Machine Learning Techniques for Intrusion Detection: A Comparative Analysis

Yasir Hamid, M Sugumaran, Ludovic Journaux

Machine Learning Techniques for Intrusion Detection: A Comparative Analysis

Yasir Hamid
Dept. of Computer Science and Engineering,
Pondicherry Engineering College,
Pondicherry, India
+91 9486971957
bhatyasirhamid@pec.edu

M. Sugumaran
Dept. of Computer Science and Engineering,
Pondicherry Engineering College
Pondicherry, India
+919488829865
sugu@pec.edu

Ludovic Journaux
Lab Le2i,
University of Burgundy,
Dijon,
France
+33 685109784
ljourn@gmail.com

ABSTRACT
With the growth of internet world has transformed into a global market with all monetary and business exercises being carried online. Being the most imperative resource of the developing scene, it is the vulnerable object and hence needs to be secured from the users with dangerous personality set. Since the Internet does not have focal surveillance component, assailants once in a while, utilizing varied and advancing hacking topologies discover a path to bypass framework's security and one such collection of assaults is Intrusion. An intrusion is a movement of breaking into the framework by compromising the security arrangements of the framework set up. The technique of looking at the system information for the conceivable intrusions is known intrusion detection. For the last two decades, automatic intrusion detection system has been an important exploration point. Till now researchers have developed Intrusion Detection Systems (IDS) with the capability of detecting attacks in several available environments; latest on the scene are Machine Learning approaches. Machine learning techniques are the set of evolving algorithms that learn with experience, have improved performance in the situations they have already encountered and also enjoy a broad range of applications in speech recognition, pattern detection, outlier analysis etc. There are a number of machine learning techniques developed for different applications and there is no universal technique that can work equally well on all datasets. In this work, we evaluate all the machine learning algorithms provided by Weka against the standard data set for intrusion detection i.e. KddCupp99. Different measurements contemplated are False Positive Rate, precision, ROC, True Positive Rate.

Keywords
False Positive, IDS, Machine Learning, Precision, ROC, True Positive

ICIA-16, August 25-26, 2016, Pondicherry, India

DOI: http://dx.doi.org/10.1145/2980258.2980378

1. INTRODUCTION
With the rapid development of business, the growth of the cyber community, hacker community and other transactions over Internet, computer security has become a critical issue [1].

No matter how much care is taken to lay down a secure system, computer security vulnerabilities surface up every day. The time between announcement of the vulnerability and reporting the first occurrence of it is diminishing. This keeps the network managers on their toes to defend and recover from any of the attacks that creep into the system. Threat prevention on the system was possible with conceivable and fairly manageable cyber community. With the development in the cyber-criminal community, the evolvement of hacking technologies and the availability of many free to use tools, prevention is simply not possible. Along these lines, 'secure' is a relative term, sufficiently given time and assets, any framework can be compromised. There are lots of threats pertaining to the system and in the networks one such threat is Intrusion.

An intrusion is an assault on the availability, integrity and confidentiality of the system [2]. Intrusions can originate from insiders such as uneducated man force, disgruntled employs or from outsiders such as hackers, crackers, cyber terrorists and hacktivists[3]. Intruders have large number of motives, greed, hatred, military, economic espionage etc. The process of evaluating the system logs to look for the footprints and uncover intervention if any is called Intrusion Detection. Over the years, intrusion detection has been reached in varied approaches like statistical, bio-inspired, fuzzy, Markov etc. Comprehensively there are two distinctive methodologies namely, anomaly and misuse for intrusion detection with the difference in what they try to model. Anomaly detection presumes that normal user behaviour is perfectly observable and adequately different from intrusive. It builds up a model for the normal user profile and the user behaviour that differs from the established one, flagged as intrusion [4]. Contrast to this misuse detection on the hand has a signature base of well-known threats and looks for the match's in the monitored data and reports an intrusion if there is a match. The added advantage of using anomaly detection over misuse detection is that novel attacks (if they are different from normal) can be flagged but with this advantage comes the problem of too many false alarms with least consideration for evolving normal behaviour. Misuse detection, on the other hand has minimal false alarms but they tend to detect only the attacks for which they have a signature in their database, but fail to recognize the ones that are not covered by their signature base. Off late to fuse both anomaly

and misuse detection in a way that they supplement each other so that the resulting system has the ability to detect the novel attacks and has lessened false alarms a considerable measure of endeavours has been put so far.

Over the years, intrusion detection has received a lot of attention and up to the moment, researchers have developed Intrusion Detection Systems (IDS) proficient of detecting attacks in several available environments. A boundlessness of methods for misuse detection as well as anomaly detection has been applied most popular of the all is using machine learning techniques. Machine learning is a vast field and has a broad range of applications including natural language processing, medical diagnosis, search engines, speech recognition, game playing and a lot more. Comprehensively machine learning strategies can be delegated as supervised or unsupervised depending on whether there is a need to train the algorithm on labelled instances. In the case of supervised learning techniques, the algorithm is trained on labelled sets and it determines a function to map instances to classes. Later in the testing stages, it has to predict the class for unforeseen instances according to the function. In the case of unsupervised techniques, there is no proper training of the learning technique on labelled data; rather they are presented with data items which they group in a set of groups with the aim of maximizing the intra-group similarity and minimizing inter-group similarity. Various machine learning techniques have been produced for different applications over the years. However, no single machine learning algorithm can be used appropriately for all learning problems rendering it impossible to create a general learner for all problems because there are varied sorts of genuine datasets that can't be taken care of by a single learner. Here in this paper, we show a relative investigation of execution of the different machine learning methods using Weka over KDDCUPP9 (benchmark data set for intrusion detection) using 10 fold cross validation and report how they perform in terms of True Positive Rate(Tp_Rate), False Positive Rate(Fp_Rate), precision, accuracy etc.

The rest of the paper is organized as follows. A brief review of the related literature is given in Section 1. Section 3 gives subtle elements on machine learning, Section 4 examines the machine learning methods provided by weka, in Section 5 the dataset utilized for the study is discussed, Section 5 describes about the performance measures taken into consideration, in Section 6 results of system on both full and diminished dataset are given and finally conclusion is given in Section 7.

2. LITERATURE REVIEW

A review of supervised machine learning techniques is given in [5]. A discussion about the Network Intrusion Detection, techniques and open issues is given in [6]. A detailed survey of the research efforts spared for intrusion detection over last few decades is given in our work [7], with the plenty of works listed in the paper the authors conclude that Hybrid Machine Learning techniques have been used vastly. Authors in [8] proposed the hybrid Audit Data Analysis and Mining, where the anomaly detection is followed by misuse detection. A model based on rule based analysis and statistical model was proposed in [9]. Authors in [10] have examined the feasibility of leaving some of the features of the dataset out, without compromising on the detection rate and had used CFS for feature selection. The detection rate of the reduced dataset was rather improved as CFS selected the best features and eliminated the redundant features. In [11] an approach of misuse detection followed anomaly detection was proposed. To process the large amount of data authors in [12] have proposed anomaly intrusion detection using improved Self Adaptive Bayesian Algorithm. A Fuzzy Rule based approach for anomaly detection was proposed in [13].In [14] authors propose a novel idea to reduce the dimensionality of the data by using triangle based k-nn approach.

3. MACHINE LEARNING

Machine learning is the study of algorithms that improve their performance with experience and are meant to computerize exercises; the machine takes every necessary step consummately furthermore in a maintained way. It is a type of artificial intelligence that provides computers with the ability to learn without being explicitly programmed[15]. It includes various learning techniques classified as supervised, unsupervised and reinforcement learning depending on the presence or the absence of labelled data. Supervised learning trains the program with labelled samples; thereby the trained program can predict similar unlabelled samples. It includes Prediction, Knowledge extraction and Compression tasks. Unsupervised learning doesn't have any training samples; it uses the statistical approach of density estimation. Unsupervised learning works by the principle of finding the hidden design of the data by clustering or grouping data of similar kind. It includes works like Pattern Recognition and Outlier Detection. Reinforcement learning is focused on software agents that need to take action in an environment so that it maximizes cumulative reward [16]. Each step of the agent is not considered individually for success or failure but on a sequence of actions taken together should have a direction towards good policy. This learning is much used in Gaming theory and Robot Navigation.

Figure 1. Classification of Machine Learning

This paper concentrates on the relative examination on the issue of Intrusion Detection in systems by applying different Prediction procedures under supervised learning. Essential to say, Intrusion detection depends on the presumption that the conduct of gatecrashers is different from a lawful user [2]. Prediction is at the important aspect of almost every scientific discipline, and the study of the forecast (prediction) from information is the focal theme of machine learning and statistics. Machine learning and statistical methods are used all over the scientific world for their use in handling the "information overload" that characterizes our current digital age [17]. This Paper centres over different prediction techniques that are utilized for examination, which include J48, Random forests, Zero R, One R, Naïve Bayes, Naïve Bayes updateable, Multilayer Perception, K star, AdaBoost, M1 and Bagging.

4. DATASET

The KDD Cup 1999 dataset, utilized for benchmarking intrusion detection issues, is used in our experiments. The dataset is a

gathering of simulated crude TCP dump data over a time of 9 weeks on a LAN. The training data was processed to about 5 million connections records from seven weeks of network traffic and two weeks of testing data yielded around 2 million connection records. A labelled set of records is given for training purpose and once the classifier is trained its effectiveness is checked on a different set of unlabelled records. Both training and testing data are taken from the different distribution with testing data containing some records which are not present in the training set [18]. The training data consists of 22 different types of attacks and 39 attacks are present in the test data [19]. This is the standard dataset for intrusion detection and for the last decade and a half and in this work, this dataset has been used for assessing the viability of different procedures.

Each record is formed of 41 attributes and 42nd being the class attribute. Data records in total consist of 22 attacks spanned across five groups, four of them being attacks i.e. DOS, R2L (Remote to Local), U2R (User to Root), and PROBE and one representing normal data. Each record of the dataset has 32 continuous, 3 categorical and 6 are nominal attributes [20]. This dataset being very large is seldom used because of computational considerations.

Most of the research efforts use only a subset of this, and hence in this work also we use 10% KDDCUP99 dataset. This dataset forms a subset of the actual KDDCUP99 dataset. Moreover, for evaluating each of the technique 10 cross-validation is used. In 10 cross validation the dataset is broken into ten subsets and in each of the iteration nine of them are used for training the classifier and 1 is used for testing. This process is repeated 10 times and the mean of results is taken for the consideration. This way the classifier is trained and tested on each of the subsets. Names and detailed description of all 41 features of the dataset are listed in the work [21]. Features can be classified into four groups i.e., Basic features consisting of all the attributes that are from TCP/IP connection, Content features used to evaluate the payload of the packet and look for suspicious behaviour, Time-based features designed to capture properties that mature over a 2-second temporal window, connection-based features computed over a historical window estimated over the number of connections [22].

It is well known that features constructed from the data content of the connections are more important when detecting R2L and U2R attack types in KDD99 intrusion dataset[23], while time-based and connection-based features are more important for detection of DoS (Denial of Service) and probing attack types [24]. Table 1 given below lists out the three variations of datasets mostly used. For each dataset, the number of different groups of attacks and number of normal connections is listed.

Table 1: Three variations of dataset mostly used

Dataset	DoS	U2R	R2L	Probe	Normal	Total
10% KDD	391458	4107	52	1126	97277	494020
Corrected KDD	229853	4166	70	16347	60593	311029
Whole KDD	3883370	41102	52	1126	972780	4898430

5. MACHINE LEARNING WITH WEKA

As discussed above machine learning techniques have diverse purposes i.e. classification, clustering, association finding. This work has considered classification techniques assessment. In many cases, more than one classifier is implemented in combination to tackle a solitary issue. With end goals, this study uses Weka. Weka is collection of the machine learning techniques for knowledge discovery, which can directly be applied to data or called from Java code. Being open source and free to use most of the researchers in data mining field and knowledge discovery have used it. In Figure2 a snapshot of weka is given.

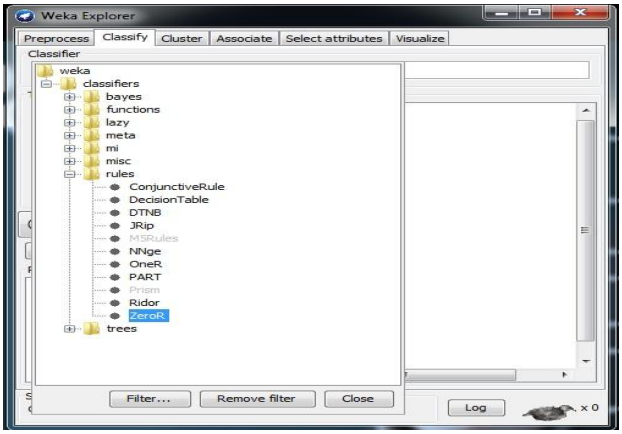

Figure 2. Weka Explorer

As evident from the snapshot that weka gathers various machine learning techniques under different tabs. Here in this work, the methods from the classify tab are used. Various classification techniques are grouped into different groups. As already mentioned that our problem is a multi-class classification problem with and the dataset available has both numerical and nominal attributes, so only those machine learning techniques are taken into consideration which is appropriate for multi class classification and feed on both nominal and numerical attributes.

In Table 2 given below the techniques employed are listed and for each technique a brief description and the class it belongs to is given.

Table 2: Machine learning techniques available in Weka

Sno	Group	Technique	Description
1.	Rule Based	Decision Table	Builds a decision table on majority class classifier.
		JRip	This class implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by William W. Cohen as an optimized version of IREP [25]
		ZeroR	Simple most classifier and is used as a baseline, ZeroR classifier simply predicts the majority category.
		OneR	OneR, short for "One Rule", is a simple, yet accurate, classification algorithm that generates one rule for each predictor in the data, and then selects the rule with the smallest total error as its "one rule".
		PART	Builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule.
2.	Bayes Rule	BayesNet	Provides data structures (network structure, conditional probability distributions, etc.) and facilities common to Bayes Network learning algorithms like K2 and B.

		NaiveBayes	The Naive Bayesian classifier is based on Bayes theorem with independence assumptions between predictors. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods [26].
3.	Functions	MultiLayerPerceptron	It is a feed foreword network that makes use of back propagation for classifying the instances, it is a directed graph with many nodes at a layer and each layer being fully connected to next layer making it appropriate for solving linearly inseparable problems[27].
		SMO	SMO automatically transforms all the nominal attributes into binary attributes, solves the multiclass problems by pair wise classification.
		Simple Logistic	Classifier for building linear logistic regression models, optimal number of LogitBoost iterations to perform is cross-validated, which leads to automatic attribute selection.
4.	Lazy Learners	IBk	Instance based with parameter k also known as K-Nearest Neighbours, or KNN, is a family of simple classification and regression algorithms based on Similarity (Distance) calculation between instances.
		Kstar	K* is one of the lazy learning approaches. These are also called memory-based methods. These methods work by learning the structure of a domain by storing examples with their classification
		LWL	Locally Weighted Learning for classification and regression uses an instance-based algorithm to assign instance weights which are then used by a specified Weighted Instances Handler.
5.	Tree	DecisionStump	A decision stump is a one-level decision tree having internal node (the root) which is immediately connected to the terminal nodes (its leaves). A decision stump makes a prediction based on the value of just a single input feature[28].
		J48	A decision tree is a predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. In WEKA C4.5 learner is implemented as J4.8.
6.	Misc	InputMappedClassifier	Wrapper classifier that addresses incompatible training and test data by building a mapping between the training data that a classifier has been built with and the incoming test instances' structure [19].

6. PERFORMANCE METRICS

All the techniques take into consideration are checked for some performance metrics derived from confusion matrix. A confusion matrix is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. All the metrics are based on the confusion matrix given in Table 3 below.

Table 3. Confusion Matrix

Actual Class	Predicted Class: YES	Predicted Class: NO
YES	True Positive(TP)	False Negative(FN)
NO	False Positive(FP)	True Negative(TN)

The metrics used are listed in Table 4 given below. For each metric a brief discussion and the calculating formulae are given.

Table 4: Performance Metrics

Sno	Technique	Description	Formula
1.	Accuracy	Proportion of classifications, over all the N examples, that were correct	$\text{Acc} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}}$
2.	Recall	Proportion of positive examples that were classified correctly	$\text{r} = \frac{\text{tp}}{\text{tp} + \text{tn}}$
3.	Precision	Proportion of correct positive classifications over all positive classifications	$\text{p} = \frac{\text{tp}}{\text{tp} + \text{fp}}$
4.	F-Measure	F-measure conveys the balance between precision and recall is actually is the harmonic mean of precision and sensitivity	$\text{FM} = \frac{2 * \text{p} * \text{r}}{\text{p} + \text{r}}$
5.	TP Rate	measures the proportion of positives that are correctly identified as positives	$\text{TPR} = \frac{\text{tp}}{\text{tp} + \text{fn}}$
6.	TN Rate	measures the proportion of negatives that are correctly identified as negatives	$\text{TNR} = \frac{\text{tn}}{\text{fp} + \text{tn}}$
7.	ROC Area	An "optimal" classifier will have ROC area values approaching 1, with 0.5 being comparable to "random guessing"	
8.	Kappa statistic	Kappa is a chance-corrected measure of agreement between the classifications and the true classes. A value greater than 0 means that classifier is doing better than chance.	
9.	Mean absolute error	The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables	

7. Results

Table 5 a given below presents the results of all the techniques applied on full dataset. It is very clear from the table that PART Classifier (Rule Based) performs best on the dataset with 99.970% correctly classified with a Mean Absolute Error of 0. Input Mapped Classifier has worst results on the dataset with 56.838% correctly classified and mean absolute error of0.0514.

Table 5: Results on full dataset

Sno	Group	Technique	CC	TP Rate	FP Rate	Precision	Recall	F-Measure	RA	KS	MAE
1.	Rule Based	DT	99.757	0.998	0.001	0.998	0.998	0.997	1	0.9959	0.0014

		CR	78.538	0.785	0.061	0.677	0.785	0.713	0.937	0.6318	0.0203
		ZeroR	56.838	0.568	0.568	0.323	0.568	0.412	0.5	0	0.0514
		OneR	98.122	0.981	0.005	0.979	0.981	0.979	0.988	0.9682	0.0016
		PART	99.970	1	0	1	1	1	1	0.9995	0
2.	Bayes Rule	BayesNet	99.670	0.997	0	0.998	0.997	0.997	1	0.9944	0.0003
		NaiveBayes	92.748	0.927	0	0.989	0.927	0.949	1	0.8802	0.0063
		NBUpdatable	92.748	0.927	0	0.989	0.927	0.949	1	0.8802	0.0063
3.	Functions	MLP	98.75	0.988	0.004	0.977	0.988	0.982	0.998	0.979	0.0011
		SMO	99.552	0.996	0.001	0.995	0.996	0.995	0.999	0.9924	0.0793
		Simple Logistic	99.941	0.999	0	0.999	0.999	0.999	1	0.999	0.0001
4.	Lazy Learners	IB1	99.941	0.999	0	0.999	0.999	0.999	1	0.9999	0.0001
		IBk	99.941	0.999	0	0.999	0.999	0.999	1	0.9999	0.0001
		Kstar	99.904	0.999	0	0.999	0.99	0.999	1	0.9984	0.0001
		LWL	98.261	0.983	0.005	0.966	0.983	0.974	0.999	0.9702	0.0058
5.	Tree	DecisionStump	78.537	0.785	0.061	0.677	0.785	0.713	0.937	0.6318	0.0203
		J48	99.960	1	0	1	1	1	1	0.9993	0
6.	Meta-Algorithm	AdaboostM1	97.859	0979	0.005	0.962	0.979	0.97	0.993	0.9636	0.0478
7.	Misc	InputMappedClassifier	56.838	0.568	0.568	0.323	0.568	0.412	0.5	0	0.0514

6.1.Feature Reduction

Feature reduction is the set of techniques that are aimed at reducing the complexity of dataset by eliminating some of the non-descriptive attributes. Works have shown that for the purpose of classification, seldom are all the features of dataset equally needed and hence we could enjoy as good results as given when all attributes are taken into consideration with very less computational resources. Table 6 given below provides the results of various feature selection techniques available in weka. Feature selection methods either select a subset of the features or rank the whole set in order of importance.

Table 6: Feature Selection

Sno	Evaluator	Search	Selected Attributes
10.	CfsSubsetEval	BestFirst	2,3,4,5,6,7,8,14,23,30,36 : 11
		GeneticSearch	2,3,4,5,6,7,8,10,12,19,23,29,30,31,33,36,38 : 17
		GreedyStepwise	2,3,4,5,6,7,8,14,23,30,36 : 11
		SubsetSizeForwardSelection	2,3,4,5,6,7,8,14,23,30,36 : 11
		LinearForwardSelection	2,3,4,5,6,7,8,14,23,30,36 ; 11
11.	GainnRatio	Ranker	8,7,13,2,11,4,10,26,25,12,3,30,39,38,36,9,5,29,14,6,35,34,33,23,22,37,24,32,40,27,41,31,28,18,1,17,16,19,15,20,21 : 41
12.	ChiSquared	Ranker	5,6,3,4,23,35,8,30,10,38,33,36,25,24,37,34,29,40,26,2,39,27,13,7,11,41,32,12,31,28,14,1,18,9,22,17,15,19,16,20,21 : 41
13.	InfoGain	Ranker	5,23,3,24,36,2,33,35,34,30,29,4,6,38,25,39,26,12,32,37,31,40,41,27,28,1,10,13,8,22,16,19,17,11,14,7,18,9,15,20,21 : 41

6.2.Results on Reduced Dataset

The Table 7 given below presents the results of machine learning techniques on the reduced dataset consisting of 11 features. In the second set of experiments 11 features given by CfsSubsetEval and BestFirst Search are provided as input to each of the technique and results of each technique are documented.

Table 7: Results on Reduced Dataset

Sno	Group	Technique	CC	TP Rate	FP Rate	Precision	Recall	F-Measure	RA	KS	MAE
8.	Rule Based	DT	99.745	0.997	0.001	0.997	0.9997	0.997	1	0.9957	0.0016
		CR	78.537	0.785	0.061	0.677	0.785	0.713	0.937	0.6318	0.0203
		ZeroR	56.837	0.568	0.568	0.323	0.568	0.412	0.5	0	0.0514
		OneR	98.081	0.981	0.005	0.978	0.981	0.978	0.988	0.9675	0.0017
		PART	99.946	0.999	0	0.999	0.999	0.999	1	0.9991	0.0001
9.	Bayes Rule	BayesNet	99.718	0.997	0	0.998	0.997	0.997	0.997	0.9952	0.0003
		NaiveBayes	96.164	0.962	0	0.99	0.962	0.973	0.999	0.9539	0.0037
		NBUpdatable	96.164	0.962	0	0.99	0.962	0.973	0.999	0.9359	0.0037
10.	Functions	MLP	99.279	0.993	0.001	0.993	0.993	0.991	0.999	0.988	0.001
		SMO	99.255	0.993	0.007	0.993	0.993	0.991	0.999	0.9874	0.793
11.	Lazy Learners	IBk	99.869	0.999	0	0.999	0.999	0.999	1	0.9978	0.0001
		Kstar	99.768	0.998	0	0.998	0.998	0.998	1	0.996	0.0003
		LWL	98.041	0.98	0.008	0.964	0.98	0.972	0.999	0.9664	0.0038
12.	Tree	DecisionStump	78.538	0.785	0.061	0.677	0.785	0.713	0.973	0.0203	0.6318
		J48	99.944	0.999	0	0.999	0.999	0.999	1	0.999	0.0001
13.	Meta-Algorithm	AdaboostM1	97.592	0.976	0.006	0.959	0.976	0.967	0.993	0.959	0.0477
14.	Misc	InputMappedClassifier	56.837	0.568	0.568	0.323	0.568	0.4214	0.5	0	0.0514

7. CONCLUSION

In this paper, a comparative analysis of various machine learning strategies for network intrusion detection was performed. The experiments were carried on benchmark dataset (KDDCUP99) for intrusion detection. We have performed two sets of experiments, one on a full dataset having 41 features and one on the reduced one with only 11 elements attributes. Experiments showed that classification algorithms doesn't depend all the 41 features thus the technique could easily get better results with an appreciable cutback in resources needed by working on the same dataset with reduced number of attributes.

8. REFERENCES

[1] J. M. Kizza, *Computer Network Security*. Springer Science & Business Media, 2005.

[2] R. Heady, G. F. Luger, A. Maccabe, and M. Servilla, *The architecture of a network level intrusion detection system*. Department of Computer Science, College of Engineering, University of New Mexico, 1990.

[3] J. Crume, *Inside internet security: What hackers don't want you to know*. Pearson Education, 2000.

[4] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer networks*, vol. 51, no. 12, pp. 3448–3470, 2007.

[5] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, *Supervised machine learning: A review of classification techniques*. 2007.

[6] C. A. Catania and C. G. Garino, "Automatic network intrusion detection: Current techniques and open issues," *Computers & Electrical Engineering*, vol. 38, no. 5, pp. 1062–1072, 2012.

[7] Y. Hamid, M. Sugumaran, and V. Balasaraswathi, "IDS Using Machine Learning - Current State of Art and Future Directions," *British Journal of Applied Science & Technology*, vol. 15, no. 3, pp. 1–22, Jan. 2016.

[8] D. Barbara, J. Couto, S. Jajodia, L. Popyack, and N. Wu, "ADAM: Detecting intrusions by data mining," in *In Proceedings of the IEEE Workshop on Information Assurance and Security*, 2001.

[9] D. Anderson, T. Frivold, and A. Valdes, *Next-generation intrusion detection expert system (NIDES): A summary*. SRI International, Computer Science Laboratory Menio Park, CA, 1995.

[10] M. A. Hall, "Correlation-based feature selection for machine learning," The University of Waikato, 1999.

[11] J. Zhang and M. Zulkernine, "A hybrid network intrusion detection technique using random forests," in *First International Conference on Availability, Reliability and Security (ARES'06)*, 2006, p. 8–pp.

[12] D. M. Farid and M. Z. Rahman, "Anomaly network intrusion detection based on improved self adaptive bayesian algorithm," *Journal of computers*, vol. 5, no. 1, pp. 23–31, 2010.

[13] S. M. Bridges and R. B. Vaughn, "Fuzzy data mining and genetic algorithms applied to intrusion detection," in *Proceedings of 12th Annual Canadian Information Technology Security Symposium*, 2000, pp. 109–122.

[14] B. Luo and J. Xia, "A novel intrusion detection system based on feature generation with visualization strategy," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4139–4147, 2014.

[15] "Machine learning - Wikipedia, the free encyclopedia." [Online]. Available: https://en.wikipedia.org/wiki/Machine_learning.

[16] S. Khan, "Ethem Alpaydin. Introduction to Machine Learning (Adaptive Computation and Machine Learning Series). The MIT Press, 2004.," *Natural Language Engineering*, vol. 14, no. 01, pp. 133–137, 2008.

[17] "Prediction: Machine Learning and Statistics," *MIT OpenCourseWare*. [Online]. Available: http://ocw.mit.edu/courses/sloan-school-of-management/15-097-prediction-machine-learning-and-statistics-spring-2012/.

[18] A. O. Adetunmbi, S. O. Falaki, O. S. Adewale, and B. K. Alese, "Network intrusion detection based on rough set and k-nearest neighbour," *International Journal of Computing and ICT Research*, vol. 2, no. 1, pp. 60–66, 2008.

[19] A. A. Olusola, A. S. Oladele, and D. O. Abosede, "Analysis of KDD'99 Intrusion detection dataset for selection of relevance features," in *Proceedings of the World Congress on Engineering and Computer Science*, 2010, vol. 1, pp. 20–22.

[20] A. Avalappampatty Sivasamy and B. Sundan, "A Dynamic Intrusion Detection System Based on Multivariate Hotelling's T 2 Statistics Approach for Network Environments," *The Scientific World Journal*, vol. 2015, 2015.

[21] P. G. Jeya, M. Ravichandran, and C. S. Ravichandran, "Efficient classifier for R2L and U2R attacks," *International Journal of Computer Applications*, vol. 45, no. 21, p. 29, 2012.

[22] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Survey on incremental approaches for network anomaly detection," *arXiv preprint arXiv:1211.4493*, 2012.

[23] "KDD Cup 1999 Data." [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[24] W. Lee, S. J. Stolfo, and others, "Data mining approaches for intrusion detection," in *Usenix security*, 1998.

[25] "JRip - Pentaho Data Mining - Pentaho Wiki." [Online]. Available: http://wiki.pentaho.com/display/DATAMINING/JRip.

[26] K. P. Murphy, "Naive bayes classifiers," *University of British Columbia*, 2006.

[27] M. Khan and S. M. K. Quadri, "Evaluating Various Learning Techniques for Efficiency," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 2, no. 2, pp. 326–331, 2012.

[28] W. Iba and P. Langley, "Induction of one-level decision trees," in *Proceedings of the ninth international conference on machine learning*, 1992, pp. 233–240.