JCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Volume-5, Issue-7 E-ISSN: 2347-2693

Enhancing Classification Accuracy using Feature Subset Selection in Intrusion Detection System (IDS)

S.A. Margaret[1*], S. Padmavathi[2]

[1*] Dept. of Computer Science, Marudupandiyar College of Arts and Science, Thanjavur, India
[2] Dept. of Computer Science, Marudupandiyar College of Arts and Science, Thanjavur, India

*Corresponding Author: margaretjulie1970@gmail.com

Available online at: www.ijcseonline.org

Received: 20/May/2017, Revised: 05/Jun/2017, Accepted: 19/Jul/2017, Published: 30/Jul/2017

***Abstract*—** Intrusion detection system (IDS) look into field has developed immensely in the previous decade. Enhancing the detection rate of client to root (C2R) assault class is an open research issue. Current IDS utilizes all information elements to recognize intrusions. A portion of the elements might be excess to the detection procedure. The reason for this experimental examination is to distinguish the vital elements to enhance the detection rate and diminish the false detection rate. The researched highlight subset choice strategies enhance the general exactness, detection rate of C2R assault class and furthermore diminish the computational cost. The exact outcomes have demonstrated a recognizable change in detection rate of C2R assault class with include subset determination methods.

***Keywords*—** Feature subset selection; classification; preprocessing; Intrusion detection system.

I. INTRODUCTION

Intrusion detection system (IDS) is a supplement of customary systems insurance procedures to be specific client confirmation, information encryption, and firewall as the principal line of resistance for PC and systems security. IDS have been perceived as extreme research territory in the previous decade attributable to the quick increment of advanced assaults on PC systems. The goal of IDS is to recognize any unknown or unordinary action as an endeavor of breaking the security arrangement of PC systems. There are three general classifications of detection approaches: 1) grouping; 2) information bunching; and 3) irregularity based approach. In the arrangement approach, we order the given informational index into various sorts of assaults. Information order is an administer machine learning strategy. In the second information bunching approach, we classify the given informational index into various classifications on the premise of closeness and divergence. Information bunching is an unsupervised machine learning procedure. In the third inconsistency based approach, we distinguish deviations from the ordinary utilization conduct examples to recognize the intrusion. When all is said in done, peculiarity based approach is semi regulated machine learning procedure. Each approach has claim points of interest and impediments over alternate methodologies.

This manuscript address the main approach i.e. information characterization for building intrusion detection demonstrate. There are many issues and difficulties in the current information characterization approaches. The first is called lopsidedness class issue, where the quantity of cases of assault class is extremely uncommon. That is the informational index circulation mirrors a huge dominant part of typical class and a minority of assault class. The second is to recognize the proper classifier for intrusion detection from countless classifiers. The third is pre-preparing the crude information with the goal that handled information can be utilized as contribution for a classifier. Precision of a classifier relies on the nature of info informational index. The nature of information relies on the quality element vector of informational collection.

The KDD Cup 1999 dataset is publically accessible benchmark for assessing of IDS procedures. KDD informational index has countless illustrations. Copy cases may negatively affect the preparation procedure of machine learning classifiers. In this manuscript, we utilized NSL-KDD Cup 99 dataset. This informational index is an enhanced form of KDD informational index and publically accessible. The informational collection has 41 highlights. Preparing informational collection comprise of 29 distinct assaults. These assaults are additionally ordered into four unique sorts: 1) disavowal of administration assaults (DoS); 2) examining assaults (Probe); 3) remote to neighborhood assaults (R2L); 4) client to root assaults (C2R); There is no information arrangement calculation can be prepared effectively with KDD informational collection to perform detection for C2R or R2L assault classifications. The quantity of cases for these assault sorts is less in the informational collection -. It has been shown that every one of the 41 elements of KDD dataset are not critical and might be wiped out, without fundamentally fall apart the execution of the IDS. Sung and Mukkamala decided just 19 highlights out of 41 highlights utilizing a trial-blunder approach. Chebrolu and Abraham et al., decided 17 and 12 highlights utilizing Bayesian system and CART classifier separately. They revealed enhanced detection rate of all assault classes with the exception of C2R sort of assaults. In addition many as of late distributed methodologies are likewise confronting low detection rate for C2R assault –. In this way the key goal of this manuscript is to enhance the general precision of the IDS and enhance the detection rate of C2R assault class.

The other issue in building a computational speedier classifier is that KDD informational collection has countless. Every one of these components are not applicable to accomplish the better detection rates of assaults. This manuscript recognizes the applicable sub set of elements so general precision is likewise held at a similar level of without decreasing the elements of KDD informational collection. The goal of the manuscript is to investigate the change of the detection rate of C2R assault class utilizing highlight subset determination procedures. This manuscript researches the reasonableness of run based classifiers to enhance the general precision of IDS. We experimentally contrast the aftereffects of two classifiers and all elements and decreased subset of components. We accomplish better execution of both classifier calculations utilizing decreased subset of components. This manuscript utilized proficient usage of existing methods accessible in WEKA information mining instrument. Whatever is left of this manuscript is organized as takes after: Section II depicts the exploratory setup. Area III presents utilized hypothesis and ideas. Segment IV presents results and dialog. Segment V presents conclusions.

II. Experimental Setup

This segment is additionally isolated into three subsections.

A. NSL-KDD Cup Dataset

The KDD Cup 1999 dataset is utilized a benchmark for assessing of IDS strategies. The lion's share of cases in this data set have been removed from the DARPA 1998 IDS assessment. KDD data set has a colossal number of repetitive cases. Copy cases may negatively affect the preparation procedure of machine learning classifiers. All through in this exact examination, we utilized NSL-KDD Cup 99 dataset. This data set is an enhanced form of KDD data set and publically accessible. In this manuscript, we have utilized KDDTrain+ and KDDTest+ data sets that have 20% of the records of the whole NSL-KDD data set. The data set has 41 highlights. These elements can be arranged into four gatherings, to be specific fundamental elements, content components, time-based elements and host-based elements. The points of interest of these 41 highlights are exhibited in table I. Preparing data set comprise of 29 distinct assaults. These assaults are additionally classified into four distinct sorts and displayed in table II. The short presentation of four distinct sorts of assaults is as per the following:

1) Denial of Service Attack (DoS): It is a class of assaults in which an assailant makes some processing or memory asset excessively occupied or, making it impossible to deal with genuine demands, or denies honest to goodness clients access to a machine.

2) Probing Attacks (Probe): It is a class of assaults in which an aggressor checks a system of PCs to accumulate data or find known vulnerabilities. An aggressor with a guide of machines and administrations that are accessible on a system can utilize this data to search for misuses.

3) Remote to Local Attacks (R2L): It is a class of assaults in which an assailant sends bundles to a machine over a system yet who does not have a record on that machine; misuses some weakness to increase nearby access as a client of that machine account on the system and can abuse powerlessness to pick up root access to the system.

KDDTest+ comprise of the (29 + 18 = 37) unique assaults. 18 novel assaults are the extra assaults in the test dataset. These assaults are not accessible in the preparation dataset. The quantity of cases of every four class of assaults in KDDTrain+ and KDDTest+ data sets is given in table III. C2R assault is just 0.04% in the KDDTrain+ and 0.037% in the KDDTest+. In the preparation stage the system develops a model utilizing the preparation data. The test data is gone through the developed model to distinguish the intrusion in the testing stage.

B. WEKA Data Mining Tool

This manuscript utilized a Waikato Environment for Knowledge Analysis (WEKA) data mining apparatus. The apparatus was planned at college of Waikato in 1993. The present rendition WEKA 3.7.11 is utilized as a part of this exact investigation. It has a gathering of best in class machine learning calculations for data mining assignments. It contains apparatuses for data pre-preparing, arrangement, relapse, grouping, affiliation standards, and perception. WEKA comprises of four applications to be specific Explorer, Experimenter, Knowledge Flow, Simple Command Line Interface. WEKA is a notable and generally utilized as a part of scholarly group because of following reasons:

- Publically accessible through GNU General Public License

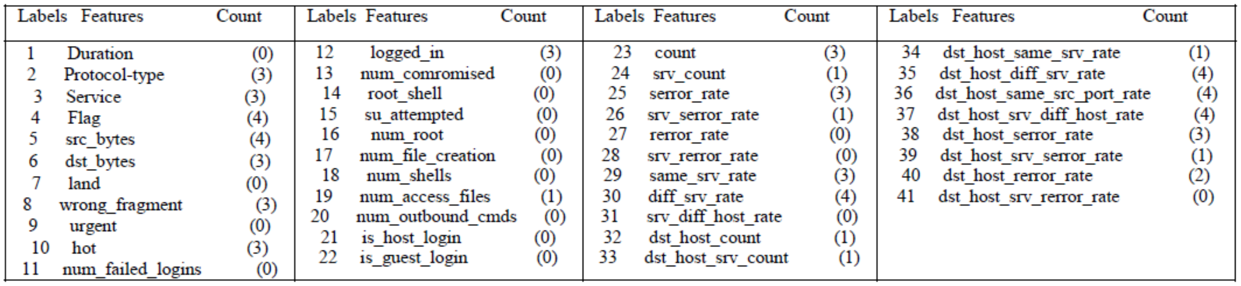
- Available for all cross –platforms
- Easily comprehended by fledgling clients
- Has the model approval office
- Based on Java dialect
- Online instructional exercises are accessible

We utilized an exceptionally proficient usage of machine learning strategies of WEKA.

C. Research Methodology

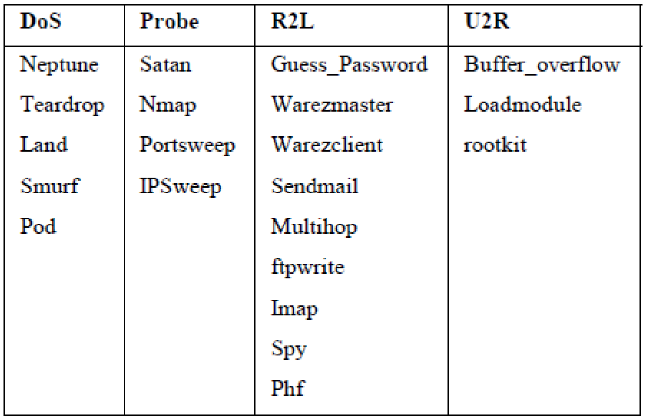
We decide the pertinent subset of components among 41 elements of KDD dataset with the goal that we can likewise hold a similar level of exactness of the classifier. We utilize wrapper strategies and channel strategy for choosing the pertinent subset of elements. The preparation and testing data records are spared with diminished arrangement of elements. We look at the general precision and detection rate of classifiers with every one of the 41 highlights and with diminished subset of components. The proposed approach enhances the general exactness of both administer based classifiers and furthermore enhances the detection rate of C2R kind of assaults.

Table 1. Complete Set of Features With Count



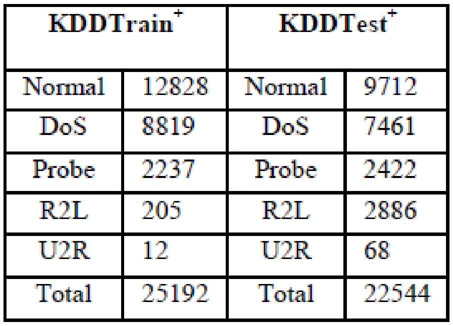
Labels	Features	Count	Labels	Features	Count	Labels	Features	Count	Labels	Features	Count
1	Duration	(0)	12	logged_in	(3)	23	count	(3)	34	dst_host_same_srv_rate	(1)
2	Protocol-type	(3)	13	num_comromised	(0)	24	srv_count	(1)	35	dst_host_diff_srv_rate	(4)
3	Service	(3)	14	root_shell	(0)	25	serror_rate	(3)	36	dst_host_same_src_port_rate	(4)
4	Flag	(4)	15	su_attempted	(0)	26	srv_serror_rate	(1)	37	dst_host_srv_diff_host_rate	(4)
5	src_bytes	(4)	16	num_root	(0)	27	rerror_rate	(0)	38	dst_host_serror_rate	(3)
6	dst_bytes	(3)	17	num_file_creation	(0)	28	srv_rerror_rate	(0)	39	dst_host_srv_serror_rate	(1)
7	land	(0)	18	num_shells	(0)	29	same_srv_rate	(3)	40	dst_host_rerror_rate	(2)
8	wrong_fragment	(3)	19	num_access_files	(1)	30	diff_srv_rate	(4)	41	dst_host_srv_rerror_rate	(0)
9	urgent	(0)	20	num_outbound_cmds	(0)	31	srv_diff_host_rate	(0)			
10	hot	(3)	21	is_host_login	(0)	32	dst_host_count	(1)			
11	num_failed_logins	(0)	22	is_guest_login	(0)	33	dst_host_srv_count	(1)			

Table 2. List of Attacks in KDDTRAIN+



DoS	Probe	R2L	U2R
Neptune	Satan	Guess_Password	Buffer_overflow
Teardrop	Nmap	Warezmaster	Loadmodule
Land	Portsweep	Warezclient	rootkit
Smurf	IPSweep	Sendmail	
Pod		Multihop	
		ftpwrite	
		Imap	
		Spy	
		Phf	

Table 3. Details of Dataset



KDDTrain$^+$		KDDTest$^+$	
Normal	12828	Normal	9712
DoS	8819	DoS	7461
Probe	2237	Probe	2422
R2L	205	R2L	2886
U2R	12	U2R	68
Total	25192	Total	22544

III. USED RELATED THEORY AND CONCEPTS

This area is additionally partitioned into four subsections.

A. Pre-handling the data set

Machine learning classifiers take input data as genuine numbers. In the pre-preparing, we need to change over all esteems relating to all components in genuine numbers. All in all, we decide the mean and standard deviation comparing to each element then we change esteems in the range from - 1 to +1. The same pre-handling process is connected on both preparing and testing data set. This pre-handling step is not connected in this experimental investigation.

B. Highlight Subset Selection Techniques

In the second phase of pre-handling, we select the pertinent subset of elements among 41 includes in KDD data set so we can hold a similar level of precision of classifier. There are two techniques for subset include determination accessible in WEKA. These are wrapper strategies and channel techniques. In the wrapper strategies, we utilize blend of characteristic evaluator and hunt technique. The accompanying characteristic evaluators are accessible in WEKA tool stash.

- CfS-Subset-Eval: Consider prescient estimation of each quality independently, alongside the level of excess among them.
- Gain-Ratio-Attribute-Eval: Evaluates quality in view of pick up proportion as for class
- Info-Gain-Attribute-Eval: Evaluate quality in view of data pick up as for the class.
- One-R-Attribute–Eval: It utilizes the preparation data for assessment or it can apply cross-approval.
- Principal-Component: Perform key part examination and change.
- Relief-Attribute-Eval: It is an example based trait evaluator.
- Symmetrical-Uncert-Attribute-Eval: It assess qualities in light of symmetrical uncertainity concerning the class.
- Wrapper-Subset-Eval: It utilizes classifier to assess ascribes and cross-approval to evaluate the precision.

The accompanying hunt strategies are accessible in WEKA tool stash

- Best first: Greedy slope moving with backtracking.
- Greedy stepwise: Greedy slope moving without backtracking; alternatively produce positioned rundown of traits.
- Ranker: Rank individual ascribes as per their assessment.

We additionally utilized regulated trait choice channel in this observational examination. The utilized mixes for highlight sub set choice are exhibited in table IV.

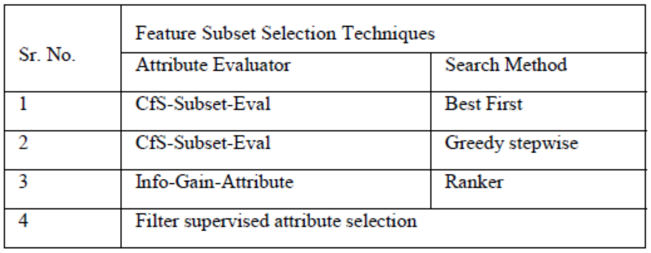
C. Administer based Classifiers

There is a substantial number of existing classifier calculations. They can be ordered into Bayesian methodologies, tree-based classifiers, control based models, work based classifiers, apathetic classifiers, multi-case classifiers and gathering approaches. We utilized Decision Table and PART govern based classifiers in this investigation. A govern based classifier creates an arrangement of IF-THEN principles for characterization. The tenets are anything but difficult to create and simple to decipher. These calculations limit the quantity of false-positive mistakes. The execution of administer based classifiers is tantamount to choice tree-based classifiers.

The concise presentation of utilized classifiers is as per the following:

1) Decision Table: It speaks to every one of the blends of conceivable conditions for a choice in forbidden frame by mapping every one of the conditions and activities in segments. It utilizes the nearestneighbor technique to decide the class for every illustration that is not secured by a choice table section, rather than the table's worldwide dominant part, in view of a similar arrangement of properties.

Table 4. Feature Subset Selection Techniques



Sr. No.	Feature Subset Selection Techniques	
	Attribute Evaluator	Search Method
1	CfS-Subset-Eval	Best First
2	CfS-Subset-Eval	Greedy stepwise
3	Info-Gain-Attribute	Ranker
4	Filter supervised attribute selection	

2) PART: It gets rules from incomplete choice tree. It utilizes C4.5's heuristics to fabricate the tree with an indistinguishable client characterized parameters from in J4.8.

D. Execution Metrics for Classifiers

This segment presents measurements for surveying how "precise" our classifier is at identifying the distinctive sorts of C2R assaults. We utilized KDDTrain+ for developing the model. We gauged the classifier's precision on a KDDTest+. Assume P is the quantity of positive cases (assaults) and N is the quantity of negative illustrations (ordinary data). The accompanying wording is utilized to characterize numerous execution measurements for classifiers.

- → Genuine positives (TP): grouping an intrusion as an intrusion.
- → Genuine negatives (TN): effectively ordering ordinary data as typical.
- → False positives (FP): erroneously ordering ordinary data as an intrusion.

→ False negatives (FN): erroneously ordering an intrusion as ordinary data.

The classifier assessment measurements can be characterized as takes after:

→ exactness = acknowledgment rate = (TP+TN)/(P+N) (1)

→ mistake rate =misclassification rate = (FP+FN)/(P+N) (2)

→ review = genuine positive rate = detection rate = TP/P (3)

→ specificity = genuine negative rate = TN/N (4)

→ accuracy = (TP)/(TP+FP) (5)

→ F-measure = (2x Precision X Recall)/(Precision+ Recall) (6)

The consequences of classifiers can likewise be contrasted with deference with root mean squared mistake (RMSE), preparing time, heartiness, versatility and interpretability.

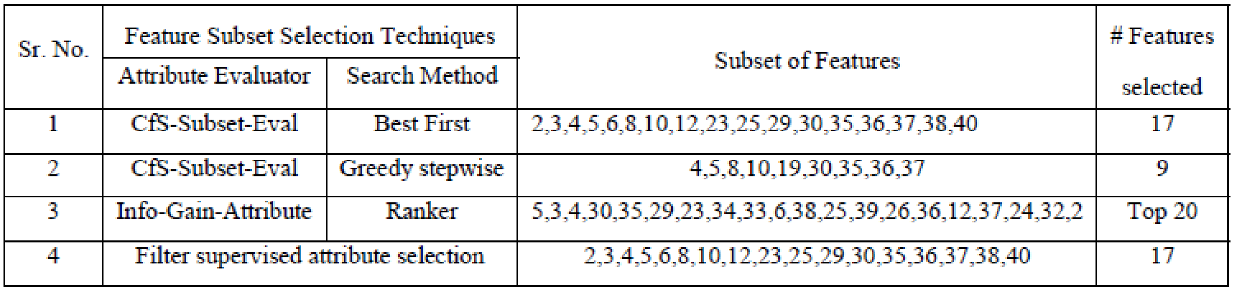
IV. Results And Discussions

We have distinguished four component sub set choice techniques. The subtle elements are given in Table IV. The aftereffects of highlight sub set choice strategies are spoken to in Table V. The section 2 and 3 in Table V demonstrate the mix of evaluators and inquiry strategy individually. Sections 4 and 5 demonstrate the sub set of chose components and aggregate number of chose highlights individually. The quantity of chose highlights is a sub set of unique 41 components of KDD data set. This is additionally broke down that the pursuit techniques specifically best-first and voracious stepwise are just good with Cfs-subset-Eval. Additionally, we display the include of each chose highlight Table I i.e. how frequently the characteristic is chosen utilizing four distinctive element sub set determination strategies. It demonstrates that the elements having check 4 and 3 are the most applicable in the detection procedure of IDS. Those elements having number 0 are not applicable by any stretch of the imagination. We can presume that we can't dispense with all components of one classification to be specific fundamental elements, content elements, time-based elements and host-based elements of KDD data set. We require a few elements of every class for enhancing detection rate of assaults. As the component sub set is chosen in the second phase of preprocessing step, the diminished preparing data set is utilized as a contribution to the classifier for preparing stage. We developed multi-class classifier. The five principle classes i.e. Typical, Probe, DoS, R2L and C2R are partitioned into these multi classes. We tried the built multi-class classifier utilizing testing data set. Amid the testing stage, we consider similar components that are utilized amid the preparation stage. We get the reenactment aftereffects of all multi-classes amid the testing stage. For quickness, we exhibit outline data in Table VI to Table IX. For correlation of the C2R assault sort, we display rundown data utilizing every one of the 41 elements and utilizing chose sub set of elements. Tables VI - IX introduce general precision, root mean square mistake (RMSE), genuine positive rate (TPR), false positive rate (FPR), F-measure. Table X shows the normal of detection rate for three assaults of C2R class. On the premise of Tables VI-X, The conclusion is as per the following:

- The general precision of the two classifiers has been enhanced essentially utilizing lessened subset of components.
- The general RMSE of the two classifiers have been enhanced fundamentally utilizing lessened subset of elements.
- Detection rate is enhanced from 0 to 0.007 utilizing Decision Table with 17 highlights. Detection rate is enhanced from 0 to 0.017 utilizing Decision Table with 9features.
- Detection rate is enhanced from 0.096 to 0.153 utilizing PART Classifier with 20 highlights. Detection rate is enhanced from 0.096 to 0.201 utilizing PART Classifier with 17 highlights.
- The detection rate of PART classifier is altogether superior to anything Decision Table classifier.
- The detection rate of PART classifier utilizing 17 highlights is altogether superior to PART classifier utilizing 17 highlights.
- The Loadmodule assault of C2R class is not distinguished in any examinations.
- The one most critical perception is that false positive rate is zero in all mix revealed here.
- The time required to build the run based classifiers are essentially diminished in every single revealed result with decreased arrangement of elements.

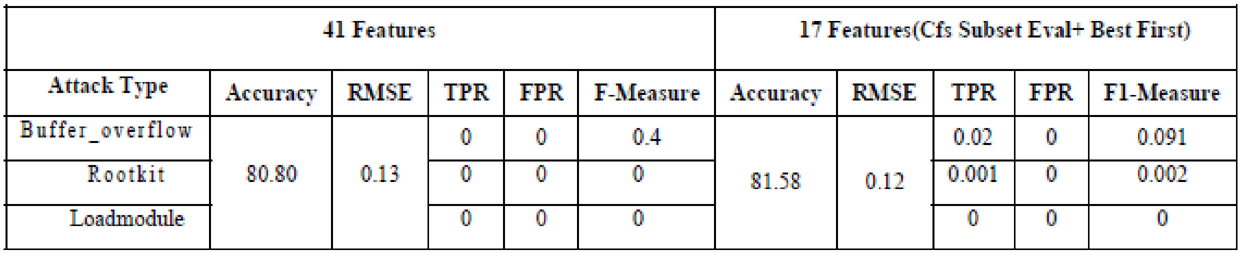
The KDDTest+ dataset has four extra assaults. These are Sqlattack, Perl, Ps, Xterm. These assaults are not distinguished in any examinations. They are additionally not announced here. This is a characteristic property of a classifier that those assaults are just recognized those are available in the preparation dataset.

Table 5. Subset of Selected Features



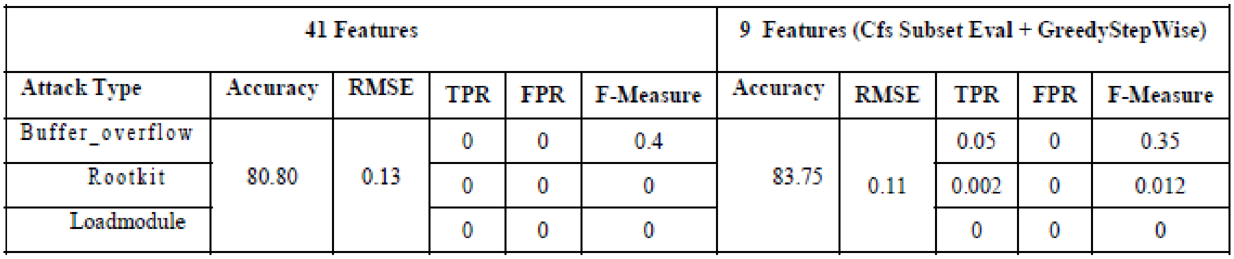
Sr. No.	Feature Subset Selection Techniques		Subset of Features	# Features selected
	Attribute Evaluator	Search Method		
1	CfS-Subset-Eval	Best First	2,3,4,5,6,8,10,12,23,25,29,30,35,36,37,38,40	17
2	CfS-Subset-Eval	Greedy stepwise	4,5,8,10,19,30,35,36,37	9
3	Info-Gain-Attribute	Ranker	5,3,4,30,35,29,23,34,33,6,38,25,39,26,36,12,37,24,32,2	Top 20
4	Filter supervised attribute selection		2,3,4,5,6,8,10,12,23,25,29,30,35,36,37,38,40	17

Table 6. Summary Results of Decision Table Classifier



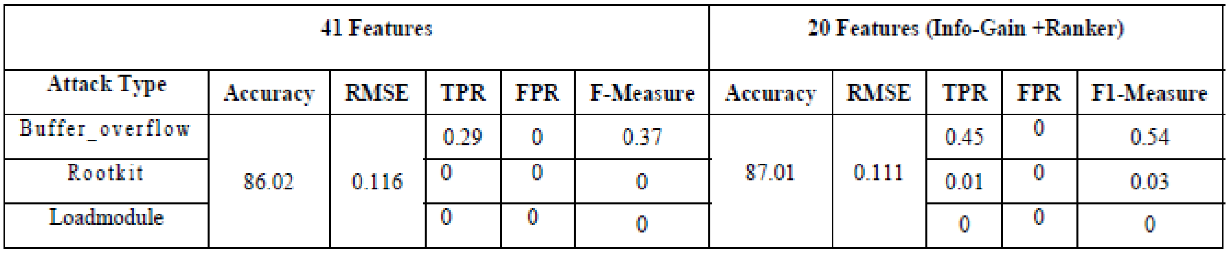
41 Features						17 Features(Cfs Subset Eval+ Best First)				
Attack Type	Accuracy	RMSE	TPR	FPR	F-Measure	Accuracy	RMSE	TPR	FPR	F1-Measure
Buffer_overflow	80.80	0.13	0	0	0.4	81.58	0.12	0.02	0	0.091
Rootkit			0	0	0			0.001	0	0.002
Loadmodule			0	0	0			0	0	0

Table 7. Summary Results of Decision Table Classifier



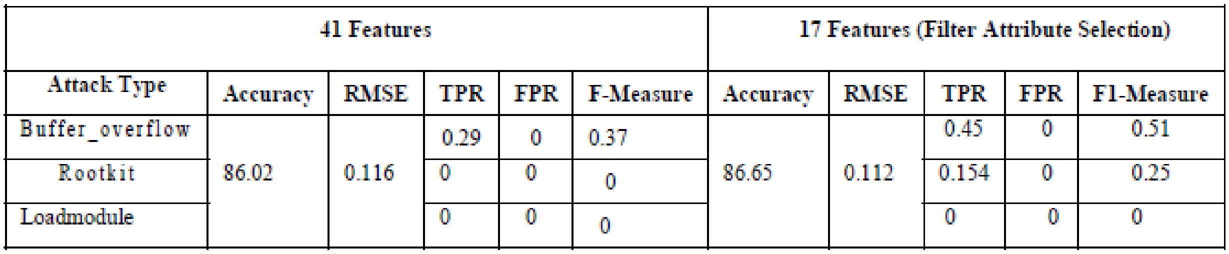
41 Features						9 Features (Cfs Subset Eval + GreedyStepWise)				
Attack Type	Accuracy	RMSE	TPR	FPR	F-Measure	Accuracy	RMSE	TPR	FPR	F-Measure
Buffer_overflow	80.80	0.13	0	0	0.4	83.75	0.11	0.05	0	0.35
Rootkit			0	0	0			0.002	0	0.012
Loadmodule			0	0	0			0	0	0

Table 8. Summary Results of Part Classifier



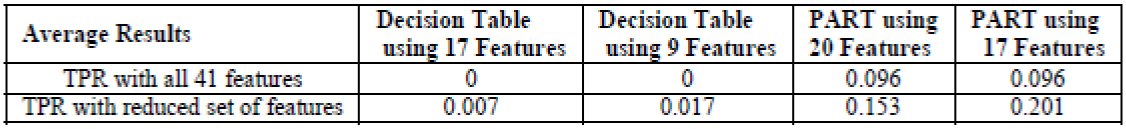
41 Features						20 Features (Info-Gain +Ranker)				
Attack Type	Accuracy	RMSE	TPR	FPR	F-Measure	Accuracy	RMSE	TPR	FPR	F1-Measure
Buffer_overflow	86.02	0.116	0.29	0	0.37	87.01	0.111	0.45	0	0.54
Rootkit			0	0	0			0.01	0	0.03
Loadmodule			0	0	0			0	0	0

Table 9. Summary Results of Part Classifier



41 Features						17 Features (Filter Attribute Selection)				
Attack Type	Accuracy	RMSE	TPR	FPR	F-Measure	Accuracy	RMSE	TPR	FPR	F1-Measure
Buffer_overflow	86.02	0.116	0.29	0	0.37	86.65	0.112	0.45	0	0.51
Rootkit			0	0	0			0.154	0	0.25
Loadmodule			0	0	0			0	0	0

Table 9. Average Summary Results of Detection Rate Of All Four Combination



Average Results	Decision Table using 17 Features	Decision Table using 9 Features	PART using 20 Features	PART using 17 Features
TPR with all 41 features	0	0	0.096	0.096
TPR with reduced set of features	0.007	0.017	0.153	0.201

The appropriateness of govern based classifiers for enhancing the detection rate of C2R class of assaults can be further examinations utilizing broad reenactment. The future research design comprises how to distinguish novel assaults of C2R class those are not accessible in the preparation dataset.

V. CONCLUSIONS

In this manuscript, we researched four techniques for determining subset of elements utilizing characteristic evaluators and pursuit strategies and channel strategy. We decided subset of existing elements on KDD dataset. We assessed the execution on diminished list of capabilities utilizing two control based classifiers. We additionally exhibited execution correlations utilizing distinctive subset of components and all current 41 highlights. We likewise gave which highlights are more important for enhancing the detection rate of C2R class. Observational outcomes uncovered that general precision and detection rate of C2R kind of assaults have been enhanced essentially.

REFERENCES

[1] Feng Yang, K. Z. Mao, Gary Kee Khoon Lee, Wenyin Tang, *“Emphasizing Minority Class in LDA for Feature Subset Selection on High-Dimensional Small-Sized Problems”*, IEEE Transactions on Knowledge and Data Engineering, Vol.27, Issue.1, PP.88 – 101, 2015.

[2] Qinbao Song, Jingjie Ni, Guangtao Wang, *“A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data”*, IEEE Transactions on Knowledge and Data Engineering, Vol.25, Issue.1, PP.1 – 14, 2013.

[3] Yong Liu, Feng Tang, Zhiyong Zeng, *“Feature Selection Based on Dependency Margin”*, IEEE Transactions on Cybernetics, Vol.45, Issue.6, PP.1209 – 1221, 2015.

[4] D. Asir Antony Gnana Singh, S. Appavu Alias Balamurugan, E. Jebamalar Leavline, *“An empirical study on dimensionality reduction and improvement of classification accuracy using feature subset selection and ranking”*, (INCOSET), PP.102 – 108, 2012.

[5] Chieng-Yi Chang, *“Dynamic Programming as Applied to Feature Subset Selection in a Pattern Recognition System”,* IEEE Transactions on Systems, Man, and Cybernetics, Vol.SMC-3, Issue.2, PP.166 – 171, 1973.

[6] Surya S. Durbha, Roger L. King, Nicolas H. Younan, *“Wrapper-Based Feature Subset Selection for Rapid Image Information Mining”,* IEEE Geoscience and Remote Sensing Letters, Vol.7, Issue.1, PP.43 – 47, 2010.

[7] A. Dastanpour, S. Ibrahim, R. Mashinchi, "*Effect of Genetic Algorithm on Artificial Neural Network for Intrusion Detection System*", International Journal of Computer Sciences and Engineering, Vol.4, Issue.10, pp.10-18, 2016.

[8] L. Boroczky, L. Zhao, K. P. Lee, *“Feature Subset Selection for Improving the Performance of False Positive Reduction in Lung Nodule CAD”*, IEEE Transactions on Information Technology in Biomedicine, Vol.10, Issue.3, PP.504 – 511, 2006.

[9] J. Yang, V. Honavar, *“Feature subset selection using a genetic algorithm”,* IEEE Intelligent Systems and their Applications, Vol.13, Issue.2, PP.44 – 49, 1998.

[10] Kashif Javed, Haroon A. Babri, Mehreen Saeed, *“Feature Selection Based on Class-Dependent Densities for High-Dimensional Binary Data”*, IEEE Transactions on Knowledge and Data Engineering, Vol.24, Issue.3, PP.465 – 477, 2012.

[11] Yongxuan Zhu, Xin Shan, Jun Guo, *“Modified genetic algorithm based feature subset selection in intrusion detection system”,* ISCIT 2005, Vol.1, PP.10 – 13, 2005.

[12] Prachi Tembhare, Neeraj Shukla, "*An Integrated and Improved Scheme for Efficient Intrusion Detection in Cloud*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.3, pp.74-78, 2017.