrid machine learning techniques for anomalies detection and classification approach. 978-1-4799-2439-4/13/ ©2013 IEEE.

Bansal, A., & Kaur, S. (2019). Data dimensionality reduction scheme for intrusion detection system using standalone classifiers. *ICACDS 2019*,CCIS 10455, pp 436-4551,2019, Springer.

Bello, F.L., Ravulakollu, K., & Amrita (2015). Analysis and Evaluation of Hybrid Intrusion Detection System models. *International conference on computers, communication and systems.*

Belouch, M., Hadaj, S.E. & Idhammad, M. (2017). A Two-Stage Classifier Approach using RepTree Algorithm for Network Intrusion Detection. (*IJACSA) International Journal of Advanced Computer Science and Applications*,Vol. 8, No. 6, 2017.

Biau, G. (2012). Analysis of a Random Forests Model. *Journal of Machine Learning Research* 13 (2012) 1063-1095.

Bolón-Canedo, V., Marono, N.S., & Betanzos, A.A. (2011). Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset. *Expert Systems with Applications* 38 (2011) 5947–5957, Elsevier.

Chen, M. (2010). Bayesian Data Mining and Machine Learning. Frontiers of Statistical Decision Making and Bayesian Analysis, 2010.

Ever, Y.K., Sekeroglu, B., & Dimililer, K. (2019). Classification analysis of Intrusion Detection on NSL-KDD using machine learning algorithms. MobiWIS 2019, LNCS 11673, pp. 111-122, 2019, Springer.

Feng, W., Hu, G., & Zhang, Q. (2014). Mining network data for Intrusion detection through combining SVMs with Ant colony networks. Elsevier, *Future Generation Computer Systems* 37(2014) 127 – 140.

Gumus, F., Sakar, C.O., Erdem, Z., & Kursun, O. (2014). Online Naive Bayes Classification for Network Intrusion Detection. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* ,ASONAM 2014.

Han, J., Kamber, M., & Pei, J. (2006). Data mining concepts and tech-niques. Second edition, San Francisco, Margan Kaufmann Publishers, USA.

Hedar A.R., Omer, M.A., Al-Sadek, A.F., & Sewisy, A.A. (2015). Hybrid evolution algorithm for data classification in IDS. *IEEE SNPD 2015*, June 1-3 2015, Takamatsu, Japan.

Hota, H.S., Shrivas, A.K., & Snghai, S.K. (2013). Artificial Neural Network, Decision Tree and Statistical Techniques Applied for Designing and Developing E-mail Classifier. *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-1, Issue-6, January 2013.

Hota, H.S., Sharma, D.K., & Shrivas, A.K. (2018). Development of an efficient classifier using proposed sensitivity-based feature selection technique for intrusion detection system. *Int. J. Information and Computer Security*, Vol. 10, No. 1, 2018.

Hota, H.S. (2015). Phishing attack detection using decision tree with proposed feature selection technique. *Journal of Global Information Technology*, Vol. 10, No. 1 & 2, 2015, pp. 28-33.

Katkar, V.D., & Kulkarni, S.V. (2013). Experiments on Detection of Denial of Service Attacks using Naïve Bayesian Classifier. *International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*, IEEE(2013).

Li, L., & Yang, D. (2010). A Novel Rule Based Intrusion detection system Using Data Mining. Proc. Of 3rd IEEE *International conference on computer science and information technology*, pp. 169-172,2010.

Mukherjee, S., & Sharma, N. (2012). Intrusion Detection using Naive Bayes Classifier with Feature Reduction. Procedia Technology, 2012.

Nejad, A.F., Kharazmi, S., & Bayati, S. (2008). Improving Admission Control Policies in Database Management Systems, Using Data Mining Techniques. *International Conference on Computer Science and Software Engineering (ICCSSE)* 2008**.**

Nema, A., Tiwari, B., & Tiwari, V. (2016). Improving accuracy for Intrusion detection through layered approach using Support Vector Machine with feature reduction. *WIR16,* March 2016, Indore, India, ACM, ISBN 9781-4503-4278-0/16/03.

Pan, S., Morris, T., & Adhikari, U. (2015). Developing a Hybrid Intrusion Detection System Using Data Mining for Power System. IEEE *Transactions On Smart Grid*, VOL. 6, NO. 6, NOVEMBER 2015.

Panda, M., Abraham, A., & Patra, M. R., (2011). A hybrid intelligent approach for network intrusion detection. Proceeding Engineering, Volume 30, 2012, Pages 1-9.

Patil, A. & Mali, S.S. (2016). Hybrid Cryptography Mechanism for Securing Self-Organized Wireless Networks. *3rd International Conference on Advanced Computing and Communication Systems (ICACCS -2016)*, Jan. 22 – 23, 2016, Coimbatore, INDIA.

Pujari, A.K. (2001). Data mining techniques. 4th edition, Universities Press (India) Private Limited.

Rajpal, R., & Kaur, S. (2018). An Efficient Hybrid Approach Using Misuse Detection and Genetic Algorithm For Network Intrusion Detection. Springer Nature Singapore Pte Ltd. 2018, M. Singh et.al (Eds): ICACDS 2018,

Revathi, M., & Ramesh, T. (2011). Network intrusion detection system using reduced dimensionality. *Indian Journal of Computer Science and Engineering (IJCSE)*, ISSN: 0976-5166, Vol. 2 No. 1 pp. 61 -67.96.

Revathi, S., & Malathi, A. (2013). A detailed analysis on NSL-KDD Dataset using various machine learning techniques for intrusion detection. *International Journal of Engineering &Technology (IJERT)*,ISSN: 2278-0181,vol. 2 issue 12. December-2013.

Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. Computers & Security, 2019.

Roy, S.S., Krishna, P V., & Yenduri S. (2014). Analyzing Intrusion Detection System: An Ensemble based Stacking Approach. 978-1-4799-1812-6/14/$31.00 ©2014 IEEE.

Shahbaz, M.B., Wang, X., Behnad, A., & Samarabandu, J. (2016). On efficiency enhancement of the correlation-based feature selection for intrusion detection systems. IEEE 7th *Annual Information Technology, Electronics and Mobile Communication Conference* (ICON), 2016.

Sharafaldin, I., Lashkari, A.H., & Ghorbani, A.A. (2018). Towards generating a new intrusion dataset and intrusion traffic characterization. *International conference on information system security and privacy,ICISSP* 2018.

Shitharth, S., & Winston, D.P. (2016). A Novel IDS Technique to Detect DDoS and Sniffers in Smart Grid. *World conference on futuristic trends in research and innovation for social welfare*.

Shrivas, A.K., Singhai, S.K., & Hota, H.S. (2013). An Efficient Decision Tree Model for Classification of Attacks with Feature Selection. *International Journal of Computer Applications* (0975 – 8887) Volume 84 – No 14, December 2013.

Sornsuwit, P., & Jaiyen, S. (2015). Intrusion Detection Model Based on Ensemble Learning for U2R and R2L Attacks. 7th*International Conference on Information Technology and Electrical Engineering (ICEE)*, Chiangmai, Thailand.

Sun, S., & Wang, Y. (2010). Research and Application of an improved Support Vector Clustering Algorithm on Anomaly Detection. *Journal of Software*, 2010.

Tao, P., Sun, Z., & Suna, Z. (2018). An Improved Intrusion Detection Algorithm Based on GA and SVM. *Special section on human-centered smart systems and technologies*, IEEE ACCESS, volume 6, 2018.

Tsai, C., Hsu, Y., Lin, C., & Lin, W. (2009). Intrusion detection by machine learning: A review. *Expert Systems with Applications*, Elsevier, 36 (2009) 11994–12000, 0957-4174/ 2009.

Xu, L., Wang, J., Yuan, J., & Jiang, C. (2014). Information security in Big data: Privacy and Data Mining. IEEE access, *The journal for rapid open access publishing*, Volume 2, 2014.

Yuan, Y., Huo, L., & Hogrefe, D. (2017). Two Layers Multi-Class Detection Method For Network Intrusion Detecton System. IEEE Symposium on Computers and Communications (ISCC), 978-1-5386-1629-1/17/$31.00©2017 IEEE.