Improved Intrusion Detection System utilizing Feature Selection Method and Ensemble Learning Algorithms

Gogineni Krishna Chaitanya[1],

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, 522502, Andhra Pradesh, India.

Uppuluri Lakshmi Soundharya[2]

[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, 522502, Andhra Pradesh, India.

Abstract—

The basic inspiration that drives Interference Area Structures (IDS) is to recognize obstacles. This type of recognition design targets critical elements in common PC-based edges to ensure electronic confirmation. The IDS model is reliably quicker and accomplishes more precise unmistakable affirmation rates by choosing the primary related features from the informative information record. The comfort of reflections can be a critical development of any ID to choose the ideal subset of reflections that improve the matching of the course of action pattern to be faster and lessen the multifaceted nature of saving or reactivating the receptivity of the painting. During this article, we have proposed a methodology based on disconnecting the dataset from the information in different subsets for each round. At that point, we developed a segment assertion strategy using the procurement channel for each subset. The game plan of ideal highlights is made by putting together the summary of the courses of action acquired for each round. The results of direct tests in the NSL-KDD educational file show that the proposed methodology to incorporate decision with less reflections improves plot accuracy and reduces multifaceted nature. Additionally, a similar report on the reasonableness of the frame is drawn for choosing highlights using a variety of mounting techniques. To reinvigorate the overall spectacle, another movement appears using Random Forest and PART to initiate a topographic structure learning calculation. The outcomes show that the less unpredictable exactness is expanded utilizing the halfway likelihood rule.

Keywords-*Intrusion Detection Systems, NSL-KDD, Feature Selection, Supervised Learning, Classification.*

INTRODUCTION

Disposed of Sensor Organizations (WSN) implant little sensors or contraptions furnished with radio, processor, memory; battery as sensor hardware. The wide circulation of these vigorously molded sensors makes customary control conceivable. These little gadgets are restricted as far as processor speed, radio reach, and noteworthy memory assets. This nature of the asset fight makes application plan structures unambiguous to producers. However, inaccessible sensor networks are shaky, and help thusly sent is misled, expanding the disappointment of assaults. WSNs are also widely recognized in sensitive applications, for example in fire assertions [1], power transmission as an array [2], boundary [3], use of armed forces [4], fundamental frameworks (CI) [5] and reduced sensor networks, remote associations (submarine WSN) [6].

The lack of true security attempts may require sending different types of attacks under sabotage conditions. These types of assaults may adhere to the standard functioning of the WSN and may disprove the point of understanding. Therefore, security is one of the unimaginable affiliations you join. the lack of means pushes manufacturers to use the standard security of the local population, as cryptography and single-course occupations work with care. The unmistakable verification of the power outage is seen on the basis that the second line of protection worked with the safety of the local population. Logically when running WSN, the outlook, spawn space should be light, adaptable, and scarce. This work proposes such systems in the context of the affirmation of the impedance of peculiarities in WSN. In this type of environment it is essential to ensure the safety of the sensor network against the dangers introduced by online insurance. Unfortunately achieving this is a regular test cycle due to the number of WSN highlights, the first success is perhaps the most fundamental one: improper resource management, covering solid areas that are cryptographic; and their assignment in wild and disposed of conditions, where it is valuable for the adversary to really arrive at the methodologies of the sensor local area, for instance, by obviously noticing the cryptographic keys from memory.

Rapidly improving progress on the web makes PC security problematic. From now on, the understanding that is tainted, the extraction of information, even when the assessments of AI, go through a specific assessment in DI with the weight of improving the accuracy of the statement as a model not vulnerable to IDS. Regardless of disclosure limits, IDS also offers additional functionality, such as audit request only. IDS plans for remote affiliation recognition are now under consideration. Two or three courses of action are suggested during the overview.

This chronicle revolves around the creation of IDs for WSN. To develop a faster clash area system model with better disclosure rates, it is essential to select the highlights from the educational information file. The decision of learning collaboration parts during model organization shows an abatement in calculation speed and improves exactness. The primary goal of this article is to pick the most suitable observable focuses to be utilized to see animosity in a KDD NSL dataset, basically comparable to the WEKA gadget utilized for the evaluation. Different execution estimations are utilized to assess the introduction of every classifier, for instance, Accuracy, Revision, Measure F, False Positive Rate, Overall Accuracy (ACC), and ROC Curvature. NSL KDD informational index is a typical illuminating rundown for identifying anomalies, particularly impedance discovery. This dataset contains 41 highlights that depend on various types of affiliate traffic. Affiliate traffic is disconnected into two classes, one is the normal class and the other is known as the quirk class.

The overall blunder class will in general blackouts or assaults beginning from the relationship while amassing the participation traffic records. As opposed to these assaults, the NSL KDD dataset is also followed through on four essential social issues, like DoS and Test. A few deals are added to the root customer (U2R), exceptionally far (R2L). DoS Assault gives the detachment of urgent relationship to set up customers through the surge of attack bundles found in the library and furthermore in the affiliation's assets. Scenes of DoS opposition join Outback and Smurf. Moreover, gets, in spite of Neptune's assaults, are likewise events of such attacks. Taking into account the phenomenal degrees of threat experienced in the various sorts of DoS assaults related to PC spending, the record basically dealt with the DoS assaults, as illustrated in the 2014 report .A DoS assault is viewed as a basic issue for real heads who reestablish ties over the Internet. DoS assaults make pioneers incomprehensible to clients by restricting participation and extra structure assets.

Notwithstanding how your participation security specialists have initiated endless audits to squash strains over DoS assaults, DoS assaults proceed to develop and at last have a more fundamental antagonistic result.

The paper report as follows. Part 2 presents a graph of the impedance area, activity identified with the survey. District 3 represents the model proposed by IDS and Section 4 is the research on the exploratory results obtained. Finally, section 5 gives the reasons.

Related work

The monitoring framework uses modernized classifiers or thought assessments to visualize standard or fantastic behaviors and generate models to help organize new traffic. Building an ideal AI-based territorial structure requires the exam to analyze the responsiveness of a single AI assessment or multiple calculations for all four rule assault classes rather than a solo assault request. We will suggest some of the assessments and procedures used by the specialists in this repository. Likewise, we will strive to focus on the scopes that used NSL-KDD to debunk the test results.

Hota and Shrivas, 2014 , proposed a model that utilized particular article attestation methods to take out irrelevant features in the dataset and make a much more brilliant and convincing classifier. The edges utilized that are related to the classifier are data acquire, affiliation, hit, and symmetric shortcoming. His testing work has been disconnected in two sections: the first is the production of a multiclass classifier that depends on various decision tree methods. At that point, the procedure to pick the improvement in the best acquired model is applied, which was here C4.5. Your bogus appraisal ended up using the WEKA gadget. The results showed that the C4.5 with Info Gain would be sharp with the results and achieve a stunning 99.68% exactness with just 17 highlights. Regardless, in view of the use of 11 highlights, in any event, weakness was envisioned with 99.64% exactness

Yin, et al., 2017 [16], examined the best approach to show IDS dependent on a meaningful learning approach using spasmodic neural affiliations (RNN-IDS) due to its ability to sacrifice better information descriptions and improve models. Recently they have managed the dataset utilizing the numbering technique thinking about how the expense of RNN-IDS data ought to be a mathematical affiliation. The outcomes showed that RNN-IDS has astonishing exactness and ID rate with a communicated pace of supposed bogus positives and customary adjustment draws near.

Highlighting choice as an essential element of any IDS can help make model organization technique less variable and faster, while ensuring or improving unmatched design execution. Shahbaz et al. [17] recommended an economically advantageous computation for the consolation of remarkable centers seen as a relationship between the direct name of the class and a subset of properties to solve in a choice the problem of dimensionality reduction and surprising representation. The results revealed that the proposed model exhibits incredibly irrelevant planning times, saving precision with precision. In addition, some collateral techniques are attempted with floating classifiers with respect to the disclosure rate. The results of the association reveal that the J48 classifier works excellently with the proposed thought confirmation system.

In addition, the overview proposed another secret IDS that reflects on the decrease in the number of highlights. Most importantly, manufacturers are heavily involved in affiliation and information retrieval. Hence, the decrease in inclusion occurs when situations acquired from both obtaining information and subsequently affiliation through another treatment technique come together, perceiving what is significant and what is not huge. These reduced highlights are inserted into a neural information relationship to plan and test the KDD99 dataset. The strategy uses a proactive organization to extract excessive and irrelevant information from the dataset to update resource utilization and reduce time of a multifaceted nature. In fact, the transparency of the diminished part diagram contrasts better with the drawing without highlighting. As the segment that improves the problems of choosing exceptional assault representations shows, SVM drop classifier specialists used to represent phenomenal assault groups and BN classifiers to package blueprints. The results of the Choice Strategy Fuse (CGFR) test showed that CGFR elective highlighting is surprising and careful in the IDS.

From this inspiration, we are attempting to diagram which of the format computations we have picked will give the least complex outcomes in the wake of picking the features that incorporate a solid relationship inside the state dataset. During this work, experts will endeavor to direct an evaluation to isolate and find commonplace and strange practices.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 8, Issue 2, 2019

3. METHODOLOGY

The fundamental objective of the test is to display a design that isolates the breaks with the least features inside the instructive record. The unarmed past has shown that solitary a subset of these striking focuses is identified with recognizable proof. Subsequently, the conviction is to decrease the dimensionality of the information assortment to pass on an unmatched classifier for a sensible time frame. The proposed approach contains 4 key stages: the fundamental stage is picking the related features for each round utilizing the thing choice procedure. Around that point, the uprooted reflections meet to accomplish the ideal reflection plane for all rounds. The last floor of the wicks has proceeded onward to the assortment stage. At last, the model is tried utilizing a test dataset. The development of the proposed hypothesis is appeared in Fig. 1.

A. Assurance of the overall capacities with regards to every association

Albeit the affiliation breaking framework handles a tremendous measure of crude information, part assurance is turning into a focal improvement in building that office. Highlight confirmation is identified with various frameworks and procedures that will in general wipe out inconsequential and outrageous highlights. The dimensionality of the data document greatly affects the intricacy of the model, bringing about low portrayal precision and high calculation time and expenses. The objective of these procedures is additionally to pick the ideal highlights that will improve the presentation of the model. There are two general classes of techniques for the decision of thought, channel procedures and backing methodologies [23]. In the channel appraises, a

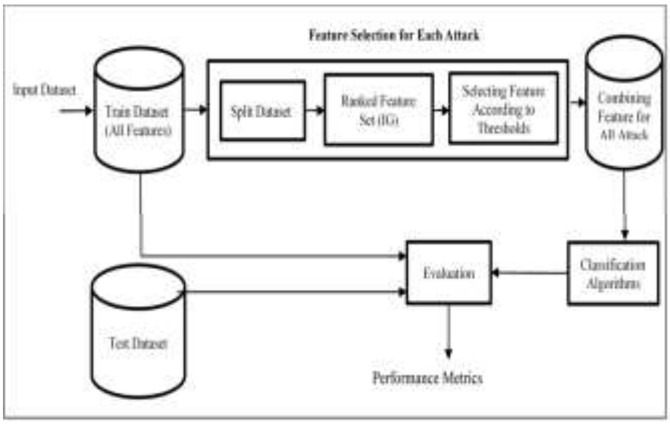

Feature Selection for Each Attack
Input Dataset
Train Dataset (All Features)
Split Dataset
Ranked Feature Set (IG)
Selecting Feature According to Thresholds
Combining Feature for All Attack
Evaluation
Classification Algorithms
Test Dataset
Performance Metrics

Fig 1. *Framework of IDS*

self-administering measure is used, which are utilized to overview the relationship of a great deal of highlights, while covering assessments utilization of one of learning calculations to make the evaluation of the segment's worth. In this assessment, Information Gain (IG) will be utilized to pick the subset of related highlights. IG is reliably cost less and quicker than the covering philosophies. Data secure is dealt with for every individual quality in the arranging dataset identified with one class. In the event that the arranged respect is high that recommends a segment is astoundingly certain this class. Something other than what's expected if the worth isn't really the destined edge, it will be executed from the part space. To get a prevalent edge respect, the dispersing of the IG respects is researched and endeavored with various edge respects on the status dataset.

The IG of an element t, by and large classes is known by condition (1).

$$IG(t) = -\sum_{i=1}^{m} p(c_i)\log p(c_i) + p(t)\sum_{i=1}^{m} p(c_i\backslash t)\log p(c_i\backslash t) + p(\bar{t})\sum_{i=1}^{m} p(c_i\backslash \bar{t})\log p(c_i\backslash \bar{t})$$

- *c*i addresses (I) classification.
- P(*c*i): likelihood that an irregular example archive has a place with class *c*i.
- P(t) and P($\bar{t}$) probability of the occasion of the segment w in a self-assertively picked document.
- P(*c*i|t): probability that a self-assertively picked report has a spot with class *c*i if document has the segment w.
- m is the amount of classes.

The confirmation highlights stage for each assault is separated into three standard undertakings as follows:

Step1: The plan dataset is allocated into 22 datasets. Each dataset report contains the records of one assault records converged with the ordinary records. On the off chance that the entire dataset is utilized without isolating, the affirmation highlights procedure will be lopsided to the most standard assaults. Hence, this development is chief to acquire more cautious outcomes.

Step2: Each record by then is utilized as a promise to IG methodology to pick the most fitting highlights of that assault. For instance, the public position usable assault has the related highlights arranged as shown in Table 1.

Step3: An arranged consolidate summary is made, and as several edges, various highlights are avoided. From the outline in Table I, it will overall be seen that the most fitting highlights for spy assault are highlights 38 and 39, in the event that we take the limit similar to 0.003. Along these lines, we can take the best two highlights and dispose of the others.

TABLE SPY RANKED RELATED FEATURES

Value Ranked	Number Feature	Name of Features
0.0012616	18	num_shells
0.0011182	15	su_attempted
0.0008254	19	num_access_files
0.0001007	2	protocol_type
0.004027	**38**	**dst_host_serror_rate**
0.0036055	**39**	**dst_host_srv_serror_rate**
0.001815	3	Service

B. Consolidate different feature sets for all attacks

In this progression, a consolidated summary of highlights is created for all rounds from the acquired subsets. For a couple of rounds, the absolute best position is chosen from the three essential highlights. Except for another assault arrangement, similar to the ground assault, one element was taken, as its range is sufficient at 1, while the positions for the different highlights were exceptionally low. which means that this component can completely separate this assault.

C. Arrangement of the training dataset

An authoritative combined subset is utilized as a commitment to the portrayal stage. The sequelae of three exceptional classifiers are accepted to shape a comparable examination. These classifiers are J48, Random-Forest (RF) and Partial Decision List (PART). Subsequent to directing the tests, the aftereffects of the two less troublesome groupings are chosen. The subsequent advance is to utilize the endeavor input method to help the introduction of the model.

The classifier □□J48: C4.5 (J48) is a calculation made by Ross Quinlan that for the most part makes a decision tree. This assessment gets notable in information depiction and the board. The extension extent technique is used during this calculation as the standard for separating the dataset. Minimalization strategies are applied to the information acquired using a "split data" measure.

Random Forest - Rumored to take care of business settled on thinking strategy that combines decision tree and tidy up procedures. Officer administration responsibilities watching out for the highlights are assembled heedlessly to make the trees made. The boondocks age cycle amasses various trees with controlled differences. The choice projecting a polling form party or weighted Democrats every now and again pick the accompanying supposition.

□□Partial Decision List (PART): PART is a fractional choice tree calculation stayed aware of an elective layout, joining the advantages of the C4.5 and PIPPER classifier. A pruned choice tree is made for all current cases, for the leaf turn by making a standard similar with the key thought, in this way, all things considered discarding the tree and proceeding.

□□Set Classifier: A set classifier is contained uniting a couple of fragile AI assessments (known as weak understudies) to improve depiction execution. The blend of weak students is reliably kept up in various procedures, for instance, the lion's offer vote, push or ending.

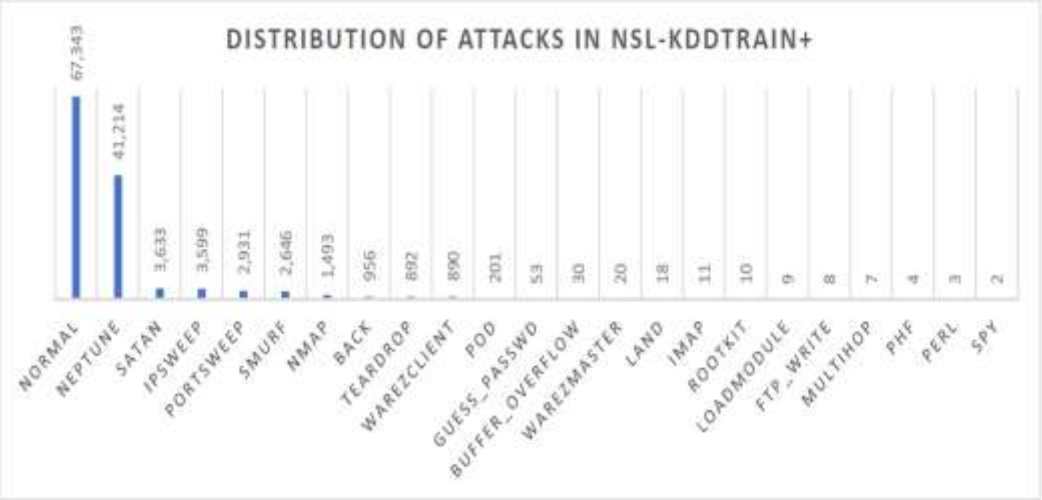


Fig 2. Distribution of Attacks in NSL-KDDTrain+

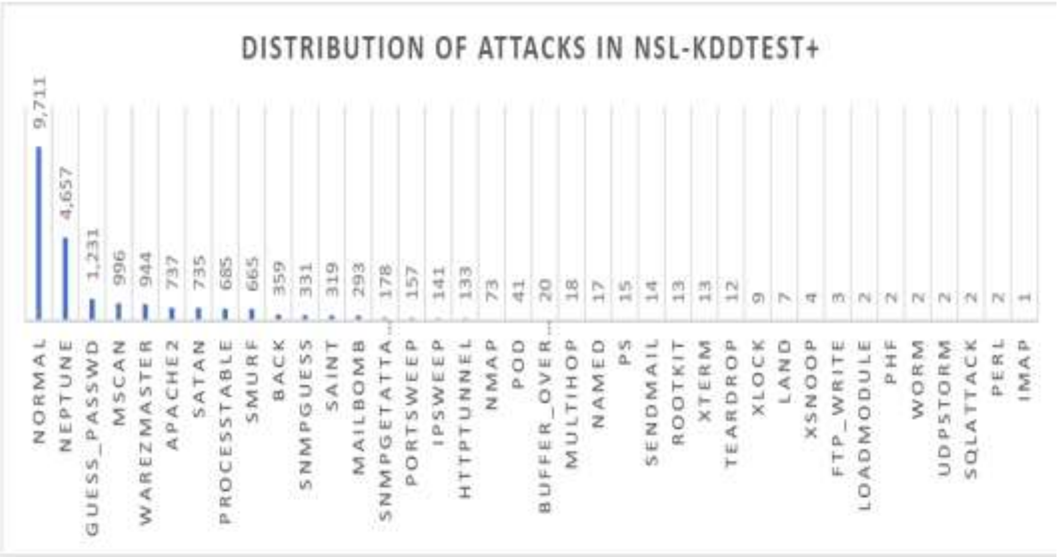


Fig 3: Distribution of attacks in NSL -KDD Test

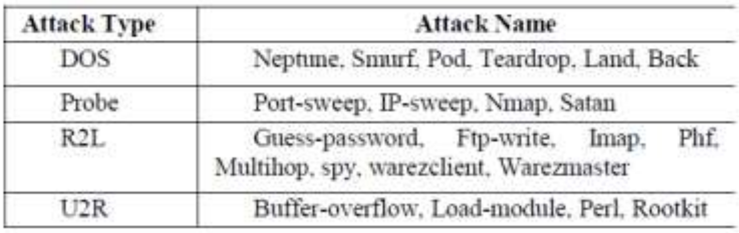
D. Model Testing

In this stage, a KDD-Test dataset is utilized to quantify the model that was made by the democratic gathering procedure. The test dataset log isn't equivalent to the arranging dataset and has an extra number of assaults. by then the presentation.Model assessment is led utilizing a few gauges like precision and district inside the ROC.

4. RESULTS AND ANALYSIS

All examinations are controlled with NSL-KDD datasets. NSL-KDD could be a refined interpretation of the KDD'99 dataset. That deals with certain typical issues inside the essential KDD dataset. Drawn-out records inside the prep set are completely disposed of together with classifiers to produce reasonable results. there is no copy information inside the general test suite. In this way, the uneven influence on the understudy show has liberally reduced. Each relationship during this dataset contains 41 highlights. Specialists during this work perform tests using information from KDDTrain and KDDTest. the various attacks are recorded

Table II. The dispersal of attacks in the NSL-KDDTrain + and NSL-KDDTest + logs is showed up in Figures 2 and 3.



Attack Type	Attack Name
DOS	Neptune, Smurf, Pod, Teardrop, Land, Back
Probe	Port-sweep, IP-sweep, Nmap, Satan
R2L	Guess-password, Ftp-write, Imap, Phf, Multihop, spy, warezclient, Warezmaster
U2R	Buffer-overflow, Load-module, Perl, Rootkit

B. Evaluation Metrics

The display evaluation of the proposed model, used assorted show estimations, for instance, precision (condition 2), audit (condition 3), F-measure (condition 4), certifiable negative rate, fake positive rate and as a rule precision (ACC) (condition 5) that known as adequately requested models (CC). Additionally, presented Received Operating Characteristics (ROC) of the system. The ROC twist is figured by drawing the association between obvious positive rate and fake positive rate in y-center point and xaxis, separately.

$$Accuracy = \frac{Number\ of\ Correct\ Classified\ Connections}{Number\ of\ Connections} \times 100\%$$

C. Results Analysis

In the wake of making various tests on the joined summary. The ideal number of merged features is comparable to 28 features.In IV Table, the accuracy and explicit evaluation appraisals with two property manners are seen wandered from utilizing the full scale dataset with 41 predictable qualities with the PART classifier with cross help of two test options and NSL-KDD Test . +. As noted, clearly precision. The introduction of the proposed structure considered utilizing the get breeze through evaluation and the major dataset. The outcome shows that amazing exactness is developed with (99.7984%) while utilizing a ton of 19 parts with cross-pass testing, while at the same time utilizing 28 features, the precision is (86.66%) while utilizing the NSL-KDD Test + dataset. Then again, the outcomes of the association between's the introduction of three estimations as per the proposed system and both the CV and the tests are introduced in Table V. As an examination, we utilize various appraisals from exceptional classifiers. These classifiers are J48, Random-Forest (RF) and Partial Decision List (PART). the best full scale exactness of the test with (86.66%) is acquired by deciding PART, yet the best by and large accuracy got by CV with (99.78%) utilizing RF. Figure 4 shows an examination of the portrayal estimations to the degree precision with see decision get breeze through and NSL-KDD + appraisal. Strong with these outcomes, the two most clear classifiers (PART and RF) are picked to control the calculation of the vote based assembling.

VI Table shows the conversation of utilizing a vote learning evaluation for Random Forest and PART to improve the got accuracy for intrude on demand settings. It has been seen that when the Random Forest and PART classifiers are utilized with different mix procedures, the accuracy of the model is improved. Table VI additionally shows that CV exactness is undefined when all of the three rules are utilized. Regardless, while utilizing the test dataset gave, amazing direct is seen for the entirety of the three models. the most easy precision is gotten when what likelihood rule is utilized. At long last, the world under the ROC turns, as demonstrated in Fig. 5, is agreed to every malevolence class inside the dataset with affirmed cross-endorsing and NSL-KDD testing. The outcomes also show that ROC checks for DoS and test assaults are basically something practically the same between the two test choices, yet the qualities change with R2L and U2R assaults.

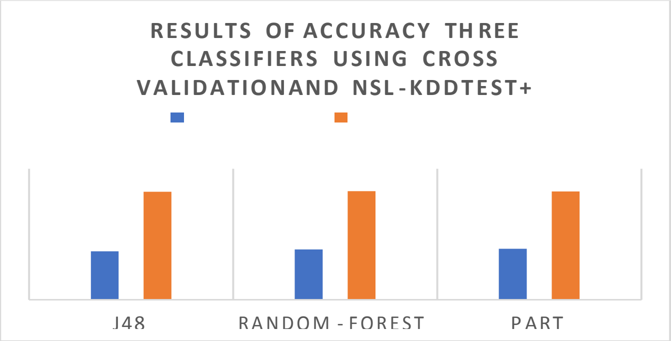


Fig 4. *Accuracy Results of Three Classifier*

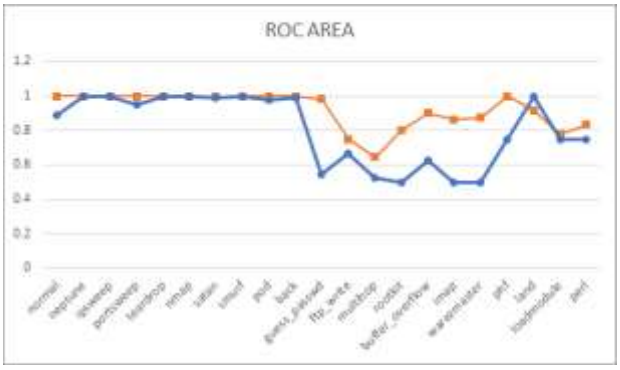
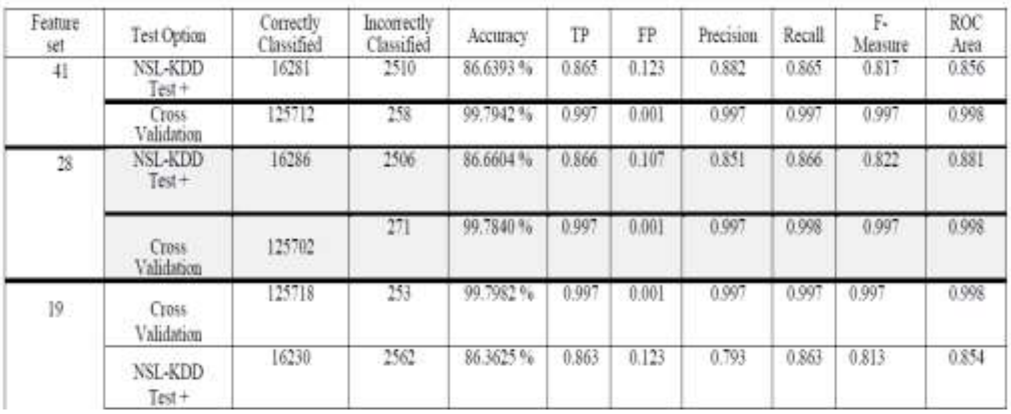


Fig 5. *Final ROC Area for each Class for CV and NSL-KDDTest+*

5. **CONCLUSION AND FUTURE WORK**

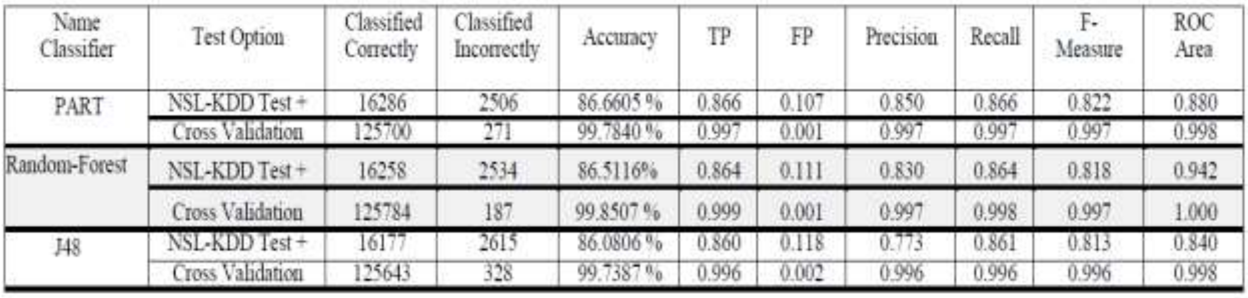
IDS is used to get PC-based systems against a large number of cyber attacks. Security towards the initial stage of the AI approach has been shown to improve supervision execution. In the research, we proposed a feature selection technique using information capture systems that were solved for each round in the NSL-KDD dataset to recognize the ideal summary of capabilities for each entered round and select these features. as shown by the specific cut-off points. It then joins the segment summary for all rounds. The test result shows that the most surprising precision was achieved using the Random Forest and PART classifiers with mixing methods, in particular the component probability rule.

IV TABLE. RESULTS WITH DIFFERENT NUMBER OF FEATURES USING PART



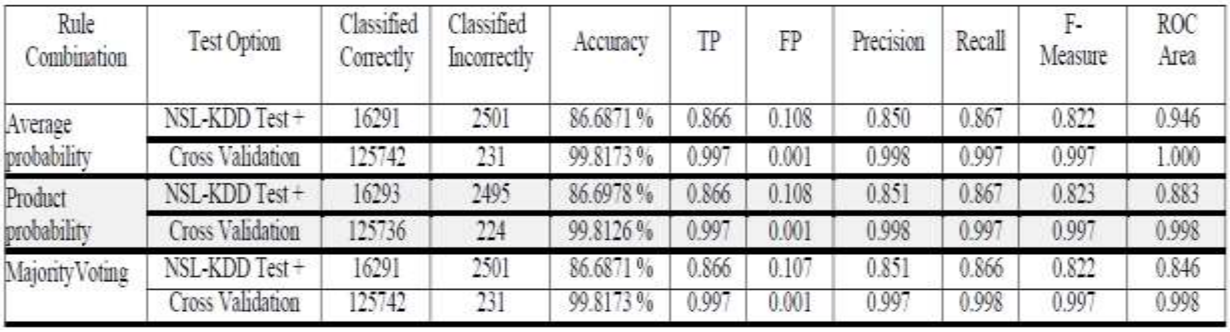
Feature set	Test Option	Correctly Classified	Incorrectly Classified	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
41	NSL-KDD Test+	16281	2510	86.6393 %	0.865	0.123	0.882	0.865	0.817	0.856
	Cross Validation	125712	258	99.7942 %	0.997	0.001	0.997	0.997	0.997	0.998
28	NSL-KDD Test+	16286	2506	86.6604 %	0.866	0.107	0.851	0.866	0.822	0.881
	Cross Validation	125702	271	99.7840 %	0.997	0.001	0.997	0.998	0.997	0.998
19	Cross Validation	125718	253	99.7982 %	0.997	0.001	0.997	0.997	0.997	0.998
	NSL-KDD Test+	16230	2562	86.3625 %	0.863	0.123	0.793	0.863	0.813	0.854

V TABLE TEST RESULTS. AND CROSS-VALIDATION OF THREE CLASSIFIERS



Name Classifier	Test Option	Classified Correctly	Classified Incorrectly	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
PART	NSL-KDD Test +	16286	2506	86.6605 %	0.866	0.107	0.850	0.866	0.822	0.880
	Cross Validation	125700	271	99.7840 %	0.997	0.001	0.997	0.997	0.997	0.998
Random-Forest	NSL-KDD Test +	16258	2534	86.5116%	0.864	0.111	0.830	0.864	0.818	0.942
	Cross Validation	125784	187	99.8507 %	0.999	0.001	0.997	0.998	0.997	1.000
J48	NSL-KDD Test +	16177	2615	86.0806 %	0.860	0.118	0.773	0.861	0.813	0.840
	Cross Validation	125643	328	99.7387 %	0.996	0.002	0.996	0.996	0.996	0.998

VI. TABLE TEST RESULTS AND CROSS-VALIDATION USING VOTE METHOD WITH (RF+PART)



Rule Combination	Test Option	Classified Correctly	Classified Incorrectly	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
Average probability	NSL-KDD Test +	16291	2501	86.6871 %	0.866	0.108	0.850	0.867	0.822	0.946
	Cross Validation	125742	231	99.8173 %	0.997	0.001	0.998	0.997	0.997	1.000
Product probability	NSL-KDD Test +	16293	2495	86.6978 %	0.866	0.108	0.851	0.867	0.823	0.883
	Cross Validation	125736	224	99.8126 %	0.997	0.001	0.998	0.997	0.997	0.998
MajorityVoting	NSL-KDD Test +	16291	2501	86.6871 %	0.866	0.107	0.851	0.866	0.822	0.846
	Cross Validation	125742	231	99.8173 %	0.997	0.001	0.997	0.998	0.997	0.998

REFERENCES

[1] P. D ́ıaz-Ram ́ırez, A., Tafoya, L.A., Atempa, J.A., Mej ́ıa-Alvarez, "Wireless sensor networks and fusion information methods for forest fire detection," Procedia Technol. 3, pp. 69–79, 2012.

[2] A. Isaac, S., Hancke, G., Madhoo, H., Khatri, "A survey of wireless sensor network applications from a power utility's distribution perspective," AFRICON 2001, pp. 1–5, 2011.

[3] B. . Mao, G., Fidan, B., Anderson, "Wireless sensor network localization techniques. Computer Networks," vol. 10, no. 51, pp. 2529–2553, 2007.

[4] V. Durisic, M., Tafa, Z., Dimic, G., Milutinovic, "A survey of military applications of wireless sensor networks," in 2012 Mediterranean Conference on Embedded Computing, MECO, 2012, pp. 196–199.

[5] L. Afzaal, M., Di Sarno, C., Coppolino, L., D'Antonio, S., Romano, "A resilient architecture for forensic storage of events in critical infrastructures.," in 2012 IEEE 14th International Symposium on High-Assurance Systems Engineering, HASE,

2012, pp. 48–55.

[6] D. Wahid, A., Kim, “Connectivity-based routing protocol for underwater wireless sensor networks,” in 2012 International Conference on ICT Convergence, ICTC, 2012, pp. 589–590.

[7] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

[8] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009., 2009, pp. 1–6.

[9] P. Institute, “2014 Global report on the cost of cyber crime,” 2014.

[10] H. S. Hota and A. K. Shrivas, “Decision Tree Techniques Applied on NSL-KDD Data and Its Comparison with Various Feature Selection Techniques,” in Advanced Computing, Networking and Informatics- Volume 1: Advanced Computing and Informatics Proceedings of the Second International Conference on Advanced Computing, Networking and Informatics (ICACNI-2014), 2014, pp. 205–211.

[11] D. H. Deshmukh, T. Ghorpade, and P. Padiya, “Intrusion detection system by improved preprocessing methods and Na #x00EF;ve Bayes classifier using NSL-KDD 99 Dataset,” in 2014 International Conference on Electronics and Communication Systems (ICECS), 2014, pp. 1–7.

[12] I. M. Y. Noureldien A. Noureldien, “Accuracy of Machine Learning Algorithms in Detecting DoS Attacks Types,” Sci. Technol., vol. 6,no. 4, pp. 89–92, 2016.

[13] M. A. Jabbar and S. Samreen, “Intelligent network intrusion detection using alternating decision trees,” in 2016 International Conference on Circuits, Controls, Communications and Computing (I4C), 2016, pp. 1–6.

[14] N. Paulauskas and J. Auskalnis, “Analysis of data pre-processing influence on intrusion detection using NSL-KDD dataset,” in 2017 Open Conference of Electrical, Electronic and Information Sciences (eStream), 2017, pp. 1–5.

[15] H. Wang, J. Gu, and S. Wang, “An effective intrusion detection framework based on SVM with feature augmentation,” Knowledge- Based Syst., vol. 136, no. Supplement C, pp. 130–139, 2017.