INTRUSION DETECTION IN WIRELESS NETWORK USING ENSAMBLE TECHNIQUES OF DATA MINING

[*]Maninder Jeet Brar
M.Tech Scholar
GKU, Talwandi Sabo

[**]Er. Chamkaur Singh
Assistant Professor
GKU, Talwandi Sabo

Abstract: Data mining is a vital method wherever intelligent strategies square measure applied extract knowledge patterns. It's the method of discovering attention-grabbing pattern and data from massive amounts of information. The information supply will embrace a information, knowledge warehouses, the web, different repositories, or knowledge that square measure streamed into the system dynamically. Data processing aims at discovering patterns in knowledge which can be gift in knowledge. The method should be automatic or semiautomatic. The pattern discovered should be important therein they result in some advantage, sometimes associate degree economic advantage. The information is invariably gift in substantial quantities. The safety attacks will cause severe disruption to knowledge and networks. Therefore, Intrusion Detection System; becomes a crucial a part of each pc or network system. Intrusion detection (ID) may be a mechanism that has security for each computers and networks. Feature choice and have reduction is a crucial space of analysis in intrusion direction system. The dimensions and attribute of intrusion file square measure terribly massive. Because of the massive size of attribute the detection and classification mechanism of intrusion detection technique square measure compromised in terms of detection rate and alarm generation. The most downside is that there'll be the distinction between the new threat discovered and signature being employed in IDS for police works the threat. Approaches of IDS supported the placement of watching square measure Network-based intrusion detection system; and Host-based intrusion detection system; During this paper, completely different the various} author's papers square measure reviewed and different issues square measure sweet-faced. The choice of illustrious and unknown attack has additionally sweet-faced a tangle of classification. There's the multiclass downside throughout the classification of information. Intrusion detection may be a downside of transportation infrastructure protection because of the very fact that pc networks square measure at the core of the operational management of abundant of the nation's transportation. of these issues square measure resolved with the assistance of Intrusion detection on Wi-Fi network victimisation KNN (K Nearest Neighbour), SVM (Support Vector Machine) and GA and ninety nine end result square measure detected with the assistance of SVM.

Keywords: IDS, ANN, Data, security, data mining etc.

I. INTRODUCTION

During the previous few years there's a dramatic increase in growth of laptop networks. There square measure numerous personal further as government organizations that store valuable knowledge over the network. This tremendous growth has posed difficult problems in network and data security, and detection of security threats, ordinarily said as intrusion, has become a really necessary and demanding issue in network, knowledge and data security. The safety attacks will cause severe disruption to knowledge and networks. Therefore, Intrusion Detection System; becomes a very important a part of each laptop or network system. Intrusion detection (ID) may be a mechanism that has security for each computers and networks. Feature choice and have reduction is vital space of analysis in intrusion direction system. The dimensions and attribute of intrusion file square measure terribly giant. Thanks to giant size of attribute the detection and classification mechanism of intrusion detection technique square measure compromised in terms of detection rate and alarm generation. For the advance of intrusion detection method numerous authors and researchers work along for feature reduction and have choice for intrusion detection system. In current situation the feature reduction and choice method concentrate on entropy based mostly technique [2]. Some authors used neural network model such Kyrgyzstani monetary unit and RBF neural network model for classification of intrusion knowledge throughout offensive mode and traditional mode of network traffic. On the mechanism of detection intrusion detection divide into 2 section host based mostly primarily based mostly intrusion detection system and network based intrusion detection system. Host based mostly primarily based mostly intrusion detection system in typically apprehend as signature based intrusion detection system. Instead signature {based based mostly primarily based mostly} intrusion detection system come back together with another variant is termed anomaly based intrusion detection. In anomaly based mostly intrusion detection numerous technique square measures used cherishes supervised learning and unsupervised learning. In network intrusion Detection, freelance and redundancy attributes ends up in low police investigation rate and speed of classification algorithms. Therefore, the way to scale back network attributes to lift performance of classification rules by applying best algorithm has become a search branch of intrusion Detection [2]. Additionally, there's typically associate in nursing initial coaching amount for Associate in nursing intrusion detector to characterize the noticeable object's behavior, and most existing ways square measure supported the idea that prime quality tagged coaching knowledge square measure without delay out there. gift a brand new approach; supported pulse Coupled Neural Networks (PCNN) to spot necessary input options for intrusion detection. Through characteristic the necessary inputs and redundant inputs, a classifier can do the reduced downside size, quicker coaching and additional correct results [1-3,1-4].

II. LITERATURE SURVEY

Yogita Gupta, amphibian genus Khudhair Abbas Ahmed , Sandeep Kautish [2017] is to gift elaborate survey of obtainable literature on recent advancements and notable contributions within the field of applications of knowledge Mining and information Management tools particularly targeted on Medical IP. once presenting introduction regarding the aim and scope of the subject, section 2 of the paper change the ideas of knowledge mining and its typical techniques i.e. Probabilistic Models, Rule Induction, Neural Networks and Analytical Learning and therefore the section ends with presenting information Management construct and its linkage with data processing and bioscience field. In section 3, all the previous relevant works of data of information of information mining and knowledge management are critically analyzed, explained and categorized on the idea of their relevancy that is followed by section four that presents discussion on all the previous works and highlight the benefits and downsides of assorted ways and tools with more scope of future analysis and limitations. The author concludes the paper whereas action on security and privacy considerations of medical knowledge and attract readers' attention towards validation of medical knowledge that is employed for medical judgments and choices.[1] Muhamad Erza Aminanto et.al.[2017] have contemplated the part coefficient techniques in existing machine students and take a goose at however they may be utilised for the precise determination of the essential highlights. thus on approve our thought, we tend to think about Wi-Fi systems since inescapable Internet-of-Things (IoT) gadgets create Brobdingnagian traffics and overcome within the in the meantime. distinctive notable and obscure assaults in Wi-Fi systems stays unimaginable testing assignments. we tend to take a look at and approve the credibility of the selected highlights utilizing a typical neural system. This investigation exhibits that the projected weighted-based machine learning model will beat alternative channel primarily based part selection models. The trial comes regarding not simply show the viability of the projected demonstrate, accomplishing ninety nine.72% F1 score, nonetheless additionally demonstrate that consolidating a weight-based part determination strategy with a light-weight machine-learning classifier that prompts basically increased execution, contrasted with the simplest outcome elaborate within the literature.[2] Aditya Shrivastava1 et.al [2013] have projected a [*fr1] and [*fr1] model for embrace selection and interruption identification. Highlight selection is significant issue in interruption identification. the selection of highlight in assault attribute and typical movement quality is testing trip. the selection of notable and obscure assault is to boot confronted a difficulty of order. PCNN is dynamic system utilised for the procedure of highlight selection in grouping. The dynamic plan of PCNN choose characteristic on determination of entropy. The characteristic entropy is high the part estimation of PCNN organize is chosen and therefore the property estimation is low the PCNN highlight selector diminishes the estimation of highlight determination. once determination of highlight the mathematician piece of facilitate vector machine is incorporated for grouping. Identification rate is high in pressure of alternative neural system model, as an instance, RBF neural system and Kyrgyzstani monetary unit organize. [3] JAYSHRI R. PATEL et.al [2013] projected a method utilizing call Trees order of Intrusion location, as indicated by their highlights into either nosey or non meddling category may be a broadly speaking examined issue. Selection trees are useful to acknowledge interruption from association records. During this paper, we tend to assess the execution of various selection tree classifiers for ordering interruption recognition info. The purpose of this paper is to explore the execution of various selection tree classifiers for positioned interruption identification info. Info Gain is employed to relinquish positioning to interruption identification info. Selection tree classifiers assessed are C4.5, CART, Random Forest and REP Tree. [4] Megha Aggarwal et.al [2013], displayed there's a sensational increment in development of laptop systems. There are totally different personal and conjointly government associations that store vital info over the system. This monumental development has postured testing problems in system and knowledge security, and identification of security dangers, often alluded to as interruption, has changed into an important and basic issue in system, info and knowledge security. The safety assaults will create extreme disturbance info and systems. During this manner, Intrusion Detection System turns into an indispensable piece of every computer or system framework Intrusion location (ID) may be a part that offers security to the 2 PCs and systems. [5]

Table 1: COMPARITIVE SUMMARY OF AVAILABLE LITERATURE

NAME OF AUTHOR(S)	***TECHNIQUE USED***	***SUMMARY***
Muhamad Erza Aminanto et.al.[2017]	weighted-based machine learning model	accomplishing 99.72% F1 score
Ms. Rohini A. Et.al. [2016]	Intrusion Detection system	Defining architecture of IDS
Dr. S.Vijayarani et.al.[2015]	ID Life cycle	Improved security in corporate world and for network users
Aditya Shrivastava1 et.al [2013]	PCNN	Identification rate is high in pressure of other neural system model
Jayshri R. Patel et.al [2013]	C4.5, CART, Random Forest and REP Tree.	Controlling the intrusion
Megha Aggarwal et.al [2013],	Intrusion Detection System (IDS)	gives security to the two PCs and systems
Gaura et al. (2013)	Used Experimental sensor data to perform monitoring non clinical health data	Research performed on non medical data.

Huang et al. (2013)	Online data mining of abnormal period patterns from medical sensor data streams	Applicable only for online tools and difficult to interpret
Clifton et al. (2013)	Proposed personalized e-health monitoring using wearable sensors	Deviations in results of different samples
Venkata Suneetha Takkellapati et.al [2012]	KNN	high precision discovery rate and less mistake rate of KDD CUP 1999 preparing informational index.
Z.Y. Zhuang (2009)	Case Based Reasoning (CBR)	Intelligent DSS for pathology test ordering by GPs with the use of Case Based Reasoning (CBR)

III. METHODOLOGY

This is to detect the intrusion from network. It is based upon weka tool. There are the programmable files containing the information about the dataset. The Intrusion detection system deals with large amount of data which contains various irrelevant and redundant features resulting in increased processing time and low detection rate. Therefore feature selection plays an important role in intrusion detection. There are various feature selection methods proposed in literature by different authors. In this a comparative analysis of different feature selection methods are presented on KDDCUP'99 benchmark dataset and their performance are evaluated in terms of detection rate, root mean square error and computational time.

The proposed step for work is :

Step 1: Process the weka tool.
Step 2: Read the KDDCups 99 dataset on the preprocessing.
Step 3: Apply the attribute selection to select the attributes.
Step 4: Note down the selected attribute and remove the unselected attribute.
Step5: Classify the selected attribute with different classifier.
Step 6: Analyze the different values after the classification.
Step 7: visualize the resulted graph with different values.
Step 8: repeat the step 3 to step 7 for different classifiers.
Step 9 : Stop

IV. RESULT& ANALYSIS

This gives the final results of the research work that is to be implemented in the weka tool. The different figures of the research works are given below

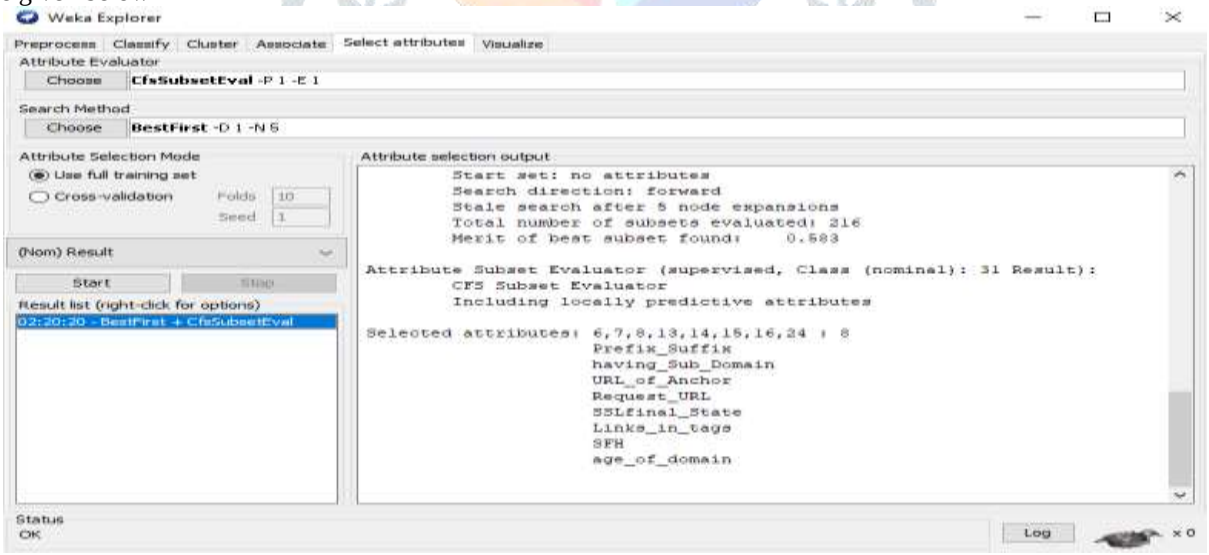

Figure 1: Selected attributes: 6, 7, 8, 13, 14, 15, 16, 24: 8

The figure 1 is the weka tool window with different number of attributes selection. Here CfsSubsetEval attribute evaluator is used to evaluator is used to evaluate the features. The best fit search method is used to search the features in AWID dataset. As you can see the 8 evaluate features like prefix _sufix and others are shown in this figure. All these attributes are classified in next step.

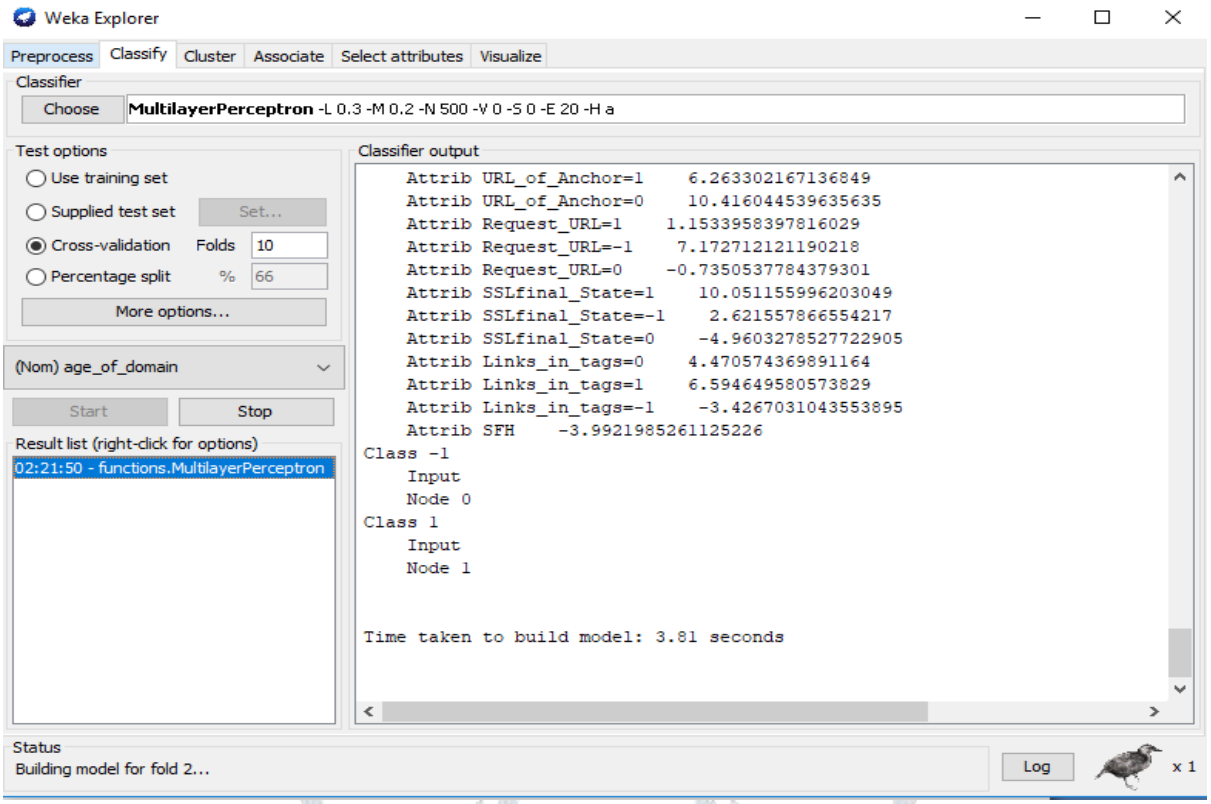

Figure 2: KNN processing on selected features

The figure 2 is the KNN processing on selected features. In this figure Multilayer perception that is the KNN classifier used to classify the selected features that are shown in figure 1. In this figure features are classified with build in time model 3.81 seconds.

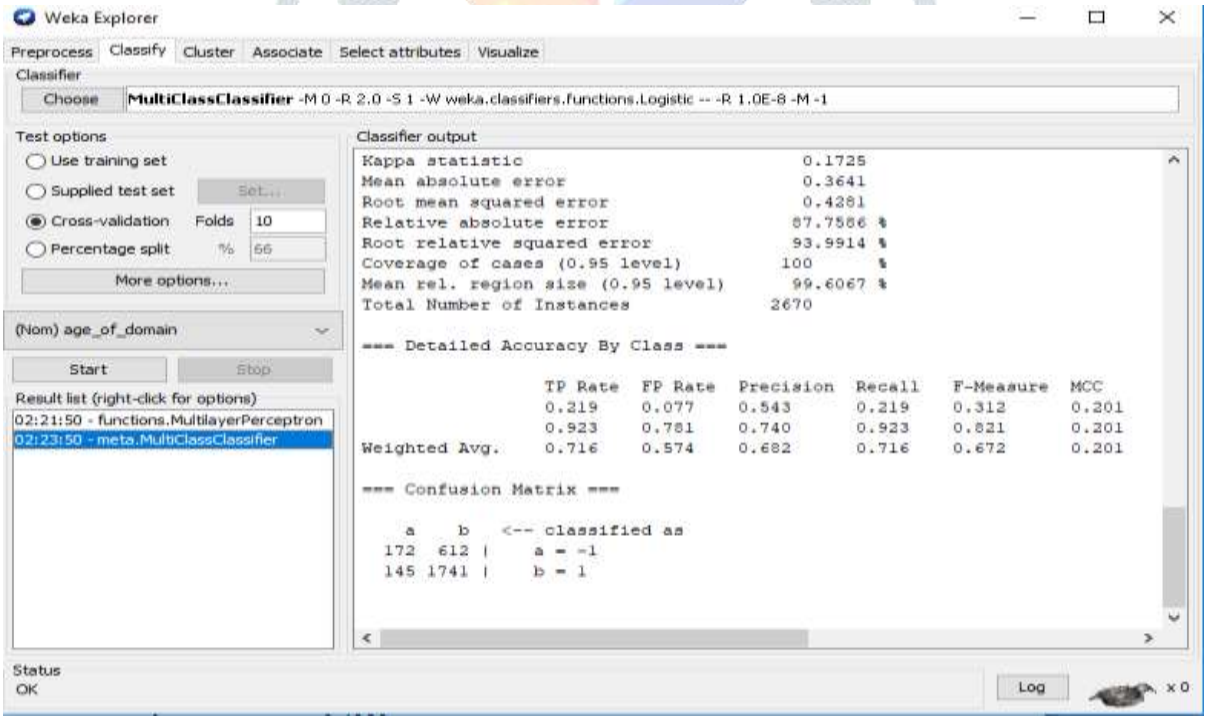

Figure 3: SVM processing on selected features

The figure 3 is the SVM processing on 8 selected features that are shown in figure 1 . In this figure different parameters are shown. These parameters are like Kappa Statistics, Mean absolute error and root mean square error etc. are displayed. It is display the TP rate, FP rate, Precision, Recall and F-measure in this figure. Here is -1and 1 value that is the intrusion and non intrusion attributes value.

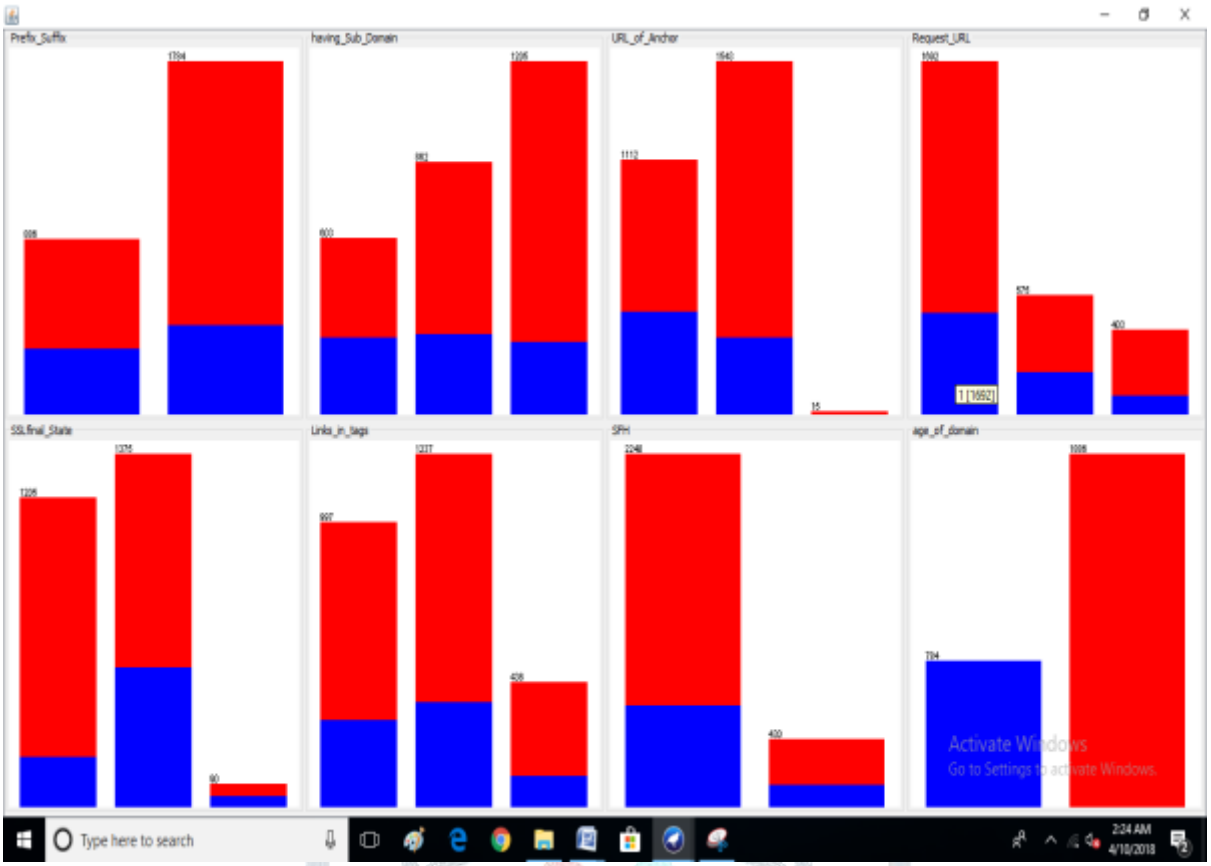

Figure 4: Color visualization of all selected features

The figure 4 is the color visualization of all selected features. It displays the normal and abnormal feature color representation of all features that are selected with the help of best fit method. In this figure red and blue color represents the intrusion detected features and non intrusion detected features.

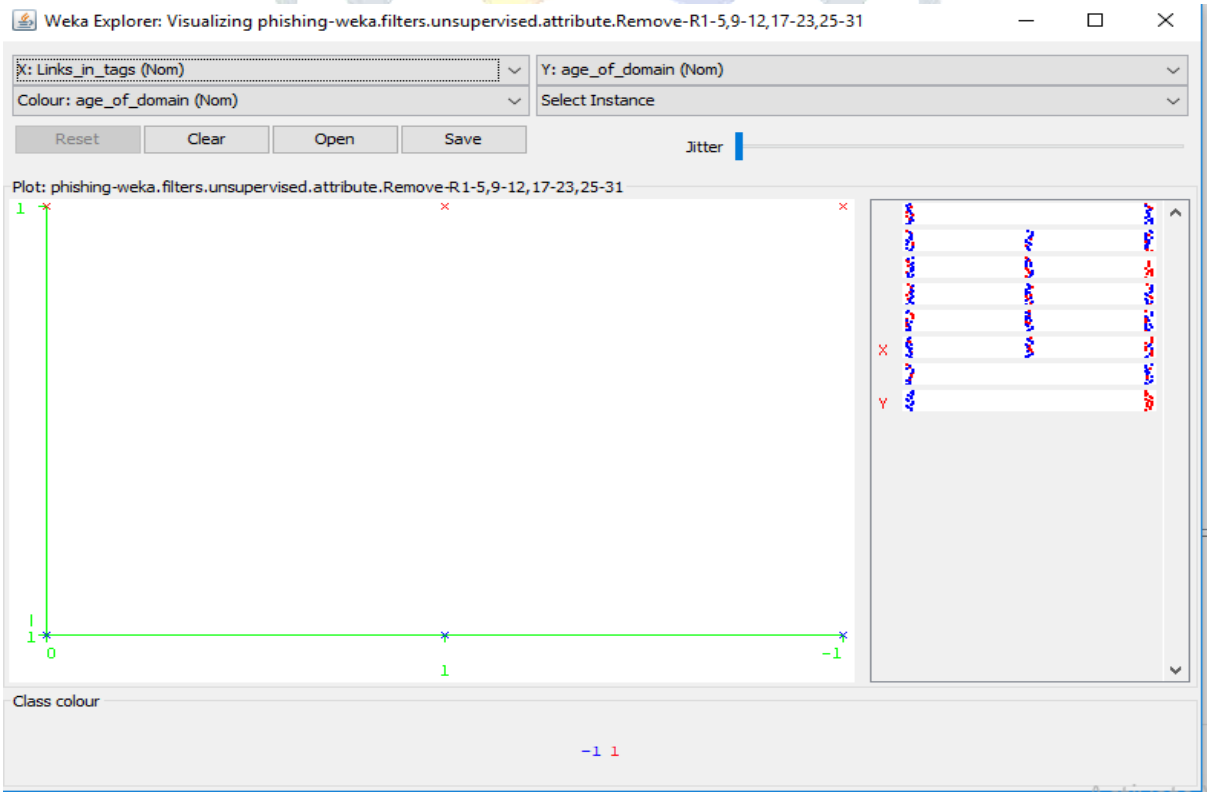

Figure 5: Link in tags feature graph visualization

The figure 5 is the link in tags feature graph visualization. It is one of the selected features. This figure defines the phishing weka filters unsupervised attribute. In this figure 0,1 and -1 is value is displayed based on attribute weight.

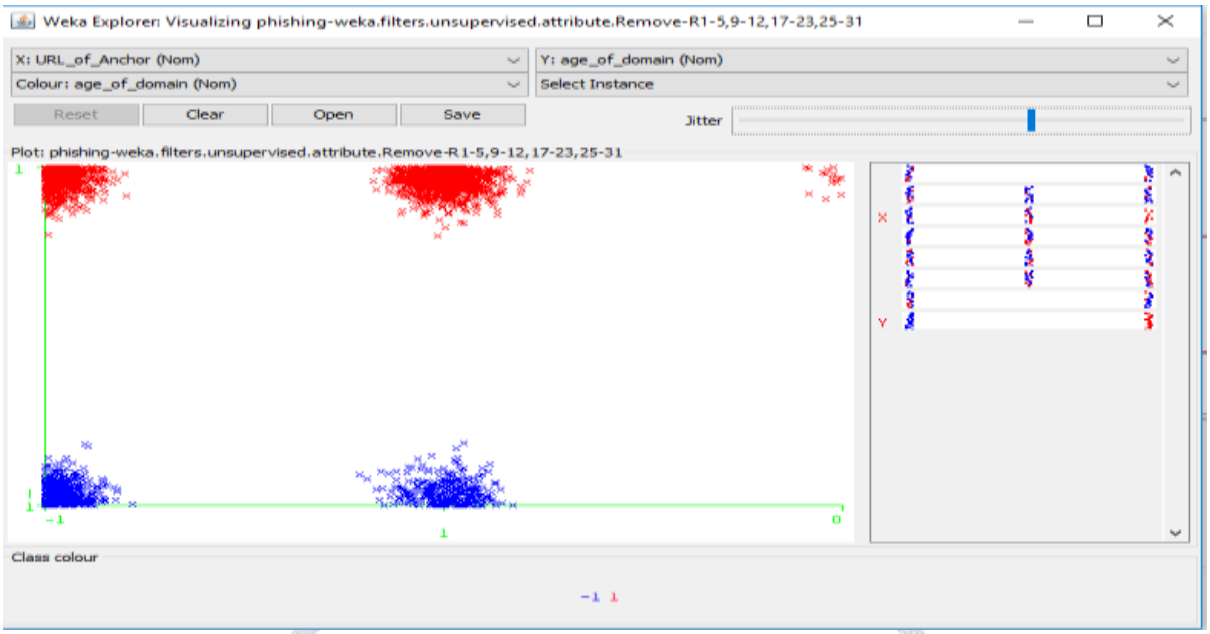

Figure 6: URL of Anchor feature graph visualization

The figure 6 displays the URL of Anchor feature graph visualization. It is also the phishing weka filters unsupervised attribute. In this figure 0,1 and -1 is value is displayed based on attribute weight.

Table 1 : KNN , SVM and GA processing Instances

	KNN		SVM		GA	
Correctly Classified Instances	1905	71.3483 %	1913	71.6479 %	1958	73.3%
Incorrectly Classified Instances	765	28.6517 %	757	28.3521 %	712	26.6%

Table 2 : KNN and SVM performance parameters

Parameters	**KNN**	**SVM**	**GA**
Kappa statistic	0.2285	0.1725	0.2071
Mean absolute error	0.3522	0.3641	0.3612
Root mean squared error	0.4383	0.4281	0.4257
Relative absolute error	84.8815 %	87.7586 %	87.0536 %
Root relative squared error	96.2335 %	93.9914 %	93.4738 %
Coverage of cases (0.95 level)	99.1386 %	100 %	100%
Mean rel. region size (0.95 level)	94.5506 %	99.6067 %	100%
Total Number of Instances	2670	2670	2670
Time taken to build model	3.81 seconds	0.16 seconds	0 Seconds

Table 3: Detailed Accuracy using KNN

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.333	0.128	0.519	0.333	0.406	0.238	0.702	0.485	-1
0.872	0.667	0.759	0.872	0.811	0.238	0.702	0.839	1

The table 1 is the KNN, SVM and GA processing instances values. In this table Correctly Classified Instances and Incorrectly Classified Instances are displayed with their values. The table 2 is the performance parameter table. It displays the different performance parameters. In this table KNN have good performance parameters. The table 3 is the detailed accuracy using KNN with TP,FP, Precision, Recall and other parameters.

Table 4: Detailed Accuracy using SVM

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.219	0.077	0.543	0.219	0.312	0.201	0.714	0.484	-1
0.923	0.781	0.740	0.923	0.821	0.201	0.714	0.850	1

Table 5: Detailed Accuracy using GA

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.222	0.054	0.63	0.222	0.328	0	0.722	0	-1
0.946	0.778	0.745	0.946	0.834	0	0.722	0	1

The table 4 and table 5 is the detailed accuracy of SVM and GA. It also displays the TP,FP, Precision, Recall and other parameters values . Based on these values the classifiers are evaluated that which one is the best classifier.

V. CONCLUSION

Data mining is additionally called information mining from information, information extraction, data/pattern analysis, information archeology, and information dredging. It involves the employment of refined information analysis tool to get antecedently unknown, valid pattern and relationships in massive information set. These tools will embrace applied math models, mathematical algorithmic rule and machine learning strategies. Therefore, data processing consists of quite assortment and managing information, it additionally includes analysis and prediction. Methoding is that the process of extracting patterns from information. Intrusion Detection System; becomes a vital a part of each pc or network system. Intrusion detection (ID) could be a mechanism that has security for each computers and networks. Feature choice and have reduction is very important space of analysis in intrusion direction system. The scale and attribute of intrusion file square measure terribly massive. Thanks to massive size of attribute the detection and classification mechanism of intrusion detection technique square measure compromised in terms of detection rate and alarm generation. During this work totally different classification algorithms square measure used and most results square measure calculated. During this it's enforced on alternative info and alternative tools. During this work GA is best as a result of here TP is 94.6 % instead of others.

REFERENCES

[1]. Muhamad Erza Aminanto et.al. “Wi-Fi Intrusion Detection Using Weighted-Feature Selection for Neural Networks Classifier” IWBIS 2017.
[2]. Megha Aggarwal et.al “ Performance Analysis Of Different Feature Selection Methods In Intrusion Detection” , International Journal Of Scientific & Technology Research Volume 2, Issue 6, June 2013.
[3]. Aditya Shrivastava et.al “A Novel Hybrid Feature Selection and Intrusion Detection Based On PCNN and Support Vector Machine” Aditya Shrivastava et al, Int.J.Computer Technology & Applications, Vol 4 (6), 922-927, IJCTA | Nov-Dec 2013.
[4]. Jayshri R. Patel et.al “Performance Evaluation of Decision Tree Classifiers for Ranked Features of Intrusion Detection” Journal of Data, Knowledge And Research In Data Technology, ISSN: 0975 – 6698| NOV 12 TO OCT 13.
[5]. Venkata Suneetha Takkellapati et.al “Network Intrusion Detection system based on Feature Selection and Triangle area Support Vector Machine” International Journal of Engineering Trends and Technology- Volume3Issue4- 2012.
[6]. Xing, Eric P., Michael I. Jordan, and Richard M. Karp. "Feature selection for high-dimensional genomic microarray data." In ICML, vol. 1, pp. 601-608. 2001.
[7]. John, George H., Ron Kohavi, and Karl Pfleger. "Irrelevant features and the subset selection problem." In Machine Learning Proceedings 1994, pp. 121-129. 1994.
[8]. Dash, Manoranjan, and Huan Liu. "Feature selection for classification." Intelligent data analysis 1, no. 3 (1997): 131-156.
[9]. Panda, Mrutyunjaya, and Manas Ranjan Patra. "Network intrusion detection using naive bayes." International journal of computer science and network security 7, no. 12 (2007): 258-263.
[10]. Nguyen, Hai Thanh, Katrin Franke, and Slobodan Petrovic. "Towards a generic feature-selection measure for intrusion detection." In Pattern Recognition (ICPR), 2010 20th International Conference on, pp. 1529-1532. IEEE, 2010.
[11]. Gong, Shangfu, Xingyu Gong, and Xiaoru Bi. "Feature selection method for network intrusion based on GQPSO attributes reduction." In Multimedia Technology (ICMT), 2011 International Conference on, pp. 6365-6368. IEEE, 2011.
[12]. Gaura, E.; Kemp, J.; Brusey, J. (2013), Leveraging knowledge from physiological data: On-body heat stress risk prediction with sensor networks. IEEE Trans. Biomed.Circuits System.
[13]. Huang, G.; Zhang, Y.; Cao, J.; Steyn, M.; Taraporewalla, K. (2013), Online mining abnormal period patterns from multiple medical sensor data streams.
[14]. L. Qu, F. Long, and H. Peng (2015), “3D registration of biological images and models: registration of microscopic images and its uses in segmentation and annotation,” IEEE Signal Processing Magazine, vol. 32, no. 1, pp. 70–77 22.
[15]. M. Attin, G. Feld, H. Lemus et al. (2015), “Electrocardiogram characteristics prior to in-hospital cardiac arrest,” Journal of Clinical Monitoring and Computing, vol. 29, no. 3, pp. 385–392