ISSN: 2663-5798 || Arab Journal for Scientific Publishing

الإصدار السابع – العدد ثمانية وستون
تاريخ الإصدار: 2 – حزيران – 2024م
www.ajsp.net

"Network Attacks and its Harm on the Security and Privacy of Network (Intrusion Detection) Machine Learning Algorithms"

Researchers:

Researcher Abrar Ali Naseer Al-Dosary, Nader Omar Fadl Al-Sayed Hamad

Department of Computer Science, College of Computing and Information Technology, University of Bisha, Bisha, Kingdom of Saudi Arabia

ISSN: 2663-5798 || Arab Journal for Scientific Publishing

الإصدار السابع – العدد ثمانية وستون
تاريخ الإصدار: 2 – حزيران – 2024م
www.ajsp.net

Abstract:

Intrusion detection system can identify attacks in a variety of situations while plays a vital role in preventing data or information from being misused. Over the years, Intrusion Detection Systems (IDS) have proven to be an effective technology for achieving security by identifying malicious actions. we propose a learning supervised and unsupervised algorithms as (RF ,SVM , Naïve Bayes and AdaBoost) to extract and classify intrusion detection data sets, KDDCup99 and the NSL KDD where they found as the most usable cited. An experiment was performed to compare the performance of several machine learning methods. The results demonstrate the most accurate strategy in terms of detection rate and false alarm rate. The RF algorithm gives the best accuracy with the two datasets but it needs one of the highest values for building model. RF gives minimum classifier errors. In addition, using the unsupervised learning, the K-means gives the best results by overall factors. The results obtained from the code in the classification process also show that the random forests recorded the best accuracy. RF gives the minimum classifier errors .While NP records the lowest value in the accuracy of the results and the highest in the emergence of errors. These results are better than those obtained in the scientific paper for R. Ravipati And A. Munther[13].

Introduction:

Intrusion detection system (IDS) is a sequence of actions aimed at preventing a malicious activity on a data, resources, and systems to protect them against assaults and unauthorized access, and control the entire access of information, And though the network security is currently a major topic of computer cyber security research [1], Intrusion jeopardize the integrity, confidentiality, or availability of any resource on a computing platform. At the moment, there are two types of intrusion detection techniques: misuse-based detection (also known as signature-based detection or knowledge-based detection) and anomaly-based detection (also known as behavior-based detection) [3], Misuse-based detection systems manually code the system's discriminative characteristics and patterns from known attacks. To detect assaults, these criteria are compared to the traffic. They are effective and efficient at detecting recognized types of attacks, with a low False Alarm rate. As a result, misuse-based detection systems are now the most used NIDSs. High detection rates, fewer false alarms, less CPU cycles, and rapid intrusion detection are all features of good intrusion detection systems. The most significant component in the effectiveness of an intrusion detection system is feature selection for intrusion detection. Artificial neural networks, fuzzy association rules, Bayesian network, clustering, decision trees, ensemble learning, support vector machine [4,5,6,7,8,9,10,11,12], and other methods have been developed to simulate network activity and detect abnormal flows. Hence intrusion prevention systems may be classify as the following four types.

- **Network Based Intrusion Prevention System** — Analyzes protocol behavior to monitor the whole network for suspicious traffic.
- **Wireless Intrusion Prevention System** — Analyzes wireless networking protocols to monitor the whole network for suspicious traffic.
- **Network Behaviors Analysis** — Analyzes the network for threats such distributed denial of service.
- **Host Based Intrusion Prevention System** - A piece of software that monitors and analyzes a host for unusual activities [13].

Classification method can be useful for both misuse detection and anomaly detection, but it is more commonly used for misuse detection. we presented a data classification for intrusion detection that can be achieved using The KDD 99 and NSL-KDD data set, thus by supplying it with The KDD 99 and NSL-KDD, containing pre-labeled normal or abnormal sequences. Different techniques like, Decision Tree, Random Forest (RF), Naïve Bays(NB), support-vector machines(SVM) or rule based methods is used to scan the network traces. Dataset will be evaluated and divided into four clusters that represent the four most prevalent forms of assaults, where the speed with which various clustering techniques are executed and investigated by analyzing the targeted dataset to discover the most susceptible protocol used by attackers to launch network-based attack.

Related works:

US. T, SUN. H, ZHU. J, WANG. S and, LI. Y. BAT used a BAT-MC models attention mechanism to Extract the relevant features for malicious traffic identification from each packet's traffic bytes using BLSTM model, then they fed the selected feature into a fully linked layer for feature fusion at the output layer where the fused features was given into a classifier. A high level of promising accuracy compared with other deep learning-based models was obtained [26].

Hossein Shapoorifard et al looked for improving classification performance in the CANN intrusion detection method and tested them using the NSLKDD dataset. CANN methodology was introduced to improve the KNN classification process based on the center of mass and the nearest neighbor , but now includes the farthest neighbor(k-FN) in conjunction with the nearest neighbor (KNN) and the second nearest neighbor when both nearest and farthest neighbors had the same class classification. These new technologies improved accuracy detection rate, and lowered alert failure rate. 2017[27].

Sandeep Gurung et al used a deep learning-based approach to detect network infiltration. The NSL-KDD dataset was used to train it. The system use a deep network to train itself on anomaly patterns and distinguish between normal connections and intrusions in network data. The technology has a higher accuracy rate than signature-based intrusion detection systems and decreases the likelihood of false positives and negatives. The system was able to installed on any server to monitor the network activity of any organization in real time. Sandeep Gurung et al. 2019 [20].

Bhupendra Inger et al suggested a decision tree-based intrusion detection approach (DT) and used a CART algorithm to classify patterns, with the Gini index as a division criterion and correlation-based feature selection (CFS) for dimensionality reduction. According to the experimental results of feature selection utilizing the feature selection subgroup (CFS) assessment approach. However, the solutions require extensive preprocessing of traffic data and extensive feature engineering. 2017 [28].

Hariharan Rajadurai1 et al, introduced multiple classifier system method called stack model, which employs a series of basic classifiers to build fresh training data in order to categorize unknown data. The NSL-KDD dataset was stacked to compare its performance to that of other common machine learning algorithms like ANN, CART, Random Forest, support vector machines. Experiments show that stacked group learning is a better strategy for identifying attacks than other currently used methods with a promoting accuracy..2020 [2].

Razan Abdulhammed et al, developed an effective intrusion detection system employing various ways for dealing with unbalanced data sets from the current Coburg Intrusion Detection Dataset (CIDDS-001). She presented deep and variable auto-encoding (VAE), voting, random forest, and stacking classifiers for machine learning to detect anomaly-based intrusion of unbalanced network traffic. Deep neural networks was used to investigate and test the effectiveness of sampling strategies on CIDDS-001.the suggested system was able to detect attacks with up to 99.99 percent accuracy.2019 [29].

Ahmad. I, Basher .M, Iqbal. M and, Raheem, presented a variety of machine learning techniques, some of which are better suited to huge data processing for network and information system intrusion detection. Different machine learning algorithms, including as SVM, RF, and ELM, are examined and compared in this paper to handle the challenge of detection accuracy. The used NSL–knowledge discovery was regarded as a standard for assessing intrusion detection algorithms2018 [30].

Dr. Siddhartha Chobe et al, utilized a stealth machine approaches dynamic security defense technology to create a comprehensive network security device when combined with existing security and support intrusion technique, where supervised intrusion detection methods were evaluated using training datasets with varied attack combinations and ratios. The simulation results show that the overall detection rates are the same in each attack group. In the maximum detection rate investigation, random forests and decision trees perform well in DoS detection, whereas Naive Bayes and Gaussian do better in the other attack categorization. 2021 [31].

Uni-Kersal-Iver et al, used intrusion detection NSL-KDD dataset with classification and three standard machine learning techniques to determine the optimal technique for the classification domain to analyze and observe the performance of different classification algorithms. Machine learning techniques were employed to establish the best technique for the classification domain. They conducted two tests. Every experiment is carried out twice. Where decision tree produced the best intrusion detection categorization results. The obtained results showed better efficiency and accuracy rates. 2019 [32].

Suchet Sabri et al, distinguished between two types of taxonomy sub-problems: binary taxonomy and qualitative taxonomy. He were able to determine that the NSL-KDD data be superior to the KDDCup99 data set in terms of quality. This is due to the fact that the classifiers trained on the KDDCup99 dataset tended to be redundant, allowing them to obtain higher accuracy. 2019 [33].

Ravipati Rama Devi and colleagues , presented an overview of various Intrusion Detection System (IDS) machine learning algorithms, including LR, KNN, Decision Tree, SVM, Naive Bayes, Multi-Layer Perceptron (MLP), Random Forest, and AdaBoost Algorithm, as well as various detection methodologies, classifiers for KDD 99 and NSL-KDD. According to the data, KNN has a high false rate and detection rate, whereas the AdaBoost approach has a very low false rate and a very high detection rate, and the algorithm runs faster than other supervised algorithms. [13].

Hafsa Bundy et al, offered a new Clustering-KNN approach for PCA-fuzzy (Principal Component Analysis Cluster and Fuzzy Clustering using K-Nearest Neighbor feature selection techniques. The efficiency and accuracy of the intrusion detection system will steadily decrease with the increase in the amount of data, according to the results of the experiment. Compared to Bayesian and QR-OMPCA methods. 2018 [21]

Guangchen Zhao et al , introduced an intrusion detection methods, a deep belief network (DBN) and a probabilistic neural network have been proposed (PNN). the model was trained and tested using the KDD CUP 1999 data set., they experimented their work using MATLAB R2010a. To measure the advantages of the algorithm, methods using DBN-PNN without PSO, PCA-PNN and conventional PNN are used as contrast experiments. This approach was superior to standard PNN, PCA-PNN, and non-optimized DBN-PNN according to experimental results.. 2017 [34].

Iram. A, Ayub. Z and, Masoodi. F, extracted features from network traffic and payloads using word embedding and a text-convolutional neural network (Text-CNN). Statistical and payload features are integrated, and the final categorization is done with a random set. The advised strategies was effective, according to extensive empirical studies. Erxue Min and colleagues (2018) [19].

Problem statement:

Reconnaissance attacks are a type of attack that collects information about a target. Where they can be either Logical reconnaissance and Physical reconnaissance. Logical reconnaissance does not require a human intervention, while Physical reconnaissance is being monitored by a network administrator with the need of intrusion detection software, or by placing other devices in the network infrastructure. Therefore, when the network is not able to transmit traffic, a type of DDOS (Distributed Denial of Service) attacks may occur, in this case, certain applications try to reformat the data on the system [16]. Therefore, instead of wasting time and possible incorrect result in the manually search for investigating log files to detect any intrusion case, automatic log file readers have been rapidly developed to extract and detecting infiltration by unauthorized personnel.

With sufficient processing speed available, it become possible to check attack patterns after the event, and monitor in 'real time' and trigger alarms if incursions are identified. And so, initial software focused on stopping out the ongoing attacks with an alert feedback, where it become a part of an existing firewall systems. But recently Many intrusion detection systems use a rule-based expert system or a statistical identifier discovery system as their detector, meanwhile new intrusions still difficult to be detected [21]. Fault detection and troubleshooting systems produce a slew of problems, According to a literature review, most researchers used a single algorithm to detect various attack categories for intrusion detection, with poor results in some circumstances. including a large number of alerts, a high incidence of false alarms, inadequate generalization, and inaccurate reporting to achieve acceptable efficiency and effectiveness, and to test and validate the proposed intrusion detection method using a skew-based intrusion detection KDD 99 & NSL-KDD datasets. In order to find the impact of Network attacks on the security and privacy of networks and to find the methods for detecting from Network attacks, thus we can train intrusion detection IDS tools to deal with Network attacks impact.

Methodology:

In our experiments, Many machine learning algorithms were applied. SVM, KNN, LR, NB, MLP, RF, AdaBoost Algorithm, and DT, to categorize data as normal or invasive after extracting the more relevant features using CPA and Invariant feature selection techniques. We evaluated our models using four different feature subsets derived from the NSL KDD datasets. Our proposed method for intrusion detection is comprised of three main steps : Pre-processing, Aspect Extraction and Aspect Type Classification, The data is pre-processed in the first stage using the following python code

```
def clean_dataset(df):
    assert isinstance(df, pd.DataFrame), "df needs to be a pd.DataFrame"
    df.dropna(inplace=True)
    indices_to_keep = ~df.isin([np.nan, np.inf, -np.inf]).any(1)
    return df[indices_to_keep].astype(np.float64) .
```

The second step is to identify the assessed model's four fundamental feature sets. To train and test the data, the third stage is to apply different machine learning classifiers. Finally, the various parameters' outcomes are examined.

KDD-99cup has a reasonable number of records dataset, it was downloaded from https://www.kaggle.com/datasets/hassan06/nslkdd. And is believed to be applied as an effective benchmark dataset to help in compare between different intrusion detection methods, we found it affordable to run in experiments after randomly selection small portions, This dataset is made up of TCP dump data from simulated network traffic gathered at Lincoln Labs in 1998. Five million connection records were used for a training set after seven weeks of traffic. A test set of two million samples was created after another two weeks of network activity.

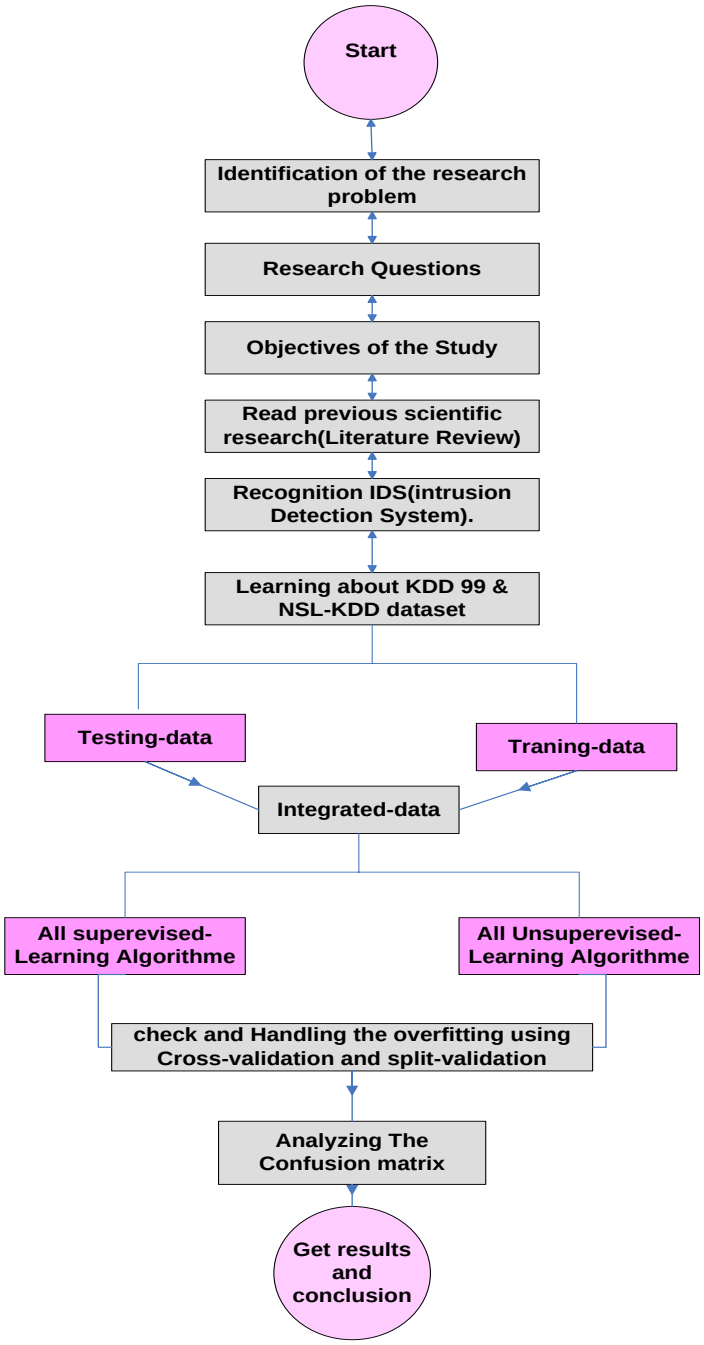

Fig 1: Roadmap for the entire work of testing and training the selected dataset

The full schedule is available on the MIT website []. This data has been filtered into KDD-99. Normal, DoS (Denial of Service), U2R (User to Root), R2L (Remote to Local), and Probe are the five types of patterns in KDD-99 (Probing Attack). Each intrusion category is further divided into subcategories based on the procedure employed to carry out the attack. Table 3 shows the distribution of patterns across target classes. It contains 41 distinct features that fall into one of three categories: There are three types of content: basic, traffic, and content.

Meanwhile NSL-KDD consists of only selected records from the entire KDD dataset and is free of any of the issues. Tavallaee et al. [23] attempted to correct KDD-99 and overcome its flaws with NSL-KDD. However, as the authors point out, there are still some issues with the dataset. Our target here is to merge the training set with testing set to increase the quality of results.

ISSN: 2663-5798 || Arab Journal for Scientific Publishing

الإصدار السابع – العدد ثمانية وستون
تاريخ الإصدار: 2 – حزيران – 2024م
www.ajsp.net

Dataset	No of records	No of classes	Normal	Malicious
KDD-99 CUP	125973	2	67343	58630
NSL-KDD	58630	58630	58630	58630

Table 1: details of dataset

Results and Discussion:

The proposed solution is trained on datasets shown in Table 1 using different algorithm such as , Logistic Regression (LR), Decision tree , Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), AdaBoost, K-nearest neighbors, then the performance of the proposed solution is assessed using various evaluations such as accuracy, mean error and time of execution, Each dataset is divided into testing and training sets. For training, 80% of the data were used with 20% for testing. We will examine the performance of the proposed solution using both WEKA , and Python code. In WEKA we implemented a cross-validation supervised learning with both 5 and 10 number of folds . The accuracy results of conducting experiments on NSL-KDD using the selected algorithms are summarized in table 2 and visualized in Fig.2, while table3 shows the False Positive (FP) rate. Many algorithms shows a promoting results as RF, J48 and K-nearest but the RF algorithm 10.

Algorithm	10 folds	5 folds
LR	95.8	95.8
J48	99.6	99.5
RF	99.8	99.6
SVM (LibLINEAR)	87.2	89.3
Naïve Bayes	87.3	87.3
Naïve Bayes Updateable	87.3	87.3
Simple Logistic	95.4	95.3
AdaBoostM1	94.1	94.3
K-nearest neighbours	99.5	99.3

Table 2: Accuracy results of NSL-KDD

Fig2: accuracy result of NSL-KDD

Algorithm	10 folds	5 folds
LR	0.992	0.992
J48	0.998	0.998
RF	1	1
SVM (LibLINEAR)	0.124	0.011
NaiveBayes	0.956	0.956
NaiveBayesUpdateable	0.956	0.956
SimpleLogistic	0.99	0.99
AdaBoostM1	0.982	0.982
K-nearest neighbours	0.978	0.978

Table 3: Fault and Positive of NSL-KDD

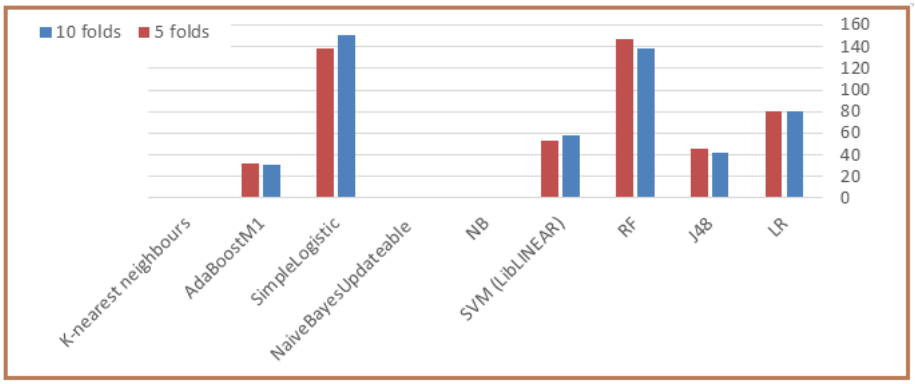

Fig 3: Run out time for each proposed algorithm NSL_KDD

The accuracy results of conducting experiments on KDD-99 CUP using the proposed algorithms shown in Table 4 and visualized in Fig.4, while table 5 shows the False Positive (FP) rate.

Algorithm	**10 folds**	**5 folds**
LR	97.5	97.5
J48	99.78	99.76
RF	99.9	99.91
SVM (LibLINEAR)	95.4	92.8
NaiveBayes	90.3	90.4

NaiveBayesUpdateable	90.1	90.2
SimpleLogistic	97.2	97.2
AdaBoostM1	94.5	94.5
K-nearest neighbours	99.7	99.6

Table 4: Accuracy result for KDD

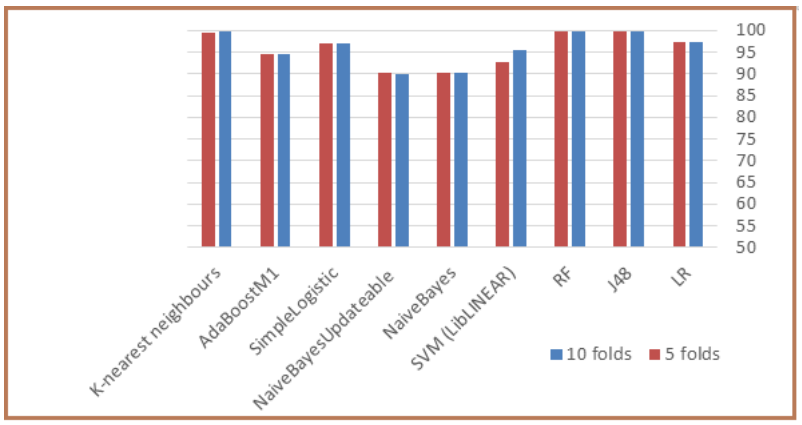

Fig 4: Accuracy result for KDD

Algorithm	10 folds	5 folds
LR	0.997	0.997
J48	0.999	0.999
RF	1	1
SVM (LibLINEAR)	0.953	0.929
Naïve Bayes	0.966	0.966
Naïve Bayes Updateable	0.966	0.966
Simple Logistic	0.99	0.99
AdaBoostM1	0.988	0.988
K-nearest neighbours	0.999	0.999

Table 5: False Positive (FP) rate of NSL-KDD

ISSN: 2663-5798 || Arab Journal for Scientific Publishing
الإصدار السابع – العدد ثمانية وستون
تاريخ الإصدار: 2 – حزيران – 2024م
www.ajsp.net

Results shows a promoting algorithms as RF, J48 and SVM but the RF algorithm with 10 folds. Meanwhile fig 5 illustrates the total runtime consumed by each proposed solution.

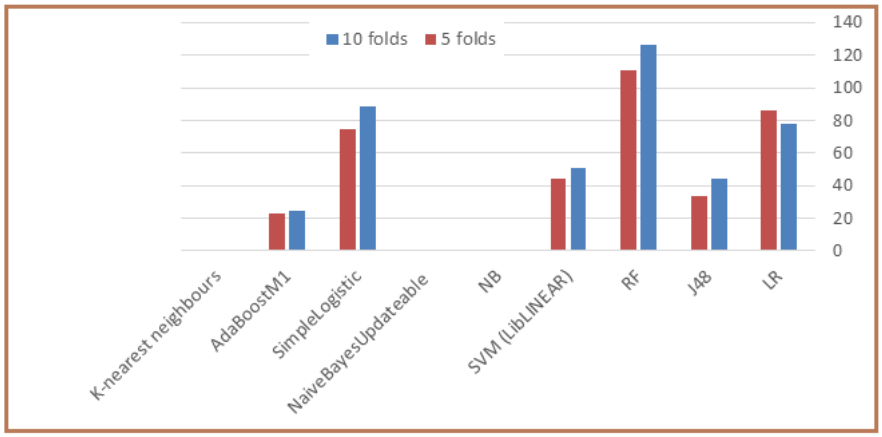

Fig 5: Total runtime for each proposed solution of KDD

The accuracy results of conducting experiments on NSL-KDD using the selected unsupervised algorithms are summarized in Table 6 and visualized in Fig.6 while Table 7 and Figure 7 show the taken time to build the model.

Algorithm	Accuracy
K-means	88.4
Farthest First	60.5
Canopy	82.3
Make Density Based Clusterer	76.4
Simple EM (expectation maximization)	83.3

Table 6: Accuracy results of NSL-KDD

Algorithm	**Time**
K-means	1.34
Farthest First	0.38
Canopy	1.58
Make Density Based Cluster	1.53
Simple EM (expectation maximization)	18.09

Table 7: Time taken to build model of NSL-KDD

The total runtime for K-,means is securing a promoting result witch is one of the best two.

The accuracy results of conducting experiments on KDD-99 CUP using the selected unsupervised algorithms are summarized in Table 8 and visualized in Fig.6 while Table 9 and fig.7 shows the taken time to build the model

unsupervised algorithms	**Accuracy**
K-means	84.6
Farthest First	81.1
Canopy	84.6
Make Density Based Clusterer	84
Simple EM (expectation maximization)	86.7

Table 8: Accuracy results of KDD-99 CUP

Algorithm	**Time**
K-means	3.66
Farthest First	0.41
Canopy	1.29
Make Density Based Clusterer	1.84
Simple EM (expectation maximization)	13.58

Table 9: Time taken to build model of KDD-99 CUP

Many algorithms give close results meanwhile the Simple EM shows better result and it consumes more time among all algorithms. So, K-means and Canopy are the best.

Python code result:

Algorithm	Accuracy	Error	Time
Random Forest	99.4	0.007	13.83
Decision Tree	98.2	0.017	10.78
NB	53.6	0.46	7.24
SVM	94.0	0.06	124.90

Table 10: Classification using NSL-KDD, 5-folds

Algorithm	Accuracy	Error	Time
Random Forest	98.1	0.018	17.55
Decision Tree	95.7	0.043	12.31

NB	52.4	0.47	8.47
SVM	90.3	0.087	133.

Table 11: Classification using KDD, 5-folds

Algorithm	Accuracy	Error	Time
Random Forest	98.8	0.012	32.13
Decision Tree	95.5	0.043	15.53
NB	52.5	0.47	17.92
SVM	91.0	0.087	350.68

Table 12: Classification using NSL, 10-folds

Algorithm	Accuracy	Error	Time
Random Forest	99.2	0.006	27.23
Decision Tree	98.1	0.017	15.81
NB	53.2	0.46	12.31
SVM	90.5	0.06	167.93

Table 13: Classification using KDD, 10-folds

Result have been recorded using previous algorithm methods , have been compared by many dimensionality reduction techniques used for supervised classification problems in machine learning, these technique can be used to project the features of higher dimensional space into lower-dimensional space to reduce dimensional costs, table 13,14 shows the obtained result.

Algorithm	Accuracy	Time (sec)
LogisticRegression	94.0	0.203
GaussianMixture	50.0	1.062
RidgeClassifier	90.0	0.140
SGDClassifier	88.0	0.124
PassiveAggressiveClassifier	86.0	0.124
KNeighborsClassifier	89.0	01.24
ExtraTreeClassifier	85.0	0.109
AdaBoostClassifier	85.0	4.012
BaggingClassifier	89.0	0.109
ExtraTreesClassifier	90.0	6.397
GaussianProcessClassifier	91.0	0.296
GradientBoostingClassifier	92.0	3.153
LinearDiscriminantAnalysis	90.0	0.109
QuadraticDiscriminantAnalysis	78.0	0.109
Kmean	52.0	4.470

Table 14: Methods dimensionality reduction techniques – NSL Dataset

Algorithm	Accuracy	Time
LogisticRegression	94.0	0.281
GaussianMixture	50.0	1.280
RidgeClassifier	90.0	0.166
SGDClassifier	88.0	0.229
PassiveAggressiveClassifier	86.0	0.109
KNeighborsClassifier	89.0	0.109

ExtraTreeClassifier	83.0	0.109
AdaBoostClassifier	85.0	4.154
BaggingClassifier	90.0	1.119
ExtraTreesClassifier	90.0	6.165
GaussianProcessClassifier	91.0	0.296
GradientBoostingClassifier	91.0	3.211
LinearDiscriminantAnalysis	90.0	0.187
QuadraticDiscriminantAnalysis	85.0	0124
Kmean	53.3	3.676

Table 15: Methods dimensionality reduction techniques – KDD Dataset

Conclusions:

Network attack types was discussed using different machine learning algorithms of Intrusion Detection System (IDS) and different detection methodologies as well as the ways of intrude detection. Carefully scrutinize the results of supervised learning accuracy, therefore a scientific contribution was summarized in by integrating the learning data with the test data and checking the accuracy using the (Cross-validation) method, RF algorithm promote all other algorithm using datasets as shown in Table. RF gives minimum classifier errors as shown in table 10. In addition, using the unsupervised learning, the K-means gives the best results by overall factors as shown in table 11. The Result have been recorded using previous algorithm methods , then compared by dimensionality reduction techniques in machine learning, where table 13,14 shows the obtained result accordingly. But still RF algorithm promote all, so as to avoid fake accuracy (Over-fitting). my additional goal is to try more machine learning algorithms to try to get a higher classification accuracy than what has been found in previous research.

REFERENCES

Kirsal Ever .Y, Sekeroglu. B, & Dimililer. K. Classification Analysis of Intrusion Detection on NSL-KDD Using Machine Learning Algorithms. 2019.

Rajadurai. H., Devi Gandhi. H. A stacked ensemble learning model for intrusion detection in wireless network. 2020.

Behrouz. A, Debdeep M. (2017), Cryptography and Network Security, (Third Edition), McGraw Hill Education (India) Private Limited.

F. Salo, A.B.Nassif, A. Essex, Dimensionality reduction with IG-PCA and ensemble classifier for network intrusion detection., Comput er Networks ,. 2019.

M. Safaldin, M. Otair, L. Abualigah. Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks, Journal of ambient intelligence and humanized computing,. 2020.

M.A. Ferrag, L. Maglaras, A. Ahmim, M. Derdour, H. Janicke.RDTIDS: Rules and decision tree-based intrusion detection system for internet of things networks.Future internet. 2020.

I.H. Sarker, Y.B. Abushark, F. Alsohami, A.I. Khan.IntruDTree: A machine learning based cyber-security intrusion detection model.2020.

K. Chisholm, C. Yakopcic, M.S.Alam, T.M. Taha,.Multilayer perceptron algorithms for network intrusion detection on portable low power hardware., in 10th annual computing and communication workshop and conference (CCWC), Las Vegas, USA, 2020.

T.B.Prasad, P.S.Prasad, K.P.Kumar, .An intrusion detection system software program using KNN nearest neighbors approach., International journal of science research and innovation engineering (IJSRIE). 2020.

S.Rajagopal, P.P. Kundapur, K.S.Hareesha, .A stacking ensemble for network intrusion detection using heterogeneous datasets, Secure Communication Networks,. 2020.

W.Wang, Y. Li, X. Wang, J. Liu, X. Zhang, .Detecting android malicious apps and categorizing benign apps with ensemble of classifiers", Future Generation Computing System, .2018.

C. Ambikavathi, S.K.Srivatsa, .Predictor selection and attack classification using random forest for intrusion detection., Journal of scientific and industrial research,. 2020.

Rama Devi.R, Abualkibash.M. INTRUSION DETECTION SYSTEM CLASSIFICATION USING DIFFERENT MACHINE LEARNING ALGORITHMS ON KDD-99 AND NSL-KDD DATASETS - A REVIEW PAPER. 2019.

http://nsl.cs.unb.ca/NSL-KDD/

http://www.cs.waikato.ac.nz/ml/weka/

Hall. P, Street. L and , Saddle River. U. NETWORK SECURITY ESSENTIALS: APPLICATIONS AND STANDARDS FOURTH EDITION. Boston. Library of Congress Cataloging-in-Publication Data. 2011.
https://blog.avast.com/avast-releases-2020-threat-landscape-report
Allen J, Christie A, Fithen W, McHugh J, Pickel J, Stoner E. State of the practice of intrusion detection technologies, Carnegie Mellon University Technical Report .2000.
Iram. A, Ayub. Z and, Masoodi. F. A Machine Learning Approach for Intrusion Detection System on NSL-KDD Dataset. 2020.
Gurung. S, Kanti Ghose. S and, Subedi. A. Deep Learning Approach on Network Intrusion Detection System using NSL-KDD Dataset.2019.
BENADDI. H, IBRAHIMI. K and, IBRAHIMI. A. Improving the Intrusion Detection System for NSL-KDD Dataset based on PCA-Fuzzy Clustering-KNN.2018.
W. Lee, S. J. Stolfo, K. W. Mok. A Data Mining Framework for Building Intrusion Detection Models., IEEE Symposium on Security and Privacy, Oakland, California, 1999.
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.
W. Lee, S. J. Stolfo, and K. W. Mok. Mining in a Data-Flow Environment: Experience in Network Intrusion Detection., In Proceedings of the 5th ACM SIGKDD, San Diego, CA, 1999.
Sapna S. Kaushik, Dr. Prof.P.R.Deshmukh, Detection of Attacks in an Intrusion Detection System, International Journal of Computer Science and Information Technologies.2011.
US. T, SUN. H, ZHU. J, WANG. S and, LI. Y. BAT: Deep Learning Methods on Network Intrusion Detection Using NSL-KDD Dataset.2020.
H. Shapoorifard and P. Shamsinejad,. Intrusion detection using a novel hybrid method incorporating an improved KNN. Int. J. Control Automat. 2017.
B. Ingre, A. Yadav, and A. K. Soni,. Decision tree based intrusion detection system for NSL-KDD dataset. in Proc. Int. Conf. Inf. Commun. Technol. Intell. Syst., 2017.
Abdulhammed R , Faezipour . M, r Abuzneid . A and, AbuMallouh A. Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic. IEEE .2019.
Ahmad. I, Basher .M, Iqbal. M and, Raheem. A. Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. IEEE Access 6:33789–33795.2018.
Diwan. T, Choubey. S and, Hota. H. A Detailed Analysis on NSL-KDD Dataset using various Machine Learning Techniques for Intrusion Detection.2021.
Ever. Y, Sekeroglu. B and, Dimililer. K. Classification Analysis of Intrusion Detection on NSL-KDD Using Machine Learning Algorithms.2019.
Sapre.S, Ahmadi. B and, Islam. K. A Robust Comparison of the KDDCup99 and NSL-KDD IoT Network Intrusion Detection Datasets Through Various Machine Learning Algorithms.2019.
G. Zhao, C. Zhang, and L. Zheng. Intrusion detection using deep belief network and probabilistic neural network. in Proc. IEEE Int. Conf. Comput. Sci. Eng. (CSE), IEEE Int. Conf. Embedded Ubiquitous Comput. (EUC), Jul. 2017.
K. Wu, Z. Chen, and W. Li, .A novel intrusion detection model for a massive network using convolutional neural networks. IEEE Access, vol. 6, pp. 50850–50859, 2018.
S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan. Costbased modeling for fraud and intrusion detection: Results from the jam project.