IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204276 |

Fraud App Detection Using Machine Learning

Mrs. A Subha(M.E)., M.Mounika, Khatiza Kubra, Sreeram C.V

Associate Professor, Dept. of CSE., JNTUA University, Kuppam Engineering college, AP, India

UG Students, Dept. of CSE., JNTUA University, Kuppam Engineering college, AP, India

ABSTRACT Objective: A novel supervised machine learning system is developed to classify network Fraud Application whether it is malicious or benign. To find the best model considering detection success rate, combination of supervised learning algorithm and feature selection method have been used. Through this study, it is found that Random Forest (RANDOM FOREST) based machine learning with wrapper feature selection outperform support vector machine (PCA) technique while classifying network Fraud Application. To evaluate the performance, NSL-KDD dataset is used to classify network Fraud Application using PCA and RANDOM FOREST supervised machine learning techniques. Comparative study shows that the proposed model is efficient than other existing models with respect to Fraud Application detection success rate.

In supervised learning, learning data comes with labels or desired outputs and the objective is to find a general rule that maps inputs to outputs. This kind of learning data is called labeled data .The learned Rule is then used to label new data with unknown outputs. It involves building a machine learning model based that is based on labeled samples. Reducing the impact of attacks; and secondly the evaluation of the system IDS. Indeed, in one hand the IDSs collect network Fraud Application information from some sources present in the network or the computer system and then use these data to enhance the systems safety. In the other hand, the evaluation of IDS is a critical task. In fact, its important to note the difference between evaluating the effectiveness of an entire system and evaluating the characteristics of the system components.

INDEX TERMS: Supervised Learning: In supervised learning, historical data with labeled examples of fraud and non-fraud cases are used to train predictive models. Algorithms such as logistic regression, decision trees, random forests, and gradient boosting are commonly employed.

Unsupervised Learning: Unsupervised learning is used when labeled data is scarce. Anomaly detection algorithms, such as isolation forests and k-means clustering, can identify unusual patterns or outliers indicative of fraudulent behavior.

Semi-supervised Learning: Combining both labeled and unlabeled data to improve fraud detection.
Feature engineering for fraud detection for credit card fraud paper, The autor examines a new approach for developing features for machine learning algorithm They address the cost-sensitivity, and the features are preprocessing to achieve improved fraud detection and savings.

Deep learning for fraud detection Credit card fraud detection deep learning algorithms are machine learning models that are trained on large datasets of credit card transactions, both fraudulent and non-fraudulent. The algorithms learn to identify the patterns and anomalies in the data that are most indicative of fraud.

I. INTRODUCTION

In this paper author is evaluating performance of two supervised machine learning algorithms such as PCA (Support Vector Machine) and RANDOM FOREST (Random Forests). Machine learning algorithms will be used to detect whether request data contains normal or attack (anomaly) signatures. Now-a-days all services are available on internet and malicious users can attack client or server machines through this internet and to avoid such attack request IDS (Network Fraud Application Detection System) will be used, IDS will monitor request data and then check if its contains normal or attack signatures, if contains attack signatures then request will be dropped.

IDS will be trained with all possible attacks signatures with machine learning algorithms and then generate train model, whenever new request signatures arrived then this model applied on newrequest to determine whether it contains normal or attack signatures. In this paper we are evaluating performance of two machine learning algorithms such as PCA and RANDOM FOREST and through experiment we conclude that RANDOM FOREST outperform

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204276 |

existing PCA in terms of accuracy.

II. ADVANTAGES OF PYTHON OVER OTHER LANGUAGES

1. Less Coding

Almost all of the tasks done in Python requires less coding when the same task is done in other languages. Python also has an awesome standard library support, so you don't have to search for any third-party libraries to get your job done. This is thereason that many people suggest learning Python to beginners**.**

2. Affordable

Python is free therefore individuals, small companies or big organizations can leverage the free available resources to build applications. Python is popular and widely used so it gives you better community support.

3. Python is for Everyone

Python code can run on any machine whether it is Linux, Mac or Windows. Programmers need to learn different languages for different jobs but with Python, you can professionally build web apps, perform data analysis and **machine learning**, automate things, do web scraping and also build games and powerful visualizations. It is an all-rounder programming language.

Disadvantages of Python

So far, we've seen why Python is a great choice for your project. But if you choose it, you should be aware of its consequences as well. Let's now see the downsides of choosing Python over another language.

1. Speed Limitations

We have seen that Python code is executed line by line. But since Python is interpreted, it often results in **slow execution**. This, however, isn't a problem unless speed is a focal point for the project. In other words, unless high speed is a requirement, the benefits offered by Python are enough to distract us from its speed limitations.

2. Weak in Mobile Computing and Browsers

While it serves as an excellent server-side language, Python is much rarely seen on the **client-side**. Besides that, it is rarely ever used to implement smartphone-based applications. One such application is called **Carbonnelle**.

III.SYSTEM ARCHITECTURE

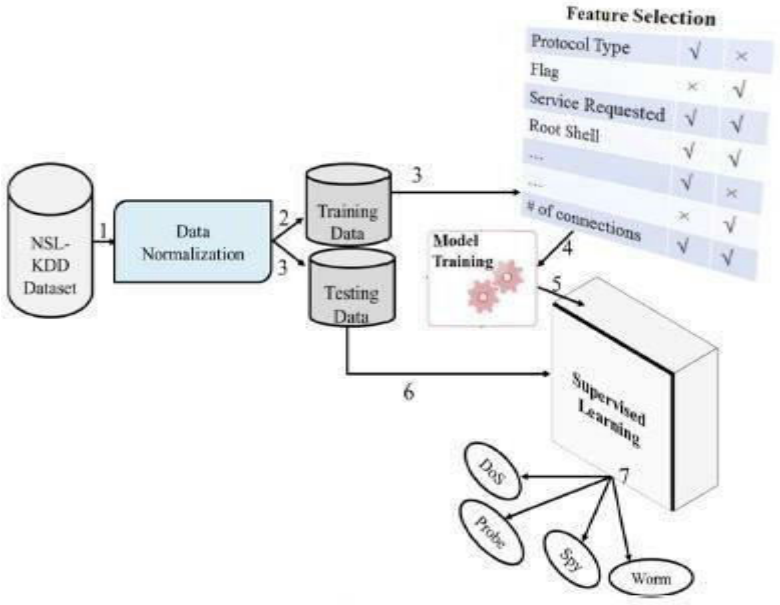


Fig 1: Proposed supervised machine learning classifier system

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204276 |

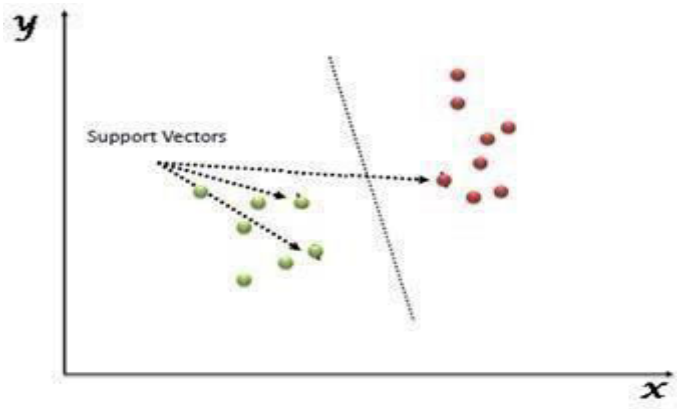
Module description

Feature selection is an important part in machine learning to reduce data dimensionality and extensive research carried out for a reliable feature selection method. For feature selection filter method and wrapper method have been used. In filter method, features are selected on the basis of their scores in various statistical tests that measure the relevance of features by their correlation with dependent variable or outcome variable. Wrapper method finds a subset of features by measuring the usefulness of a subset of feature with the dependent variable. Hence filter methods are independent of any machine learning algorithm whereas in wrapper method the best feature subset selected depends on the machine learning algorithm used to train the model. In wrapper method a subset evaluator uses all possible subsets and then uses a classification algorithm to convince classifiers from the features in each subset. The classifier consider the subset of feature with which the classification algorithm performs the best.

IV. ALGORITHMS USED IN THIS PROJECT

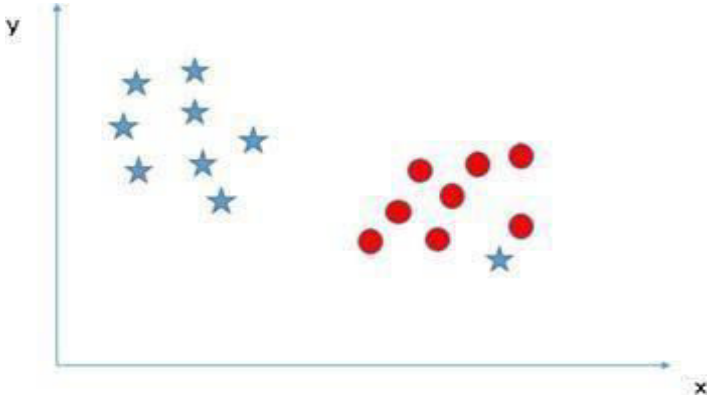
Support Vector Machine:-

"Support Vector Machine" (PCA) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the PCA algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).



Support Vectors are simply the co-ordinates of individual observation. The PCA classifier is a frontier which best segregates the two classes (hyper-plane/ line).You can look at support vectormachines and a few examples of its working here

Can we classify two classes (Scenario-4)?: Below, I am unable to segregate the two classes usinga straight line, as one of the stars lies in the territory of other(circle) class as an outlier.



As I have already mentioned, one star at other end is like an outlier for star class. The PCA algorithm has a feature to

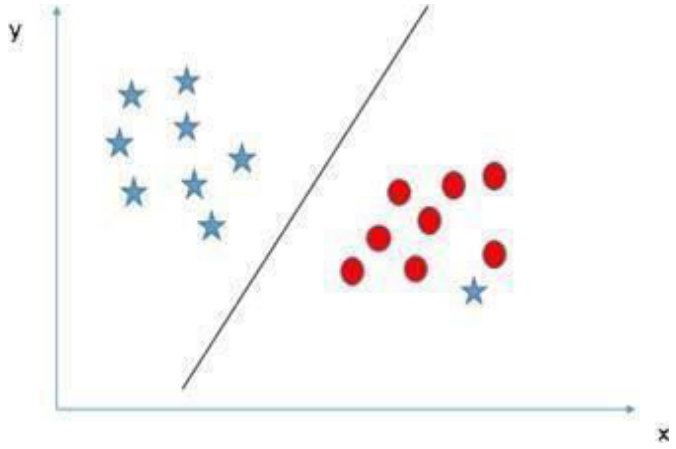
International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

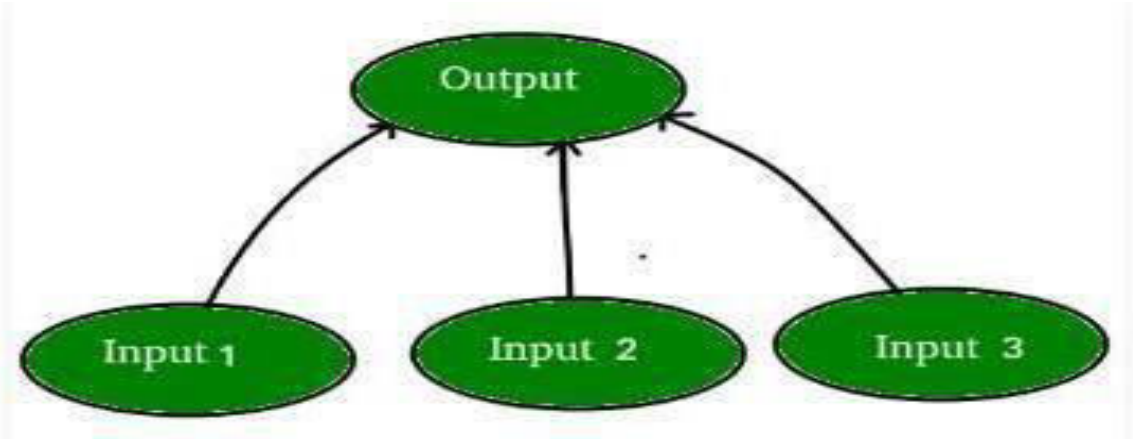
| DOI: 10.15680/IJIRCCE.2024.1204276 |

ignore outliers and find the hyper-plane that has the maximum margin. Hence, we can say, PCA classification is robust to outliers



Artificial neural network:

Artificial neural networks are one of the main tools used in machine learning. As the "neural" part of their name suggests, they are brain-inspired systems which are intended to replicate the way that we humans learn. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something that the output layer can use. They are excellent tools for finding patterns which are far too complex or numerous for a human programmer to



Let's say we have a problem where we want to predict output given a set of inputs and outputs as training example like so:

Input 1	Input 2	Input 3	Output
0	1	1	1
1	0	0	0
1	0	1	1

Fig 2: Training Examples

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204276 |

Now we want to predict the output the following set of inputs:

1	0	1	?

Fig 3: Test Example

Note that the output is directly related to third column i.e. the values of input 3 is what the output is in every training example in fig. 2. So for the test example output value should be 1.

V. CONCLUSION & FUTURE WORK

In this paper, we have presented different machine learning models using different machine learning algorithms and different feature selection methods to find a best model. The analysis of the result shows that the model built using RANDOM FOREST and wrapper feature selection outperformed all other models in classifying network Fraud Application correctly with detection rate of 97.02%. We believe that these findings will contribute to research further in the domain of building a detection system that can detect known attacks as well as novel attacks. The Fraud Application detection system exist today can only detect known attacks. Detecting new attacks or zero day attack still remains a research topic due to the high false positive rate of the existing systems.

VI. FUTURE WORK

Research further in the domain of building a detection system that can detect known attacks as well as novel attacks. The Fraud Application detection system exist today can only detect known attacks. Detecting new attacks or zero day attack still remains a research topic due to the high false positive rateof the existing systems.

REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based Fraud Application detection system using limited labeled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178– 184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the Fraud Application detection system," in International Conference on Networked Systems, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based Fraud Application-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of Fraud Application detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for Fraud Application detection," arXiv preprint arXiv:1312.2177, 2013.

[7] N. Chakraborty, "Fraud Application detection system and Fraud Application prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229– 6166, 2013.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details