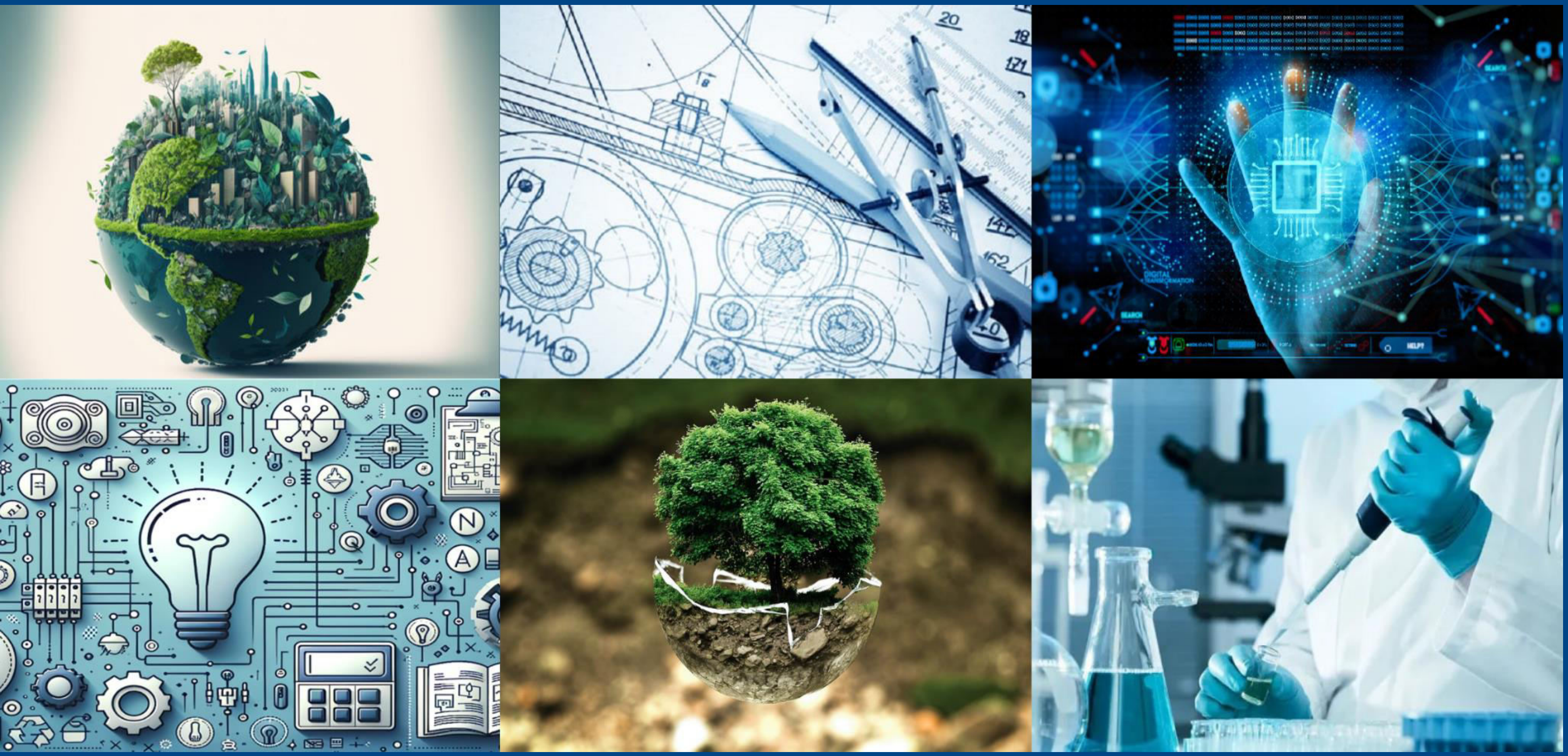
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 5, May 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Network Intrusion Detection using Machine Learning

Thanigai Kumar K V, Dr.Balaji Kannan

AssistantProfessor, Department of Computer Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies, Chennai, India

Student, Department of Computer Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies,Chennai, India

ABSTRACT: This project focuses on developing a machine learning-based Network Intrusion Detection System (NIDS) designed to classify network traffic as either normal or malicious. It begins with comprehensive data preprocessing steps including cleaning, encoding, scaling, and feature selection to prepare the data for effective model training. The system is trained and tested using labeled datasets containing various types of network activity. Several supervised learning algorithms such as K-Nearest Neighbors, Logistic Regression, Decision Trees, and Naive Bayes are implemented and evaluated. Performance metrics are used to identify the most accurate and reliable model for intrusion detection. The chosen model is saved along with preprocessing tools like the scaler and encoder to ensure consistent results in deployment. This integration makes the system accessible and practical for non-technical users. The saved models and interface work together to provide a complete, deployable solution. By detecting intrusions at an early stage, the system contributes significantly to enhancing cybersecurity. Overall, it offers a robust, efficient, and scalable approach for monitoring and preventing network threats..

KEYWORDS—NIDS, Network Intrusion Detection

I. INTRODUCTION

For the past few years, network has played a significant role in communication. The computer network allows the computing network devices to exchange information among different systems and individuals. The services of various organizations, companies, colleges, universities are accessed throughout computer network. This leads to a massive growth in networking field. The accessibility of internet has acquired a lot of interest among individuals. In this context, security of information has become a great challenge in this modern area. The information or data that we would like to send is supposed to be secured in such a way that a third party should not take control over them. When we are talking about security, we have to keep three basic factors in our mind: Confidentiality, Integrity and availability. Confidentiality means privacy of information. It gives the formal users the right to access the system via internet. This can be performed suitably along with accountability services in order to identify the authorized individuals. The second key factor is integrity. The integrity service means exactness of information. It allows the users to have self- assurance that the information passed is acceptable and has not been changed by an illegal individual.

An Intrusion Detection System (IDS) is used to watch malicious activities over the network. It can sort the unfamiliar records as normal or attack class. First monitoring of the network traffic is done, and then the IDS sorts these network traffic records into either malicious class or regular class. It acts as an alarm system that reports when an illegal activity is detected. The exactness of the IDS depends upon detection rate. If the performance is high for the IDS, then the correctness of detection is also high. Some of the intrusion detection systems are marketed with the ability to stop attacks before they are successful. They are used to shield an association from attack. It is a relative concept that tries to identify a hacker when intrusion is attempted. Ideally, such a system will only alarm when a successful attack is made. Intrusion detection system is not a perfect solution to all attack types. The various goals that can be accomplished with an Intrusion

II. SCOPE OF THE PROJECT

This project encompasses a broad range of services including:

- Software Development
- Engineering Solutions
- Systems Integration
- CRM and IT Consulting
- Product Development
- E-Commerce Solutions
- IT Outsourcing

We aim to combine innovation with responsibility to:

- Address current business challenges effectively.
- Create future-ready opportunities for clients.

Our approach is based on:

- A strategic framework called **AIM** – Architect, Integrate, Manage.
- Proven offshore development methods minimizing client effort.

Reusable frameworks to reduce time and cost.

III. METHODOLOGY

We propose below methodology for solving the problem. Raw data collected would be pre- processed for missing data, anomalies and outliers. Then an algorithm would be trained on this data to create a model. This model would be used for forecasting the final results. ETL stands for Extract, Transform and load. It is a tool which is a combination of three functions. It is used to get data from one database and transform it into a suitable format. Data preprocessing is a data mining technique used to transform sample raw data into an understandable format. Real world collected data may be inconsistent, incomplete or contains an error and hence data preprocessing is required.

EXISTING AND PROPOSED SYSTEM

Network Intrusion Detection system is a mechanism that is used within the network to identify the malicious event. It uses K- Nearest Neighbor algorithm for intrusion Detection. The network traffic is monitored in the network that is in the sub- net. If an attack is observed it matches the traffic with the known attack list. Then an alert is passed to the administrator. Network Intrusion Detection System (NIDS) and Host Intrusion Detection System (HIDS) are the two most widely used systems for intrusion detection. NIDS is installed in router to identify the passage of network traffic. HIDS runs on an individual system. The functions of two IDSs are the same. HIDS also monitors the unauthorized activity. It takes a short review of the existing files in the system. Then it matches it with the old system files. If it finds an intrusion or changes in the system, then an alert is passed to the administrator. The intrusion can be detected as if a file is modified or deleted, then it means malicious activity is reporte.

In Proposed system supervised method is used for detecting the Intrusion in the system. In order to increase the detection ability of IDS and prevent the service providers from attack, we propose an efficient ML based IDS using Light gradient boosting method and Random Forest algorithms. In order to overcome the problem of class imbalance, feature selection based on CFS-BA is used to determine a subset of the original features to eliminate irrelevant features. The detection framework of the proposed ML- Based consists of three stages including: feature selection, build and train the ensemble classifier and attack recognition. Detailed information about the framework.

IV. IMPLEMENTATION

Python is a popular programming language. It was created by Guido van Rossum, and released in 1991. The most recent major version of Python is Python 3, which we shall be using in this tutorial. However, Python 2, although not being updated with anything other than security updates, is still quite popular.

Python concepts

If you're not interested in the haws and whys of Python, feel free to skip to the next chapter. In this chapter I will try to explain to the reader why I think Python is one of the best languages available and why it's a great one to start programming with.

Open source general-purpose language.
Object Oriented, Procedural, Functional
Easy to interface with C/Object/Java/Fortran
Easy to interface with C++ (via SWIG)
Great interactive environment

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

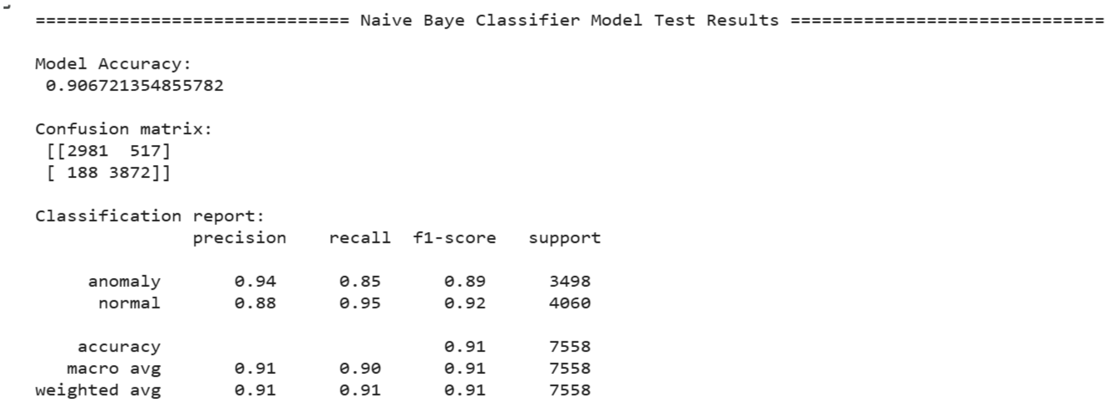
Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.

Python is Interactive − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

Python is Object-Oriented − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.

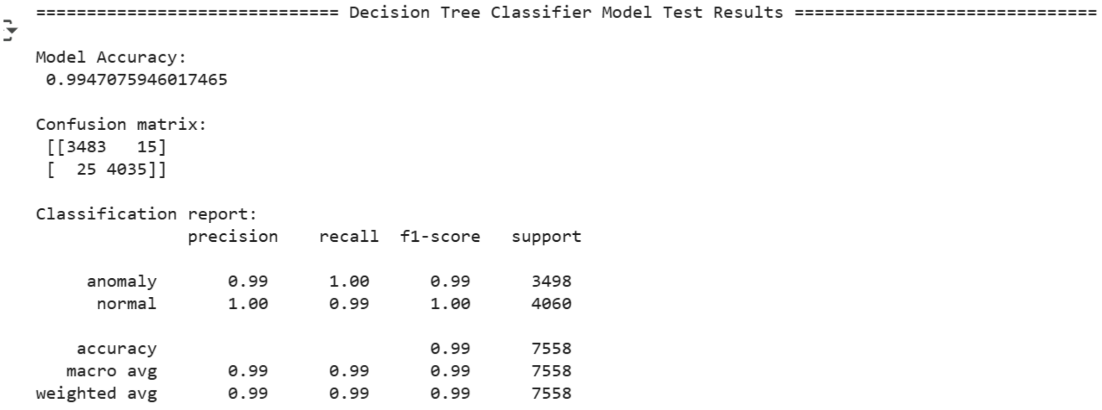
Python is a Beginner's Language − Python is a great language for the beginner- level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

V. RESULT



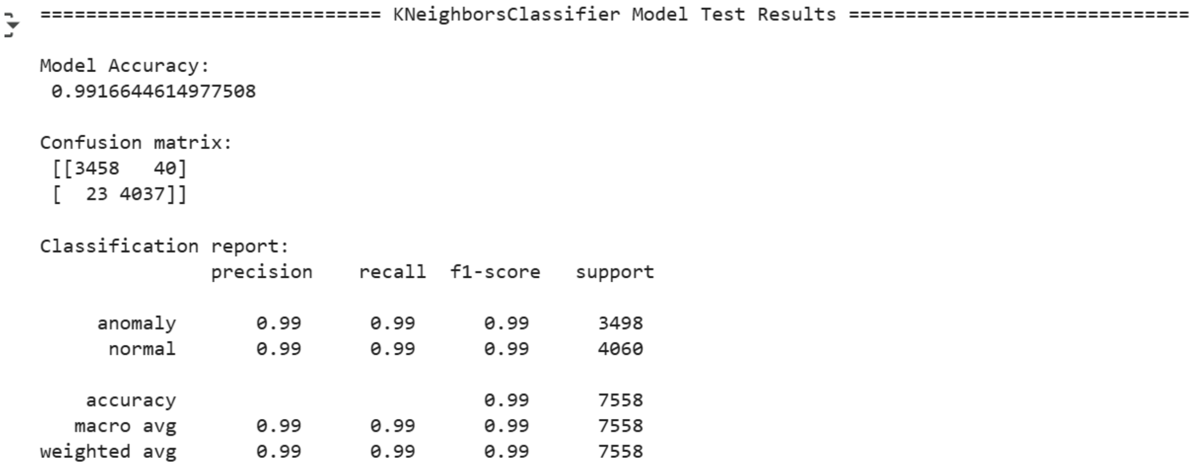
```
============================== Naive Baye Classifier Model Test Results ==============================

Model Accuracy:
 0.906721354855782

Confusion matrix:
 [[2981  517]
 [ 188 3872]]

Classification report:
               precision    recall  f1-score   support

     anomaly       0.94      0.85      0.89      3498
      normal       0.88      0.95      0.92      4060

    accuracy                           0.91      7558
   macro avg       0.91      0.90      0.91      7558
weighted avg       0.91      0.91      0.91      7558
```

Fig 1 Naive Baye Classifier Model Test Results



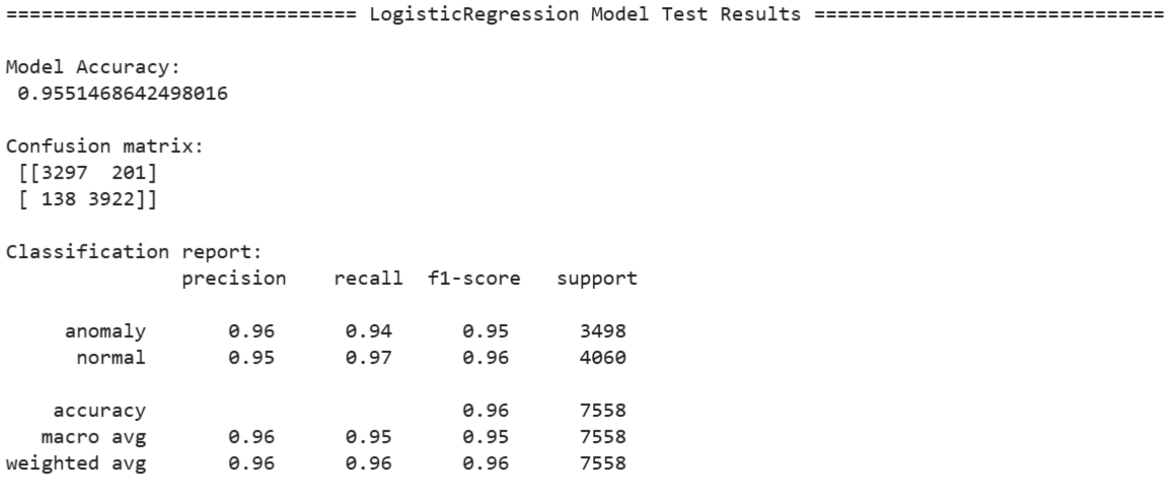
```
============================== Decision Tree Classifier Model Test Results ==============================

Model Accuracy:
 0.9947075946017465

Confusion matrix:
 [[3483   15]
 [  25 4035]]

Classification report:
               precision    recall  f1-score   support

     anomaly       0.99      1.00      0.99      3498
      normal       1.00      0.99      1.00      4060

    accuracy                           0.99      7558
   macro avg       0.99      0.99      0.99      7558
weighted avg       0.99      0.99      0.99      7558
```

Fig 2 Decision Tree Classifier Model



```
============================= KNeighborsClassifier Model Test Results =============================

Model Accuracy:
 0.9916644614977508

Confusion matrix:
 [[3458   40]
 [  23 4037]]

Classification report:
               precision    recall  f1-score   support

     anomaly       0.99      0.99      0.99      3498
      normal       0.99      0.99      0.99      4060

    accuracy                           0.99      7558
   macro avg       0.99      0.99      0.99      7558
weighted avg       0.99      0.99      0.99      7558
```

Fig 3 KNeighborsClassifier Model



```
============================= LogisticRegression Model Test Results =============================

Model Accuracy:
 0.9551468642498016

Confusion matrix:
 [[3297  201]
 [ 138 3922]]

Classification report:
               precision    recall  f1-score   support

     anomaly       0.96      0.94      0.95      3498
      normal       0.95      0.97      0.96      4060

    accuracy                           0.96      7558
   macro avg       0.96      0.95      0.95      7558
weighted avg       0.96      0.96      0.96      7558
```

Fig4 LogisticRegression Model

VI. CONCLUSION

In this project, we addressed limitations associated with the LightGBM (Light Gradient Boosting Machine) algorithm, particularly its assumption of strong feature independence among the attributes in the dataset. While LightGBM is efficient and powerful for structured data, its core methodology treats each feature independently, which may not be ideal for intrusion detection tasks where feature interdependencies can play a significant role.

To overcome this limitation, we proposed a new approach that approximates the interactions between features using conditional probabilities. This method provides a more nuanced view of how features relate to each other and contributes to better decision-making by the learning algorithm.

The performance of the proposed algorithm was evaluated against these classifiers using standard metrics such as accuracy, precision, recall, and F1-score. The results demonstrated that the proposed approach performed better or comparably to existing classifiers, especially in detecting complex intrusion patterns.

Overall, the application of conditional probability-based feature interaction modeling, combined with effective feature selection, led to a more robust and accurate intrusion detection system. This approach not only enhances model performance but also provides a foundation for future research into more advanced hybrid machine learning algorithms for cybersecurity applications.

REFERENCES

1. Gnes Kayack, H., Nur Zincir- Heywood, A., and Heywood, M. I.: Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets. Third Annual Conference on Privacy, Security and Trust, (2005).
2. Balakrishnan, S., and Kannan, V.K.:Intrusion Detection System Using Feature Selection and Classification Technique. International Journal of Computer Science and Application (IJCSA), vol. 3, issue 4, (2014).
3. Vinchurkar, D. P., and Reshamwala, A.: A Review of Intrusion Detection System Using NN and Machine Learning Technique.Debar, H, Dacier, M., and Wespi, A, A Revised taxonomy for intrusion detection systems, Annales des Telecommunications Vol. 55, No.7–8, 361–378, 2000.
4. Sommer, R., and Paxson, V.:Outside the Closed World: On Using Machine Learning For Network Intrusion Detection. IEEE Symposium on Security and Privacy, pp. 305-316, (2010).
5. J Shun and HA Malki, A neural network-based network intrusion detection system. Proc. Fourth IEEE Int Conf Nat Comput 5, 242–246 (2008).ICNC’08.
6. MM Kabir and MM Islam, K Murase, Using a Neural Network, a new wrapper feature selection approach. NeuroComputing 73, 3273–3283 (2010). Elsevier.
7. G Xiantai, J Weidong, Z Dao, In Metropolitan Area Networks, a Multi-Agent Scheme for Identification and Containment. J. Electron. (China) 23(2),259–265 (2006).
8. F Amiri, MMR Yousefi, C Lucas, A Shakery, N Yazdani, For intrusion detection systems, feature selection is based on shared knowledge. J. Network Comput. Appl 34, 1184–1199 (2011).
9. H Liu, L Yu, Towards combining grouping and clustering feature collection algorithms. IEEE Trans. Knowl. Data Eng. 17, 491–502 (2005).
10. Scarfone, K., and Mell, P.:Guide to Intrusion Detection and Prevention Systems (IDPS). National Institute Of Standards and Technology. Special Publication February-2007.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com