ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

Fuzzy Based Feature Selection for Intrusion Detection System

[1] P.Indira priyadarsini, [2] Ch.Anuradha, [3] P.S.R. Chandra Murty
[1] Department of CSE, KKR & KSR Institute of Technology & Sciences, Guntur-522017, AP,India.
[2] Department of CSE, PVP Siddhartha Engineering college, Vijayawada-52007, AP,India.
[3] Department of CSE,Acharya Nagarjuna University college of Engineering, Guntur-522510, AP,India

Abstract:- **An Intrusion Detection System (IDS) gathers and evaluates information from different locations, and finds potential security risks that include exterior as well as inside of the organization. It contains an enormous volume of data with irrelevant and redundant features which result in longer processing time and poor detection rate. So, feature selection should be empowered as an important characteristic for better performance on massive datasets. Feature selection refines the high dimensional data sets by removing over fitting and curse of dimensionality problems mainly in the domain of machine learning. The perceptive of feature selection lies in increasing the accurateness. In this paper, Fuzzy_Chi_Euc algorithm was given for selecting best features in KDD Cup 99 data set. In this algorithm integration of two filtering methods is done. The fuzzy inference rules are applied for selecting the features. The classification is carried out for finding intrusion and normal data using Support Vector Machines (SVMs). From the experiments conducted it is shown, most significant and relevant features are thus helpful for classification, which, in turn, reduce the time of training with better classification accuracy.**

Keywords:- **Intrusion Detection System, Machine learning, over fitting, Chi square distance, Euclidean distance, Fuzzy inference rules, Support Vector Machine.**

I. INTRODUCTION

Nowadays, the Internet and worldwide connectivity are extending very fast, the damage to computer systems in businesses and organizations are unresolved. Even though in several anti-virus software's, many secured network protocols exist they are insufficient to give assurance for security. Therefore, Intrusion Detection System (IDS) is becoming acute component to the security infrastructure in organizations [1]. Constructing highly efficient IDSs is notably challenging issue and incorporated as a noteworthy area of research, while it is theoretically impossible to build a system without defects [2]. There are several machine learning algorithms exploited for identifying intrusions such as neural networks, genetic algorithms, Support Vector Machines, Bayesian networks and Ant colony optimization.IDS inspects all the data for identifying intrusions, leading to a pitfall in detecting suspicious behavior [3]. Examining all the data will result in poor detection rate, so in order to excel performance; data to be processed should be reduced. This reduction can be achieved by feature selection or feature reduction processes. The machine imported data are prodigious to search and analyze it. Hence feature selection can handle the dynamically shifting environments. A structured data mining scheme was set up for exploring audit data and building proficien intrusion detection models [4]. The key motivation of feature selection techniques lies in removing over fitting of the data, therefore further examination will be viable. Reducing dimensionality leads to effectual removal of irrelevant and redundant data, better learning capability and speeding up results [5]. The foremost functionality of the researcher depends on exploring the best feature selection algorithm and constraints for a specific classifier for a given data set. The best subset acquired comprises a minimum number of dimensions that which in turn contributes an improvement in learning accuracy [6]. The aspect of feature selection is to depute a subset of the features instead of using all the features which make more complex for classification. Many researchers have shown that feature selection is adopted as a tool for building effectual data models [7]. This wider use of data preprocessing techniques results in remodeling of known models for allied schemes or absolutely new proposals. However, the latest growth of dimensionality of data produces a major difficulty for numerous current feature selection techniques in accordance with effectiveness and performance. Since the filtering methods have low costs and low accuracy rates, merging two methods possibly

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

improve the accuracy [8]. So in this paper, integrated feature selection methods are exploited for choosing reliable features and eliminating irrelevant features thus improving accuracy for intrusion detection process. Hence in our work, feature selection process was carried out by fuzzy_chi_euc algorithm for reducing KDD cup 99 dataset and produced good results. After the data set has been reduced using proposed algorithm, Support Vector Machines (SVMs) are exploited for classifying intrusion and normal data. The left over portion of the paper is ordered as follows. Section 2 provides prior works done on building IDSs using feature selection. Section 3 gives an outline of SVMs, the mathematical overview of chi square and Euclidean algorithm.2 Section 4 gives the description of the KDD cup 99 data set. The proposed IDS model with its architecture is given in Section 5. Section 6 provides experimental work conducted and outcomes obtained. The last section provides conclusion and future scope

II. RELATED WORKS

Chi Square feature selection was applied on SAGE data set to reduce feature set and achieved better results using Naive Bayes, SVM, C4.5, RIPPER and Nearest Neighbor techniques [7]. Recently Mohammed [8] projected a mutual information based algorithm for selection of optimal features for classification and shown that their algorithm selects more critical features by using Least Square Support Vector Machine and other state-of-the-art methods. Abraham [9] has built proficient IDS by two feature selection algorithms containing Bayesian networks (BN) and Classification and Regression Trees (CART) and an integration of BN and CART achieved better accuracy results and lower computational costs. An information gain has exploited in feature selection and shown notable results [10]. Filtering methods are proposed using a hybrid intelligent technique, by combining filters approach and a classifier for making a clever choice and accordingly achieved excellent results [11]. Saurabh [12] have built effective and efficient intrusion detection system by using Feature vitality based reduction method (FVBRM) for finding a reduced set of significant features using NSL-KDD data set. In another related work, Factor analysis was applied to get best features through which efficient IDS was obtained [13] [14]. K.Bajaj [15] showed how feature reduction can improve the detection accuracy; they reduced the features using information gain, gain ratio and correlation. While M. Sharma [16], used feature Quantile filter and Chi-Squared for condensing the number of features. Others introduce Genetic Algorithms along with Linear Discriminant Analysis as a hybrid feature selection method [17]. Correlation based feature selection as a filter method was given to find top significant features and achieved good classification accuracy by applying to five high dimensional data sets [18, 19]. Experiments showed that merging feature selection methods could possibly progress classification accuracy [20].

III. MACHINE LEARNING APPROACH: SCOPE

3. 1. Classification: Support Vector Machines

Nowadays SVMs are remarkable areas for research and also became powerful tools in machine learning. Support Vector Machines (SVMs) are the building machines for classification process using support vectors. They are presented by Vapnik [21]. These are made using Statistical Learning Theory (SLT).They are precise on training samples and have a superior generalization ability on testing samples. With SVMs both linear and nonlinear decision boundaries are identified through optimization problem [22]. Mostly SVMs solve a two-class problem through splitting the dataset by drawing a maximal marginal hyper plane well-defined by a set of support vectors.The dataset is separated using a hyper plane in such a way that maximizes the margin (solid line in the figure below), this can be done by extending both the marginal lines at both sides. The solid line is called Maximal Marginal Hyper plane (MMH). Then the support vectors are taken as a subset of training dataset which plays a crucial role in classification process; hence the process is named as Support Vector Machines. If SVM is not able to separate into two classes, then it solves by mapping input dataset into high dimensional dataset employing a kernel function. Then in the high dimensional space, it is able to classify with good accuracy. There are several kernel functions used in SVM classification like linear, polynomial and Gaussian.

3.2 Chi Square

The chi-square approach is a statistical method of independence to determine the dependence of two variables. It is basically used in correspondence analysis and canonical correspondence analysis. From the definition of chi-square it can be easily deduced for the application as feature selection. In some cases, it is treated as weighted Euclidean distance. Here calculating chi-square distance is done for every feature variable and the class label. Based on the expected frequency and observed frequency, chi-square distance is calculated [23]. The chi-square distance, denoted by $\chi 2$, amid two points x = [

$x_1, x_2 \ldots x_m]$ and y = [$y_1, y_2 \ldots y_m$] is defined as:

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

$$\chi^2 \sum_{i=1}^{j} (f_0 - f_e)^2 / f_e \qquad (1)$$

Where f_0 is the actual data in the data set and f_e is the expected data given as $f_e = \frac{Sr\ Sc}{GT}$ and Sr is the row sum, Sc is the column sum and GT is Grand total.

3.3 Euclidean Distance

Euclidean Distance is the trivial distance metric mainly used as filter approach in the field of data mining. It is also called as a Euclidean norm or Euclidean metric. It is mainly based on Pythagorean Theorem from the basic mathematics [24].So it is also referred as a Pythagorean metric. The formulation is specified as the square root of the summation of the squares of the differences among the subsequent coordinates of the two points. The Euclidean distance among two points with n dimensions $P = (p_1, p_2, p_3 \ldots p_n)$ and $Q = (q_1, q_2, q_3 \ldots q_n)$ is given as:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2} \qquad (2)$$

We compute this distance for every attribute in the dataset. i.e. calculating the distance between each attribute and class label. This is done for all the 41 attributes.

4. KDD Cup 99 DATA SET

The KDD Cup 99 dataset used in the experiments has been taken from the Third International Knowledge Discovery and Data Mining Tools Competition. Each connection record is given with 41 features. The list of attributes contains both continuous type and discrete type variables, which are statistical distributions, crooked intensely from each other, so the identification of intrusions becomes a very tough task. There are 22 categories of attacks from the following four classes: Denial of Service (DoS), Probe, Remote to Local (R2L) and User to Root (U2R) [25].The dataset has 391458 DoS attack records, 97278 normal records, 4107 Probe attack records, 1126 R2L attack records and 52 U2R attack records. This is the dataset taken from only 10 percent of the original data set. The 41 features are given in the order A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z, AA,AB,AC,AD,AE,AF,AG,AH,AI,AJ,AK,AL,AM,AN,AO and the class label AP

V. INTRUSION DETECTION SYSTEM: MODEL

Computationally capable Intrusion detection system is built by fuzzy based feature selection algorithm and SVM as a classification tool. The proposed IDS model presents an entire structure for selecting the best sub set of KDD cup 99 dataset which will proficiently characterize normal traffic and abnormal traffic. Here, a fuzzy based feature selection algorithm works by combining two filtering methods for fastening the training process. The main intention of this approach is to apply fuzziness to the filtering methods. This was given based on chi-square distance and Euclidean distance metrics. The proposed Framework of the IDS containing the following components:

- Attainment of MN_KDD data set
- Pre-processing the data set
- Fuzzy based feature selection
- Detection using SVM
- Evaluating results

It mainly includes two phases. In the first phase, feature selection is done based on the fuzzy chi-square and fuzzy Euclidean techniques. The scores obtained will be sorted and accordingly, fuzziness is applied for getting best features. In the second phase, classification of attack and normal data is done using Support Vector Machines.The mechanism for merging several methods is a Fuzzy Inference System (FIS) [26] which permits, formulating the inference rules in a linguistic approach, by therefore following the natural reasoning. The main aim of FIS lies in the rapidity and proficiency to apply on both large and small datasets satisfactorily. The process begins with complete data set containing N features and M records. The algorithm for fuzzy_chi_euc is shown in figure 1. For all the features chi-square distance is calculated and sorting is done. And three degrees of membership is given to chi_val and they are namely low, mid, high based on some threshold value. Call them as f_chi.In the same way, for each feature, Euclidean distance is calculated and sorting is done. And three degrees of membership is given to euc_val and they are namely low1, mid1, high1.Call them as f_euc.It is shown in line 11 and line 12 of the algorithm. Afterwards discard low, low1 as these are ineffectual in the selection procedure. Then fuzzy if-then rules are applied to the med, high, med1, high1 scores. This was given in line 14 of the algorithm. Next, the reduced feature set is traced from the resultant four intermediate feature sets. This will be the final optimistic subset obtained. The resulting sub set contains n features. The reduced data set is now trained and tested using SVM classifier. The evaluation metrics for the proposed agenda are detection rate, false alarm rate; the number of seconds taken for constructing the model and also accuracy.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

Fuzzy_Chi_Euc Algorithm

Input: Data set (MxN),chi_val, euc_dist, class,f_chi, f_euc, k_1, k_2, k_3 k_4, n

Output: Classification measures {detection rate, false positive rate, Accuracy}

Start:
1. For i=0; i<M;i++
2. For j=0; j<N;j++
3. for each feature
4. calculate chi_val= CHIVAL(feature, class)
5. end
6. for each feature
7. calculate euc_dist=EUC(feature, class)
8. end
9. sort chi_val
10. sort euc_dist
11. f_chi={low,med,high}
12. f_euc={low1,med1,high1}
13. delete low,low1;
14. if(f_chi ==med and f_euc==med1) then
k_1= # of features obtained
if(f_chi ==med and f_euc==high1)
then k_2= # of features obtained
if(f_chi ==high and f_euc==mid1)
then k_3= # of features obtained
if(f_chi ==high and f_euc==high1) then
k_4= # of features obtained
15. n= {k1Uk2U k3Uk4}
16. Reduced dataset=D(MxN,Mxn)
17. Classify (dataset, SVM)
End
End

Fig 1: Proposed Fuzzy_chi_euc algorithm

VI. EXPERIMENTAL WORK AND RESULTS

We conducted all the experiments and obtained results using Java 1.6 and Weka 3.6.9 on the platform Windows 2007 with 3.40 GHz CPU and 2.0GB of RAM. WEKA is an open source Java code produced by researchers at the University of Waikato in New Zealand [28]. In the entire experiments conducted 10 fold cross validation is made. The dataset is partitioned at random into 10 equal parts in which the classes are taken approximately as equal size as in the full dataset. Each part is kept out in turn and the training is performed on left over 9 parts, then its testing is conducted on hold out set. The training is done in a total of 10 times on different training sets and lastly, the 10 error rates are averaged for attaining overall error estimate.

6.1 Procurement of Data set

Collection of 10% KDD cup 99 dataset is done, it has been normalized. The researchers in their works have used the portion of the dataset from the KDD cup 99 data set for building IDSs not including the complete train or test dataset [27].So, we have taken a subset of KDD cup 99 containing 14207 records and call as "MN_KDD dataset". The size of the dataset is taken in proportion to the relative size of the KDD cup 99 dataset and R2L,U2R records are taken as usual from the original data set. MN_KDD data set contains 3000 normal records, 10000 DoS records, 574 Probe records, 401 R2L records and 52 U2R records. In the dataset all the symbolic attributes are converted to numeric. It includes protocol_type, service, flag and class label. So, for the class labels, we have assigned 1, 2,3,4,5 values for U2R, R2L, Probe, DOS, Normal respectively.

6.2. Feature Selection and assessment of results obtained

After executing all the preprocessing techniques appropriately, the fuzzy_chi_euc algorithm is applied. With the 41 features in the training dataset, they are defined as F= $\{F_1, F_2 \ldots F_{41}\}$ and its corresponding class label C. Let any feature Fj= $\{x_{j1}, x_{j2} \ldots x_{jn}\}$ where j is the total number of features and n is the number of training instances and C= $\{c_1, c_2, c_3, c_4, c_5\}$.The chi-square distance is calculated as mentioned in section 3 for feature1 and corresponding class label and likewise to all the remaining 40 features. Subsequently, they are sorted as low, med, high based on some threshold value. Now, Calculate the Euclidean distance for each feature based on the formula given in equation (1) in section 3.Then the Euclidean distance metric becomes

$$d_j(F_j, C) = \sqrt{(x_{1j} - c_i)}\,2 \qquad -(3)$$

Then we get the corresponding distances for each feature. Hence a total of 41 features has obtained the Euclidean distance. Based on this they are arranged in the order highest to lowest. Accordingly, they are taken as low1, med1, high1. For illustration, for the feature1, Euclidean distance is calculated as $\sqrt{\sum}$ (feature1-class label)2 .That is given as 1)calculating the difference for every vector point in Attr1 and class label and it is squared.2)Then Step (1) is done for all the 14027 instances 3)Sum of 14027 values is taken 4)Then square root of the sum obtained. This is for calculating the Euclidean distance for feature1.Above four steps have to be done for all the 41 features. Then for each feature, we get some desired value. The work flow of the proposed IDS model is shown in figure 2 below. In total, three experiments were done on MN_KDD data set containing 14207 records and 41 attributes with five different class names such as U2R, R2L, Probe, DoS, Normal i) The dataset was taken containing 14027 records with no feature selection (fs), i.e. Taking 41 attributes and then applied SVM. ii) In the second experiment Euclidean distance is applied as feature selection metric, through which 29 attributes are

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

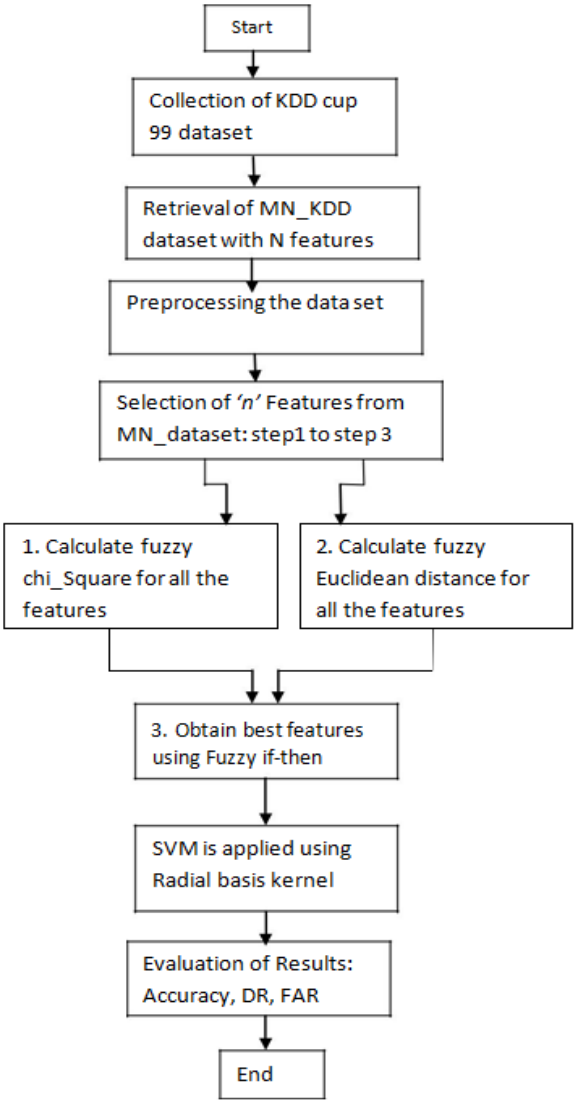

Fig 2: Work flow of proposed IDS model

selected and then SVM is applied for classification. The features chosen here are C,D,F,G,H,I,J,K,N,O,P,Q,R,S,T,U,W,X,Y,Z,AA,AB,AC,AF, AG,AL,AM,AN and AO. iii) In the third experiment, proposed Fuzzy_chi_euc algorithm given in section 5 is applied, through which best features are obtained. The meticulous score of Euclidean distance, chi square distance are given in the corresponding tables Table I, II. The values of the low, low1 are discarded since they are useless for the selection process. Then fuzzy if-then rules are applied to the resultant med, high, med1, high1 values. According to this, we get k1, k2 , k3, k4 intermediate feature set, each one containing a different number of features. From proposed algorithm, we get nine features as the best ones. They are given in Table III. Hence nine different features are selected from 41 features, and then applied SVM classifier. With this reduced number, rapidity of the learning process gets increased. In the Experiments conducted we performed four-class classification. We separate the data as two classes namely "Normal" and "Others" (DoS, Probe, U2R, R2L) patterns, where the Others is the group of four classes of attack instances in the data set. The idea is to divide normal and attack traffic. Reiterating the same process for all classes using SVMs is done. Training is conducted with the RBF (radial basis function) kernel. The training procedure is carried out with 10 fold cross validation.

Table I: Chi-Square score of med and high

Sno	Chi Square Score	Feature
Sorted as *Med*		
1	10135.89	H
2	10083.38	AD
3	9837.2	AH
4	9437.68	AJ
5	9369.56	F
6	9249.53	AG
7	8604.19	AL
8	8204.91	X
9	7885.6	Y
10	7122.5	B

Sorted as *High*		
1	30215.7	E
2	14892.3	C
3	12397.4	W
4	11837.3	Al
5	11695.2	D

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

Table II: The Euclidean score of med1 and high1

Sno	Euclidean Score	Feature
Sorted as *Med1*		
1	491.57	G
2	491.70	H
3	491.74	I
4	491.68	N
5	491.75	O
6	491.45	P
7	491.29	S
8	491.75	T
9	491.74	U
10	491.60	Y
11	491.58	Z
12	491.55	AA
13	491.16	AB
14	491.58	AL
15	491.64	AM
16	491.24	AN
17	491.08	AO

Sorted as *High1*		
1	2221.57	C
2	701.02	D
3	1299.58	F
4	492.21	Q
5	20414.29	W
6	20446.24	X
7	29178.79	AC
8	21435.73	AF
9	28030.43	AG

Table III: Features selected based on the proposed algorithm

S.no	Features selected based on the proposed algorithm	Feature Name
1	C	service
2	D	flag
3	F	dst_bytes
4	H	wrong_fragment
5	W	count
6	X	srv_count
7	Y	serror_rate
8	AG	Dst_host_srv_count
9	AL	Dst_host_serror_rat

To evaluate Fuzzy_chi_euc Algorithm, the main metrics used are the Accuracy (ACC), the detection rate (DR) and the false alarm rate (FAR). The Accuracy is defined as the percentage of instances that are classified correctly. They are interpreted in the table IV. The assessment of proposed algorithm with Euc+SVM and with no feature selection using SVM is done and given in table V. Comparatively the proposed approach with fuzzy_chi_euc algorithm routs the remaining two approaches.While the time is taken to build proposed model is 667 sec using SVM. It produced results with 96.4% accuracy, classifying 13696 instances correctly out of 14027 instances. The accuracy of all the three models is compared in figure 5 below. It shows model versus accuracy in percentage.

VII. CONCLUSION AND FUTURE ENHANCEMENTS

Building efficient IDS is the vital process which will be made by preferring the pertinent feature selection procedure as a module of preprocessing, as the KDD cup 99 dataset is a massive data set with nearly five million records. In the experiments done, Fuzzy_Chi_Euc algorithm was developed for selecting best features works by merging two filtering methods, the fuzzy chi-square and fuzzy Euclidean distance. Since these days it is worth paying attention in using dimensionality reduction techniques for improving and building well proficient Intrusion Detection Systems (IDSs), upcoming research seizes alterations of the proposed scheme

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 11, November 2017

and upgrading it to attain enhanced performance and automation by developing classifiers which are more precise for the detection of attacks. The problem with KDD Cup 99 data set is, it is imbalanced, and subsequently, the outcomes obtained cannot be appropriate. As further enhancements, research can be extended inspecting detection rates of both U2R and R2L attacks and probably increasing the detection rates and also firm choice of essential features. Since the 41 feature set contains 241-1 possible subsets, the task is tough selecting the optimistic subset which contributes in increasing accuracy rate. In future work, exploring robust techniques for choosing best features to develop light weight Intrusion Detection System models can be done.

Table IV: Detection rate and FAR of Fuzzy_chi_euc method in comparison with Euc+SVM and SVM with no feature selection

Evaluation metrics/Algorithm		Fuzzy_ chi_ euc	Euc+S VM	SV M
Nor mal	DR (%)	99.8	95.6	93. 5
	FAR(%)	1.63	10.3	19
DoS	DR (%)	99.2	90	79. 4
	FAR(%)	1.56	6.4	11. 2 6
Prob e	DR (%)	93.7	84.3	78
	FAR (%)	1.82	1.33	1.3 6
R2L	DR (%)	52.3	13.2	11
	FAR (%)	16.3	0.10	0.4 4
U2R	DR (%)	61.5	17.3	9.6
	FAR (%)	12.7	0.08	0.1 2

Table V: The comparison of proposed approach with two methods: Euclidean and no feature selection

	Fuzzy_c hi_ euc Algorith m	Euc+SVM	SVM
Total number of instances	14207	14207	14207
Correctly classified	96.4%	87.34%	79%
Incorrectly classified	3.5%	12.58%	20.8%
Time taken to build the model	667 sec	823 sec	982 sec
Overall Detection rate (avg)	81.3%	60.09%	54.32 %

REFERENCES

[1] Hasan, M.A.M., Nasser, M. and Pal, B.:On the KDD'99 Dataset: Support Vector Machine Based Intrusion Detection System (IDS) with Different Kernels, IJECCE, vol. 4, 1164-1170, (2013).

[2] Adebayo, O.A., Shi, Z., Shi, Z. and Adewale, O.S.: Network Anomalous Intrusion Detection Using Fuzzy-Bayes.Intelligent Information Processing III,vol.228,525-530, (2006). doi:https://doi.org/10.1007/978-0-387- 44641-7_56.

[3] Chen, Y., Abraham, A. and Yang, J.:Feature Selection and Intrusion Detection Using Hybrid Flexible Neural tree. Advances in Neural Networks—ISNN 2005, 439-444. (2005).doi.http://dx.doi.org/10.1007/11427469_71.

[4] W. Lee, S. J. Stolfo, and K. W. Mok.: Adaptive intrusion detection: A data mining Approach.AI Review, vol.14,I(6):533 – 567, (2000), doi.http://dx.doi.org/10.1023/A:1006624031083.

[5]Khalid,S,Khalil,T, & Nasreen S.: A survey of feature selection and feature extraction techniques in machine learning. Science and Information Conference (SAI) 372–378, (2014).

[6]L. Ladla and T. Deepa .:Feature Selection Methods and Algorithms. International Journal on Computer Science and Engineering (IJCSE), Vol.3 I (5), 1787-1797, (2011).

[7] Xin Jin, Anbang Xu,Rongfang Bie1, Ping Guo.: Machine Learning Techniques and Chi-Square Feature Selection. Springer-Verlag Berlin Heidelberg LNBI 3916, pp. 106 – 115, (2006).

[8] Mohammed A.Ambusaidi, Xiangjian He, Priyadarsi Nanda,Zhiyuan Tan.:Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm", IEEE Transactions on Computers ,Vol. 65, I(10), Oct(2016).

[9] A Abraham, S Chebrolu, J P. Thomas .:Feature deduction and ensemble design of intrusion detection systems. Computers & Security,Vol. 24, I(4), Pages 295-307, (2005). http://dx.doi.org/10.1016/j.cose.2004.09.008.

[10] Zhe Gao,Yajing Xu,Fanyu Meng,Feng Qi,Zhiqing Lin .: Improved information gain based feature selection for text

categorization. 4th International Conference on Wireless Communications, Vehicular Technology, Information Theory and Aerospace & Electronics Systems (VITAE), (2014).doi: 10.1109/VITAE.2014.6934421.

[11] Panda, Mrutyunjaya, Ajith Abraham, and Manas Ranjan Patra.: A Hybrid Intelligent Approach for Network Intrusion Detection,International Conference on Communication Technology and System Design 2011,pp 1-9 ,Procedia Engineering 30 (2012).

[12] Saurabh, Mukherjee, and Neelam Sharma.:Intrusion detection using naive Bayes classifier with feature reduction," Procedia Technology vol.4, 119-128, (2012). doi: 10.1016/j.protcy.2012.05.017.

[13] Eunhye kim, seungmin lee, kihoon kwon.:feature construction scheme for efficient intrusion detection system, journal of information science and engineering vol.26, 527-547 (2010).

[14] P Indira Priyadarsini, I Ramesh Babu.:Building Efficient Intrusion Detection System using Factor Analysis and Support Vector Machines, in International Journal of Engineering Research and Technology, Vol3, (2014).

[15] K.Bajaj and A. Arora.:Dimension Reduction in Intrusion Detection Features Using Discriminative Machine Learning Approach. In International Journal of Computer Sciences.vol. 10, no. 4, pp. 324–329, (2013).

[16]M. Sharma, K. Jindal, and A. Kumar.:Intrusion Detection System using Bayesian Approach for Wireless Network, Int. J. Comput. Appl., vol. 48, I(5), pp. 29– 33, (2012).

[17] H. M. Imran, A. Bin Abdullah, M. Hussain, and S. Palaniappan.:Intrusions Detection based on Optimum Features Subset and Efficient Dataset Selection, IJEIT,vol. 2, I (6), pp. 265-270, (2012).

[18] Vimal Kumar Dubey,Amit Kumar Saxena.:Hybrid classification model of correlation-based feature selection and support vector machine, IEEE International Conference on Current Trends in Advanced Computin (ICCTAC) (2016), doi: 10.1109/ICCTAC.2016.7567338.

[19] M. Hall, Correlation Based Feature Selection forMachine Learning, Doctoral Dissertation, The University of Waikato, Department of Computer Science, (1999).

[20] Z. Karimi and A. Harounabadi.:Feature Ranking in Intrusion Detection Dataset using Combination of Filtering Methods,Int. J. Comput. Appl. (0975 – 8887), vol. 78,I(4), pp. 21–27, (2013).

[21] Cortes C.,Vapnik V.:Support vector networks, in Proceedings of Machine Learning pp.273–297, (1995).

[22] P Indira priyadarsini, Nagaraju Devarakonda,I Ramesh Babu,"A Chock Full Survey on Support Vector Machines", International Journal of Computer Science and Software Engineering,Vol 3,I(10),(2013).

[23]http://web.pdx.edu/~newsomj/da1/ho_chisq.pdf

[24]http://www.econ.upf.edu/~michael/stanford/maeb4.pdf

[25] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani .:A Detailed Analysis of the KDD CUP 99 Data Set, Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defence Applications, (2009).

[26] Timothy J. Ross.:Fuzzy Logic with Engineering Applications (3rd Edition), John Wiley & Sons, New Jersey, USA,.

[27] Chebrolu, S., Abraham, A. and Thomas, J.P. (2004) Hybrid Feature Selection for Modeling Intrusion Detection Systems. Neural Information Processing, 1020-1025, (2010). http://dx.doi.org/10.1007/978-3-540-30499-9_158.

[28] M. Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer.: The WEKA data mining software: an update," ACM SIGKDD Explorations Newsletter, vol. 11, I(1), pp. 10-18, (2009).