# Cuckoo Search Algorithm and BF Tree Used for Anomaly Detection in Data Mining

Shubhi Kulshrestha, Ankur Goyal
*M.Tech Scholar, Assistant Professor*
*YIT, jaipur, India*

***Abstract*** *— Anomaly detection is the new research topic of this new generation of researcher's today. Defect detection is a domain, that is, the key towards impending data mining. The term mining model refers towards methods & algo's that allow data towards be extracted & analyzed so that the rules & patterns that characterize the data can be found. Towards learn more about hidden structures & connections, the technology of data mining(DM)may be practical towards any type of data. In the current world, a lot of data is transported from one place towards another. Data that is transmitted or stored is subject towards attack. Although there are several techniques or applications available towards secure data, there are still ambiguities. It helps towards analyze resulting data & determine several types of attack data mining techniques that are open towards occurrence. Anomaly detection is technique of DMto detect astonishing or unanticipated behavior concealed in data that increases probabilities of penetrating or attacking. This learning proposes a model aimed at structure a NW penetration detection system by a machine learning (ML) algo & the system is primarily a fault-based penetration detection. In this paper, the feature selection Cuckoo search algo is performed, & the classification is mainly performed by the BF tree. It shows that the proposed work produces better results & the detection error attributes are more accurate.*

**Keywords** — *Data mining, Intrusion Detection System, Anomaly IDS, K Means, SMO, Genetic Algorithm, Cuckoo Search Algorithm and BF tree.*

## I. INTRODUCTION

A huge amount of information is available in the information industry; this information is not valuable until it is converted in towards valuable data. From concealed information, we may gain significant and valuable information. As information can lead us towards a rich source aimed at information discovery. In the mining procedure, we treat formerly stored information aimed at additional forecasting & forecasting. In the age of the data society, network-based computer systems have become a fundamental role, so they have develop target of intruders & criminals. Firewalls, client verification, data security, and data encryption techniques that prevent infiltration fail towards fully protect the networks & systems behavior from growing attacks & malware. Infiltration detection systems (IDS) are considered towards defend computers & networks after several cyber-attacks & viruses. IDS is a system that monitors NW or system operations aimed at malicious activity & yields reports proceeding a management station. [1] Since data mining is a key application region of data mining based on algo detection, it aims towards solve problems in analyzing large volumes of information. IDSs construct effective clustering and classification models towards discriminate usual behavior after abnormal behavior by data mining (DM) techniques. This study lays the groundwork aimed at research & exploration & implements the penetration detection model system built proceeding DM technology. Infiltration is activities that violate system's security procedure. Infiltration detection is procedure utilized towards detect infiltration. Network security has become a key issue into modern years, as computer network (CN) is encompassing dramatically. Data systems and NWs are issue towards electronic attack & risk of penetration is so high. Intrusion Detection System (IDS) isscheme aimed at identifying intrusion & reporting towards Authority or NW Administration. Data mining (DM) methods have been effectively useful into several areas including NW management, education, biology, marketing, industrial, and procedurecontrolleras well as fraud detection. DM aimed at IDS is a method that may be utilized mostly towards detect anonymous attacks and raise alarms after safety breaches are identified. [1].

Computing as well as networking technologies allow humans to enter net along with the net. In connecting, the amount of Internet users is collective rapidly, high-recognition of global-wide influences has given rise to some safetydifficulties. Traditional strategies similaroperatorverification, information encryption as well as firewall are utilized extensively to protect PC security. When a PIN is tampered with, the right to enter unauthorized access to consumer certification cannot be stopped; Firewalls cannot save intruders or attackers from other malicious intentions. In the case of a firewall, Intrusion Detecting Systems (IDS) [2], which employs special analytical method (s) to stumble on the attacks, two identities of intrusion detection (ID) Models of fashion misuse and discrepancy are [3]. With the predetermined pattern, it will shape the unknown pattern after which it does not forget that these miles are regular or odd.

However, the latter also knows about the behaviors that are prepared with everyday styles. The specific advantages and disadvantages of both models can be used to detect misuse as a new naval attack, nonetheless it cannot perceive a novel occurrence. On alternative pallets, there is low accuracy and high false alarm for discrepancy detection for anomaly-based methodology usagesarithmetical strategies to analyze packages, as well as it may detect novel occurrences.

Many algorithms in DM and device have been used primarily in mastering. Although K in this letter means SMO and Genetic Algorithm used to discuss the IDS. Thus, there is a small evaluation of both algorithms, & other usuallyutilized algo related to discussion of IDS.

## II. ANOMALY DETECTION (AD)

ID is a big problematic that has been deliberate in specific study domain names as well as application arenas. Different strategies of AD were definitely tilted for the assured software sectors, at the same time as the others are normal. AD is defined due to the difficulty of the defining statistics styles, which do not agree with expected behavior. These designs are usually shown inconsistencies, outliers, disagreeing comments, exceptions, aberrations, surprises, strange, or contaminated substances and many others. In exceptional use areas Of these, the anomalies as well as outlays are 2 phrases that are commonly used to maximize the approach of AD; now and again. AD Special programs include credit playing cards, coverage or fraud detection for health care, ID for cyber security, detection of fault in important systems, and common usage in military monitoring for enemy games. Imposes. Importance of AD is due to this claim that the anomalies in the information decode important and frequently important, active facts in the inclusive category of the software domain. [4].

AD is defined as procedure of patterning in a dataset, whose conduct isn't always regular before predictable. This is the documentation of information factors, gadgets, comments or opportunities that aren't conformed to anticipated or recognized sample ofassumed organization. These inconsistencies arise actual difficult nonetheless can also identify large as well as big threats such as cyber intrusion or fraud. AD is used in further techniques of behavioral analysis and analysis, which help to detect, detect and study the predictions of these anomalies. AD is roughly speaking a process of DM as well as it is utilized to control pattern of discrepancies in a given statistics set as well as define details around their occurrence. This is appropriate in the domain containing schemediscovery, id, fault detection, machine health tracking as well asoccurrence detection system in device network. In scenario of deceptionas well as id, inconsistencies or thrilling genres are not unavoidably rare objects, nonetheless those are amazing tones of deeds. Such anomaly does not accept meaning of anomalies or outlars as rare events; many such advertising strategies no longer work in examples unless they were definitely combined or educated. . Therefore, in these cases, a cluster evaluation set of rules can be additionalappropriate for perceiving micro cluster patterns produced through those record factors. The important point related to AD approach is illustrated in Figure 1.

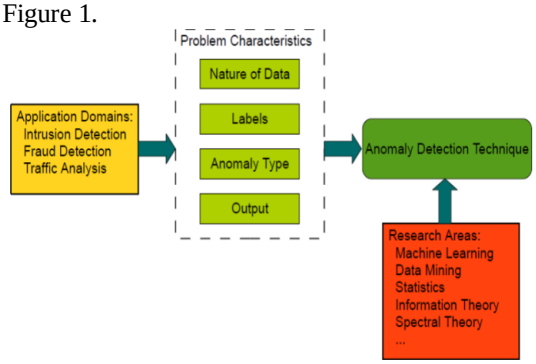


**Fig.1. Key components related with an anomaly detection method**

## III. ALGORITHMS USED IN DATA MINING

There are various algorithms used in DM and theses are discussed in this section such as:

### *A. K-Means Algorithm (KMA)*

Clustering is a DM approach that creates a costly cluster of gadgets. The clustering technique describes the training and holds the objects in each class, inside the classification errors, the gadget goes to predefined training. The K approach is entirely one of the algorithms based on clustering. Network Typing Elegance label is divided into 4 key commands, which can be DoS, Check, U2R as well as R2L. Through Fig 2 (A) Fig2 (D) demonstrations steps related to K-Means clustering method. After Fig2, the application of classification approach will show very final overall results. The purpose of using K-Means Clustering (KMC) is to generalize and split the information in time of attack and to create a group. KMC legislators, in each cluster, split the datasets recorded in the K-clusters conferring to pre-value recognized as B-factors in factor centroids (cluster centers), that is suggested value of numerical information limitedinsideeverycluster. In our case, we take exactly = three so that it is contained in three groups (C1, C2, C3). Meanwhile U2R as well as R2L attack patterns are clearly comparable to regular examples, an additional cluster is used for organization of U2R as well as R2L attacks. Back to Fig (B), every input will be deposited consuming square distance between entry information factors as well as centers where centriole will be deposited. Through the calculation of the suggestions standards of assumed input set aimed atevery cluster, novel centroids will be producedaimed atevery cluster as

shown in picture (C). Phase (B) as well as (C) in fig are repeated pending end consequence as shown in picture (d).

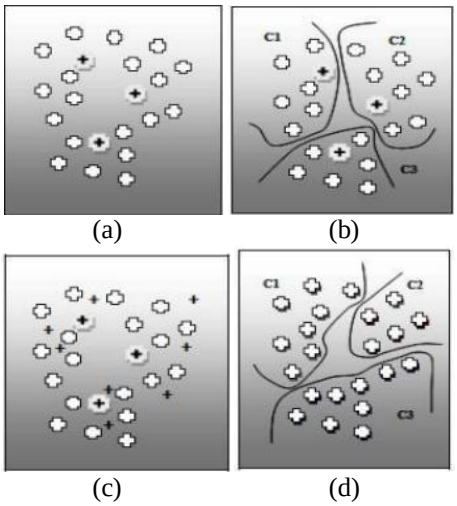


**Fig.2. K-MeansClustering**

The KMA mechanism such as shadows:

- Excellent preliminary centers of K clusters.
- Replication steps 2 via three till cluster membership steadies.
- Make a novel partition by way of transmission every information to its closest cluster centers.
- Calculatenovel clusters as centroids of clusters [5].

### *B. Sequential Minimal Optimization (SMO)*

SMO sets a specially set of rules to fix problems related to optimization, the SMO set of rules usages vectors in every phase of the answer, through usage of all Vector Systems (SVM) optimization It container be resolved, it may be solved. SMO, deprived of the use of a big scale matrix, Excellence SMO is the only way to solve the use of 2 Lagrange, choose the SMO2 Lagrange which are optimized at the same time and find or find top-rated values To update the updated cost which are used for SVM because the most modern optimum value. SMOs resolveproblematic of quadratic programming associated with the optimization of the analytic parameters, parameters have been optimized and the parameters of the system data are allowed according to the parameters, the income from SMO is now largely matrix from a large scale matrix Is not without use. The use of SVMs can be stored in memory of the education facts in education system, so this method may be run additional rapidly as well as computing time of SMO can be reduced. [6].

### *C. Genetic Algorithm (GA)*

GA is an evolutionary approach which solves each restricted as well asunimpeded optimization difficulties. It is based on natural choice which energiesorganicdevelopment. GA time also again adapts a populace of person solutions. At every step, GA picks samples at random from the cutting-edge population to be parents as well as uses them to produce changed offspring aimed at subsequent compeers. Over successive generations, populace "evolves" towards a top-rated solution. Here each individual within populace of GA signifies a chain of actions, which, if implemented to prototype of statistics set, creates every other prototype of the records set.

1) Fitness Function (FF): The FF or the prediction accuracy standards is used to evaluate the high-quality of the prototype which is generated in each generation.

2) Selection: Selection is the method of figuring out which people in the populace to apply for reproduction, and replacement for the subsequent generation.

3) Multi Point Crossover: This paper used multi factor crossover strategies to rearrange the extraordinary characteristics of an man or woman and create offspring out of mother and father inheriting the characteristics immediately.

4) Mutation: Mutation alternatives define how the genetic set of rules makes small random modifications within the individuals in the populace to create mutation offspring. Mutation generates genetic diversity and allows the genetic algorithm to go looking a broader space [7].

## IV. LITERATURE REVIEW

This paper [8] We suggest a combination of machine learning technique aimed at network penetration detection created on K-media clustering as well as sequential minimal optimization (SMO) organization. It presents a hybrid technique that helps detect false positive alarm rates as well as untrue negative alarm rates, improve detection rates, as well as detect zero-day invaders. NSL-KDD dataset was utilized in futuremethod. Organization was achieved consuming Sequential Minimal Optimization. After learning the specific technique, projectedmethod (K-medium + SMO) attained a positive detection rate of 94.48% as well as a false alarm rate of 1.2. %) As well as Accurateness (97.3695%).

In this paper [9], A combination of anomaly-based fully formatted ID technique is created on DT and KNN. To improve presentation of the futuremethod, a piece option method is used to extract customized records from NSL-KDD datasets. Experimental effects confirmed that projected method has attained positive identification rate of 99.7%, false alarm charge of 0.2% as well as ninety nine percent..

This paper [10] It has proved that excessive accuracy can be preserved while dropping counterfeit positivity consuming proposed version made from SVMs, Decision Trees (DT), and Na, ve Bayes (NB). Firstly, SVM is primarily based on a novel binary organization, which is brought to dataset to stipulate that example is an attack or normal traffic. 2nd, visitors of attack are routed through DT for class. 3rd, NB as well as DT will vote on severalunsystematic attacks.
In this paper [11], To produce network intrusion detection system (NIDS) produced two levels of mining algorithms and to decreaseuntrue alarm rate, NB algorithm in the first level used to classify unusual activity in the main four attack types from normal behavior. Is done for. In the second level the ID3 DT algorithm is used to classify four attack types in children (22) of attacks by normal behavior. To evaluate the presentation of two proposed algorithms by using Kdd99 dataset IDS and evaluation metric accuracy, precision, DR, F-measurement. New results prove that the proposal system lowered the 99% high identification rate (DR) and false positive (FP) for various types of network intrusions.

This study [12] A version is proposed for the creation of NID using the device that obtains knowledge of set of rules called DT. This machine mainly detects an anomaly based intrusion. The graded capabilities of the Dataset Alteration Identifiers (CCIDS) 2017 in this model are prearrangedconsuming label encoder. Some pleasant capabilities are chosen using recursive-feature-elimination (RFE). This information is then distributed in school education and testing information. Training facts are used to create a DT-Model in which each leaf shows conceivable result. Organization models are used to see data as malicious or benign by using school information. Calculatingaccurateness of classifier on a future record moderately than facts beyond is a dominant element. The accuracy of classifier found on check information is ninety nine. The accuracy of projected gadget designates that True-Positive-Rate (TPR) is 99.Nine% as well as false-positive-rate (FPR) is zero .1%. proposed model uses fixed day-to-day statistics for education records and examines the statistics associated to customary structures which have been modeled consuming KDD-CUP-99 notification usual. Apart from this, differentnew structures, it does not usagesome fact-mining tool similar Weka. This work is done as foundation of somenovel set of rules using the Dataset CCIDS 2017.

This paper [13] Provides a framework on SPARC's main technologies. We propose future trends of IDS in Spark using the NB classifier with the NSL-KDD dataset. The time it takes to make the model is 0.03 seconds. The accuracy of the detection of the attack was 99.777% with high identification rate and fewer false alarm rates. Our future work with this initial task is to combine association rules and other classifier models for better accuracy.

## V. PROPOSED MEHODOLOGY

Cuckoo Search Algo (CSA) created on levy flight conduct as well as issue parasites. CSA provides excellent consequences in controlled optimizations. Cuckoo natures survey a destructive limitation procedure. Females take control as well as untrained their inseminated eggs in cages of other birds. When host bird understands that it does not own egg, it either throws alien egg or refuses nest as well as builds another in its new location. Every egg in a nest signifies a possible result, as well as every cookie egg signifies a novel result. The idea here is discovery novel as well as improvedresolutions (Cuckoo) to replace old resolutions in cages. For simplicity's sake, let's assume that everycase has an egg. CSA may be protracted to additional complex situations as soon asevery nest comprisesmanifold eggs on behalf of a set of solutions. The CSA may be abridged by succeeding 3 main rules:

- C lays an egg in every Cuckoo, as well as putting it in a vaguely selected cage.
- The best nests that carry the best quality eggs (solutions) will move to the next iteration.
- A has a secureno. Obtainable host nests as well as host-birds may identify a distant egg bypossibility pa ϵ [0, 1]. host-bird may either leave foreign egg in cage before leave nest to make a novel nest in a novelplace. In recurrence, 3rd hypothesis may be that novel probabilistic solutions of n nests can be considered new arbitrary solutions instead of pa. In example of optimization problematic, advantage or fitness of aanswer is straightconnected to impartial function. In CSA, levy flights are determined by steps of a beak. Levy follows the flight to create novelexplanations for xi (t + 1) aimed atith cuckoo:

$$x_i(t+1) = x_i(t) + \alpha \oplus Lévy(\lambda)$$

Where α > 0 represents phase length that need be connected to degree of problematic worried. Product symbol ⊕ stipulates entry-wise increases. This broadside has measured a Lévy flight wherever step-lengths are scattered as per possibility dissemination assumed as:

$$Lévy\, u = t^{-\lambda}, \quad 1 < \lambda \leq 3$$

Which consumes an limitless alteration.Successivephases of cuckoo effectively

establish a random walk proceduresubmitting a power-law step-length delivery with a weighty tail. Cuckoo Search algo is obtainable in Figure3.

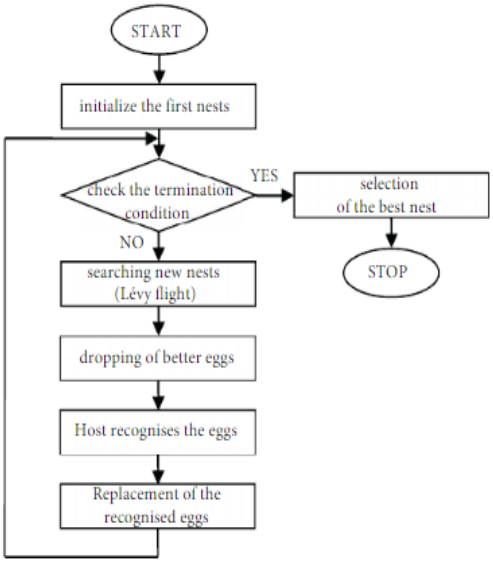


**Fig. 3 Block Diagram of Cuckoo Search Algo**

In Best-First (BF)
Best-First (BF) Decision in Tree Algo, Tree applies time through selecting node to decrease impurity of partitioning of all current nodes. In this algo, pollution can be intended by Gini index as well as data gain. The BF tree is made up of a divide-Cuckoo style parallel to normal depth 1st decision trees. Simple stage of making best 1st tree is assumed below.
.

- Excellent characteristic to place in Root node as well as build some branches aimed at this attribute created on certainstandards.
- Divide training series in subsets, one aimed at every branch extending since root node.
- N construction procedure spending all protuberances are cleaned or specified. Reached extension.

In the previous work, Genetic Algo has been used with ConsistencySebsetEvel to select the appropriate features from the dataset. Then K means used for the clustering of an attributes selected to decrease the time as well as complexity. Instead of this, it has not reduced much false positive value so this can be improved by the proposed work. In our proposed work, initially preprocessing of the dataset (NSL-KDD) performed by using Cuckoo Search Algo as well as ConsistencySebsetEvel. This process is generally used to select appropriate features as well as eliminate duplicate, incomplete as well as extra features by grouping it. After this step, it produces 18 attributes from the overall dataset. These attributes are taken for clustering for reducing its complexity. BF Tree is used to divide these attributes into clusters as it performs the greatest split in tree created on boosting algo. Final step is to realize organization of these clusters to generate the two sets of data such as normal as well as anomaly. The classification has been done by BF Tree to differentiate the normal attributes from anomalous to prevent the intrusion. The proposed algo is explained below to understand the proposed work.

Proposed Algorithm:

Step 1: Start
Step 2: Import NSL-KDD dataset
Step 3: Get training dataset
Step 4: Perform pre-processing over the dataset
Step 5: Apply Cuckoo Search Algorithm and ConsistencySebsetEvel for feature selection
Step 6: Obtain features and perform clustering using BF Tree
Step 7: Apply BF Tree for classification of the data into anomaly and normal data
Step 8: Get the final result
Step 9: Exit

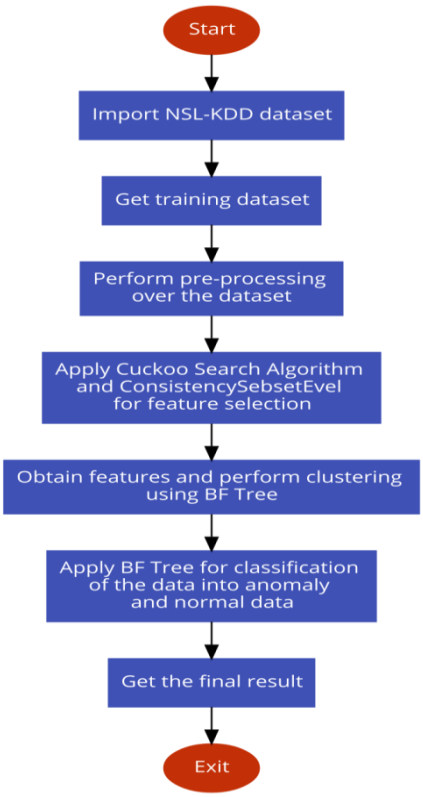


**Fig.4 Proposed Flowchart**

## VI. RESULT ANALYSIS

In this work, WEKA tool is utilized aimed at study of proposed work. Firstly explain the basic terms understand well as the concept as well as the

following procedures are demonstrated in the form of steps.

.

Confusion Matrix (CM):
Confusion is a moment of consequence of prediction in a classification problem. The no. of true as well as false predictions is summarized with values as well as each class is divided. This is important to Matrix of Misunderstanding. Confusion matrix demonstrationsconducts in which your organization model can be disordered once making predictions. It not only delivers insights about errors that make a classifier nonetheless rather importantly, that it does errors.

| | Class 1 Predicted | Class 2 Predicted |
|---|---|---|
| Class 1 Actual | TP | FN |
| Class 2 Actual | FP | TN |

Here,
• Class 1 : Positive
• Class 2 : Negative

Definition of the Terms:

- Positive (P): Observation is positive
- Negative (N): Observation is not positive
- True Positive (TP): Observation is positive, as well as is predicted to be positive.
- False Negative (FN): Observation is positive, but is predicted negative.
- True Negative (TN): Observation is negative, as well as is predicted to be negative.
- False Positive (FP): Observation is negative, but is predicted positive.

Detection Rate (DR):
DR is given by the relation:

$$DR = (TP / TP + FN) * 100$$

False positive rate (FPR):
FPR is assumed by relation:

$$FPR = FP / (TN + FP) * 100$$

Accuracy:
Accuracy is assumed by relative:
$$Accuracy = (TP + TN) / (TP + TN + FP + FN) * 100$$

### *A. Training dataset preparation*

Initially training dataset used to perform Cuckoo Search Algo + ConsistencySebsetEvel at 10-fold cross-validation test manner &produce topographies which are described below:

Instances: 25192
Attributes: 18
service
src_bytes
dst_bytes
land
num_failed_logins
logged_in
su_attempted
num_file_creations
num_outbound_cmds
is_host_login
is_guest_login
count
dst_host_count
dst_host_srv_count
dst_host_same_srv_rate
dst_host_diff_srv_rate
dst_host_srv_diff_host_rate
dst_host_srv_serror_rate

BF Tree is used which generate the tree of size: 151 and N0. of Leaf Nodes: 76. It takes time to figure model: 18.56 seconds.

=== Stratified cross-validation ===
=== Summary ===

| | |
|---|---|
| Correctly Classified Instances | 25081 99.5594% |
| Incorrectly Classified Instances | 111 0.4406% |
| Kappa statistic | 0.9911 |
| Mean absolute error | 0.0056 |
| Root mean squared error | 0.0639 |
| Relative absolute error | 1.1248 % |
| Root relative squared error | 12.8141 % |
| Total Number of Instances | 25192 |

=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---|---|---|---|---|---|---|---|---|
| 0.997 | 0.005 | 0.995 | 0.997 | 0.996 | 0.991 | 0.996 | 0.994 | normal |
| 0.995 | 0.003 | 0.996 | 0.995 | 0.995 | 0.991 | 0.996 | 0.995 | anomaly |
| 0.996 | 0.005 | 0.996 | 0.996 | 0.996 | 0.991 | 0.996 | 0.994 | Weighted Avg. |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 13402 | 47 | a = normal |
| 64 | 1679 | b = anomaly |

***B. Test Data Preparation***

=== Evaluation on training set ===

Time occupied to test classical on training information: 0.08 seconds

=== Summary ===

Correctly Classified Instances 25143 99.8055%
Incorrectly Classified Instances 49 0.1945 %
Kappa statistic 0.9961
Mean absolute error 0.0035
Root mean squared error 0.0421
Relative absolute error 0.7116 %
Root relative squared error 8.4356 %
Total Number of Instances 25192
=== Detailed Accuracy By Class ===

| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class |
|---|---|---|---|---|---|---|---|---|
| 0.998 | 0.002 | 0.998 | 0.998 | 0.998 | 0.996 | 0.999 | 0.999 | normal |
| 0.998 | 0.002 | 0.998 | 0.998 | 0.998 | 0.996 | 0.999 | 0.998 | anomaly |
| 0.998 | 0.002 | 0.998 | 0.998 | 0.998 | 0.996 | 0.999 | 0.999 | Weighted Avg. |

=== Confusion Matrix ===

| a | b | <-- classified as |
|---|---|---|
| 13424 | 25 | a = normal |
| 24 | 11719 | b = anomaly |

Tables 1 demonstrates the comparison among the base and propose and then the graphs are plotted to show the difference.

**Table 1: Comparison Between Base & Propose**

| Parameters | Base | Propose |
|---|---|---|
| FPR | 1.4% | 0.2% |
| DR | 95.8% | 99.8% |
| Accuracy | 95.8% | 99.8% |
| Time | 36s | 0.08s |
| Correctly classified instances | 97.3047% | 99.8055% |

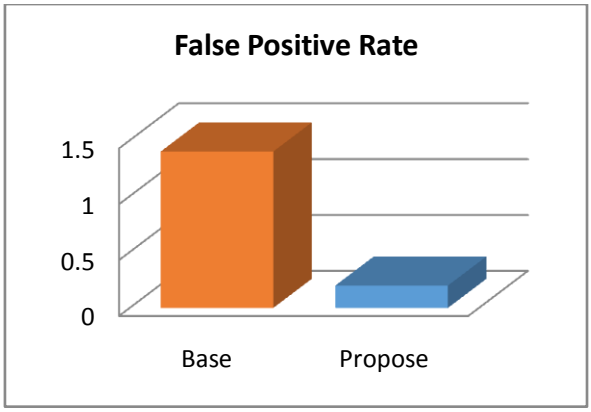


**Fig.5 Comparison of FPR between Base & Propose**

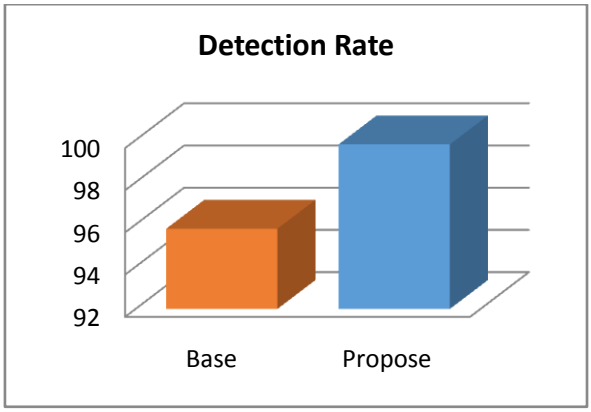


**Fig.6 Comparison of DR between Base & Propose**

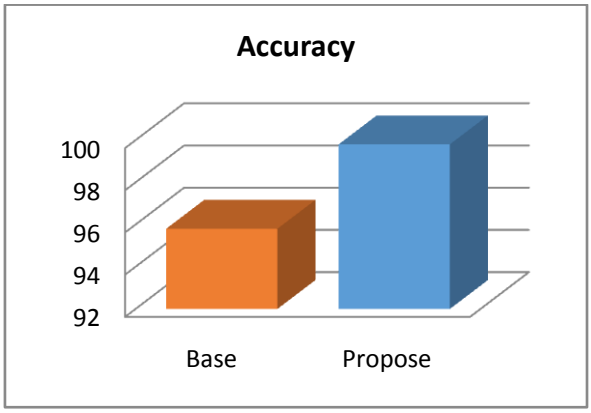


**Fig.7 Comparison of Accuracy between Base & Propose**

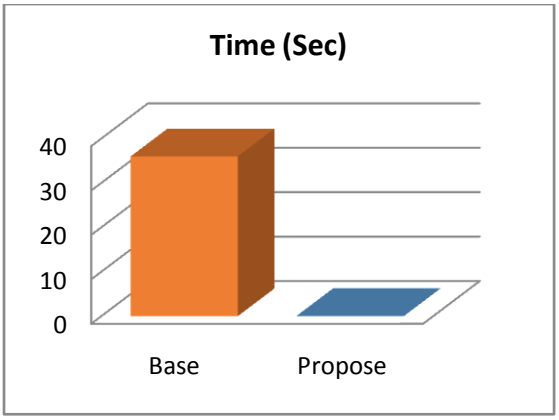


**Fig.8 Comparison of Time required to test model between Base & Propose**

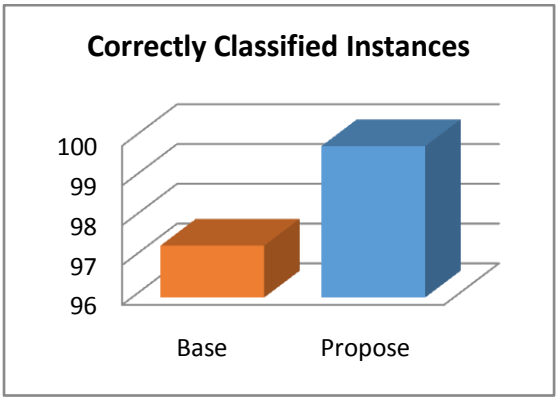


**Fig.9 Comparison of Correctly classified instances between Base & Propose**

## VI. CONCLUSION

The suggested model usages current data set aimed at exercise records and check facts in comparison to the conventional structures which have been modeled using KDD-CUP-ninety nine records set. Moreover, not like different schemes, it does not usagesome data-mining device similar Weka. Previously KDD-Cup99 Dataset became taken into consideration as the benchmark dataset for intrusion-detection but Nowadays, the network and the attack methods have changed drastically. The approach based on Classification of dataset is presented and discussed to develop an efficient intrusion detection model. The trialoutcomes demonstrate that the proposed approach can be used to develop an ID-Model having high detection rate, high accuracy, low FPR, less time required to test model and highly correctly classified instances .

## VII. REFERENCES

[1] J. Huysmans, B. Baesens, D. Martens, K. Denys And J. Vanthienen, New Trends in Data Mining, TijdschriftvoorEconomieen Management, Vol. L, 4, 2005: 1-14.

[2] Lee, W., Stolfo, S. J., and Mok, K. W. 1999. A data mining framework for building intrusion detection models. In Proceedings of IEEE Symposium on Security and Privacy. 120–132.

[3] Patcha, A. and Park, J.2007. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Comput. Netw. 51, 12 (August 2007), 3448-3470. DOI=http://dx.doi.org/10.1016/j.comnet.2007.02.001

[4] Varun Chandola, Arindam Banerjee and Vipin Kumar, Anomaly Detection: A Survey, ACM Computing Surveys, Vol. 41, No. 3, Article 15, 2009: 1-58.

[5] Sahil Sanjay Tanpure, JayrajJagtap, Gunjan D. Patel, ApashabiPathan, Zishan Raja, Intrusion Detection System in Data Mining using Hybrid Approach, International Journal of Computer Applications (0975 – 8887) National Conference on Advances in Computing, Communication and Networking (ACCNet – 2016).

[6] Dedy Kurniadi, Sam FarisaChaerulHaviana, Data Mining Sales Optimizations Using Sequential Minimal Optimization Algorithm, Journal of Telematics and Informatics (JTI) Vol.4, No.2, September 2016, pp. 39~44 ISSN: 2303-3703.

[7] Singha, T., &Goswami, S. (2017). Classification in data mining using POEMs/GA algorithm. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS).doi:10.1109/icecds.2017.8390187.

[8] Saad Mohamed Ali Mohamed Gadal, Rania A. Mokhtar, Anomaly Detection Approach using Hybrid Algorithm of Data Mining Technique, 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), Khartoum, Sudan.

[9] Foroushani, Z. A., & Li, Y. (2018). Intrusion Detection System by Using Hybrid Algorithm of Data Mining Technique. Proceedings of the 2018 7th International Conference on Software and Computer Applications - ICSCA 2018.doi:10.1145/3185089.3185114

[10] Goeschel, K. (2016). Reducing false positives in intrusion detection systems using data-mining techniques utilizing support vector machines, decision trees, and naive Bayes for off-line analysis. SoutheastCon 2016. doi:10.1109/secon.2016.7506774.

[11] Sarah M. Shareef, Soukaena H. Hashim, Intrusion Detection System Based on Data Mining Techniques to Reduce False Alarm Rate, Engineering and Technology Journal Vol. 36, Part B, No. 2, 2018.

[12] Riyazahmed A. Jamadar, Network Intrusion Detection System Using Machine Learning, Indian Journal of Science and Technology, Vol 11(48), DOI: 10.17485/ijst/2018/v11i48/139802, December 2018.

[13] BlessyBoaz ,Kavitha.N, Anomaly Detection Based Intrusion Detection System Using Machine Learning Under Parallel Processing Framework, International Journal of Pure and Applied Mathematics Volume 118 No. 24 2018.