INTRUSION DETECTION IN CYBER SECURITY: MACHINE LEARNING CLASSIFIER PERFORMANCE EVALUATION

Dr SK Mahaboob Basha[1*], P. Tejaswi[2], K. Pranith Reddy[2], B. Manoj Reddy[2], G. Deepak[2]

[1,2]Department of Computer Science and Engineering(Cyber Security), Sree Dattha Group of Institutions, Sheriguda, Hyderabad, Telangana.

*Corresponding author: Dr. SK Mahaboob Basha

ABSTRACT

Industry 4.0 refers to the fourth industrial revolution, characterized by the use of smart technologies, data exchange, and automation in manufacturing. The Industrial Internet of Things (IIoT) is a crucial component of Industry 4.0, where devices and machines are connected to the internet for enhanced communication and data exchange. As industries embrace these technological advancements, there is a corresponding increase in the potential vulnerabilities to cyber threats and intrusions. Securing the IIoT environment becomes paramount to ensure the smooth and secure operation of industrial processes. The need for an innovative approach to combat intrusion in IIoT arises from the critical nature of industrial processes. Cyber-attacks on IIoT systems can lead to disruptions in production, compromise of sensitive data, and even pose threats to human safety. Traditional security systems often rely on static and rule-based approaches, which may not effectively adapt to the dynamic and sophisticated nature of modern cyber threats. Legacy security measures might lack the agility and intelligence needed to counter advanced persistent threats and zero-day vulnerabilities. Therefore, there is a need to move beyond traditional security paradigms and embrace innovative approaches that leverage cutting-edge technologies. The problem at hand is the susceptibility of IIoT systems to cyber threats and intrusions. These threats may include unauthorized access, data breaches, malware attacks, and other forms of cyber-attacks that can compromise the integrity, confidentiality, and availability of industrial data and processes. The challenge is to develop a robust and proactive model that can detect, prevent, and combat intrusions in real-time within the context of Industry 4.0. Therefore, this research aims to implement intrusion combat model for IIoT., where the significance of the proposed innovative approach lies in its potential to enhance the security posture of IIoT systems in the context of Industry 4.0. By developing a robust intrusion combat model, the industrial sector can mitigate the risks associated with cyber threats, ensuring the continuity of operations, safeguarding sensitive data, and protecting the overall integrity of the industrial processes. The significance extends to the broader implications for the adoption and successful implementation of Industry 4.0 technologies, as a secure IIoT ecosystem is fundamental to realizing the full benefits of the fourth industrial revolution.

Keywords: Intrusion Detection System (IDS), Cyber Security, Zero-Day Attacks, Proactive Threat Mitigation, Cyber-Attack Prevention

1. INTRODUCTION

The concept of Industry 4.0, heralding the fourth industrial revolution, traces its roots to the early 2010s when discussions around the integration of digital technologies into manufacturing gained traction. This period marked a significant shift from traditional manufacturing methods to smart, interconnected systems characterized by automation, data exchange, and artificial intelligence. The term "Industry 4.0" itself was coined in Germany, where the government initiated the "Industrie 4.0" project to promote digital transformation in manufacturing. The evolution of Industry 4.0 is closely intertwined with advancements in technology, particularly in areas such as the Internet of Things (IoT), cloud computing, and big data analytics.

Fig. 1:Intrusion Detection in Cyber Security.

These technological innovations laid the foundation for the Industrial Internet of Things (IIoT), which represents a crucial component of Industry 4.0. The IIoT enables machines, devices, and sensors to communicate and share data over the internet, facilitating real-time monitoring, analysis, and control of industrial processes.Over the years, Industry 4.0 has witnessed rapid adoption across various sectors, including automotive, aerospace, electronics, and healthcare. Organizations worldwide have embraced smart manufacturing practices to improve efficiency, productivity, and flexibility while reducing costs and time to market. However, alongside these benefits come challenges, particularly in terms of cybersecurity.

2. LITERATURE SURVEY

An Anomaly Detection System (ADS) is considered an essential security management system, functioning as a sniffer and decision driver for routing traffic and spotting suspicious activities. It operates as a packet capture and decoding engine for ensuring security and recognizing anomalous behavior. Both visible and invisible (zero-day) threats can be tracked, with a focus on creating patterns from standard data and treating any variance from it as an intrusion. For instance, research by [1, 2] focused on using Particle Swarm Optimization (PSO) techniques to optimize the performance of the One-Class Support Vector Machine (OCSVM) method by harvesting Modbus/TCP message network streams for testing and verifying the system. Another study by [3] built an IDS/ADS based on this design, which learned from offline data from a SCADA setting using network traces.In another approach, [4] constructed an IDS centered on the Modbus/TCP protocol setting using a K-NN classifier. While these mechanisms performed well in certain cases, they were designed for particular configurations with a strong false positive rate (FPR). Similarly, [5] proposed an improved intrusion detection system (IDS) for matching the diverse structures of SCADA schemes using diverse OCSVM frameworks to efficiently identify multiple assaults. However, this approach consumed a large amount of computational power and had a high false warning rate for identification.

In [6], authors suggested an ADS for detecting Modbus/TCP protocol-infiltrated assaults by using SCADA mechanisms to obtain different aspects of contact events and using an SVM algorithm to identify attacks. However, the detection method was ineffective in detecting irregular behaviors.To address factors associated with the OCSVM's ability to successfully track network attacks, [7] merged the OCSVM method with the recurrent -means clustering algorithm. Additionally, [8] proposed a critical infrastructure intrusion detection system centered on an artificial neural network (ANN) method. They trained a multiperceptron ANN to identify anomalous network activity using fault back-propagation and Levenberg-Marquard features. Using a virtual network, [9] employed an ANN to detect DoS/DDoS attacks in IoTs, while [10] proposed a decentralized IDS based on artificial immunity for IoT devices. Another group of researchers in [11] projected a Possibility Risk Identification-centered Intrusion Detection System (PRI-IDS) method for detecting replay attacks by inspecting Modbus TCP/IP protocol network traffic. However, these schemes had a high rate of false alarms and had trouble identifying certain new attacks.

In a different approach, [12] created a learning firewall that received tagged samples and automatically configured itself by writing conservative preventive rules to avoid false alerts. They introduced a novel classifier family called classifiers, which focused on zero false positives as the decision-making criterion. This classifier, based on CART, was used to create a firewall for a Power Grid Monitoring System and was tested on the KDD CUP'99 dataset, showing the efficacy of the strategy.IDSs have been analyzed utilizing subsurface networks for identifying irregular findings from host and network-based systems by several researchers. Deep learning techniques have also been explored, with some researchers proving the effectiveness of shallow and deep networks. However, the swiftness at which system signals are converted into massive datasets poses a significant obstacle to IDS architectures' ability to analyze the subsequent large amounts of data for processing. For intrusion detection, various machine learning techniques such as SVM, ant colony optimization, decision trees, and neural networks have been proposed. These methods aim to address the dynamic patterns of attacks and the need for effective cyberattack classification and prediction. Researchers have also suggested innovative feature selection techniques and hybrid models to enhance IDS performance, particularly in detecting IoT-related threats. From the existing related work, it is evident that DL algorithms can significantly improve the efficiency of IDS for IIoT by achieving high prediction performance while maintaining a low false alarm rate. Motivating the use of DL models with hybrid rule-based techniques for automatic feature selection and sensing anomaly trends in data as suspect vectors using data transmission depth coverage. The proposed DFFNN based with hybrid rule-based feature selection model contains a rule-based model using a genetic search engine to select relevant features and the DAE-DFFNN algorithm to classify IIoT network by classifying the constraint values of the DAE. This model can find a good approximation for communication networks and transform high-dimensional data to low-dimensional data using the decreased layer of the DAE-DFFNN model.

3. PROPOSED METHODOLOGY

The Python-based application is designed for intrusion detection in Industrial Internet of Things (IIoT) environments and features a user-friendly GUI built with Tkinter. Users can upload network traffic datasets, which are then preprocessed to handle missing values and encode categorical variables. The data is split into training and testing sets for model development. The application supports two machine learning algorithms—logistic regression and random forest—for training intrusion detection models. Upon training, performance metrics such as precision, recall, F1-score, and accuracy are computed and displayed within the GUI, enabling users to assess model effectiveness. Additionally, the application allows users to upload new data for intrusion prediction using the trained models, with results shown in the interface. A bar graph visually compares the performance of both classifiers, assisting users in selecting the most suitable algorithm for their intrusion detection needs.

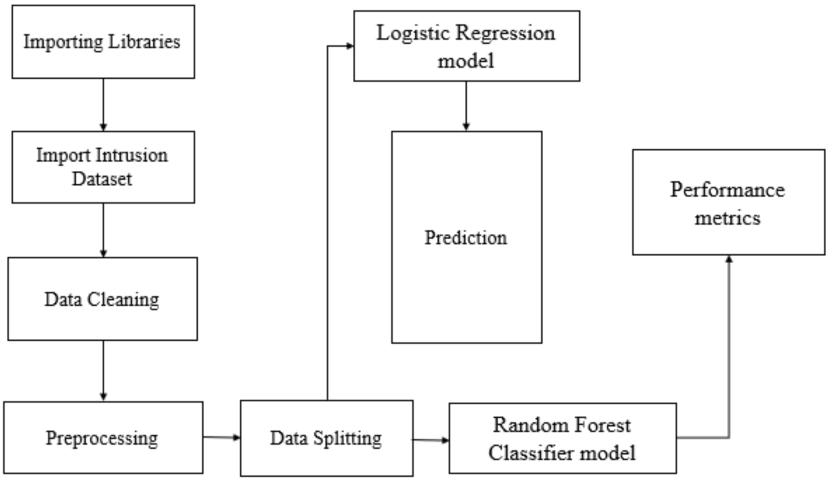


Fig. 2: Block diagram of proposed diagram.

This Tkinter-based Python application is designed for intrusion detection in IIoT environments using Decision Tree and Random Forest classifiers. It offers a user-friendly GUI that enables users to upload datasets, preprocess data (handling missing values and encoding), and transparently split it into training and testing sets. The application supports logistic regression, decision tree, and random forest algorithms, allowing users to compare model performance through key metrics like accuracy, precision, recall, F1-score, confusion matrix, and ROC curves. Predictions can be made on new data, and a comparison graph visually illustrates the performance differences between classifiers.

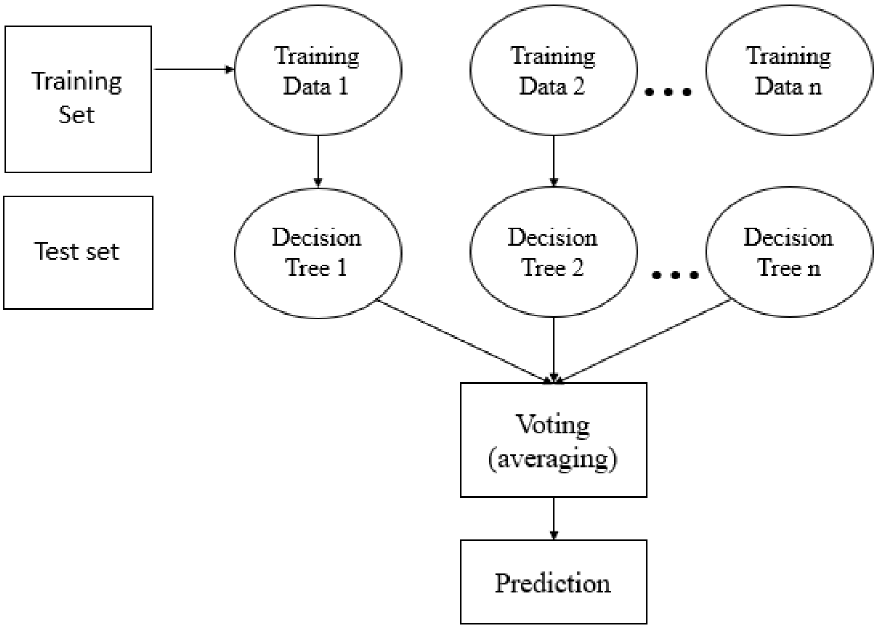


Fig. 3: Random Forest Algorithm Working.

The Random Forest algorithm, based on ensemble learning and bagging, enhances prediction accuracy by combining outputs from multiple decision trees. Key advantages include fast training, resilience to missing data, reduced overfitting, parallelization, and support for high-dimensional data. The modular structure ensures adaptability and scalability, making it ideal for both practical deployment and educational purposes in machine learning and cybersecurity.

4. RESULTS AND DISCUSSION

Figure 4: The image displays the Intrusion Combat Model for Industrial IoT (Internet of Things). This model outlines a framework or approach designed to identify and mitigate security threats in industrial IoT systems. It include components such as data collection, anomaly detection, threat analysis, and response mechanisms tailored specifically for industrial environments. Figure 5: This figure showcases the uploaded dataset and the preprocessing steps applied to it. It may include processes such as data cleaning, normalization, feature selection, or transformation to prepare the data for analysis and model training. The preprocessing steps ensure that the dataset is suitable for input into machine learning algorithms.

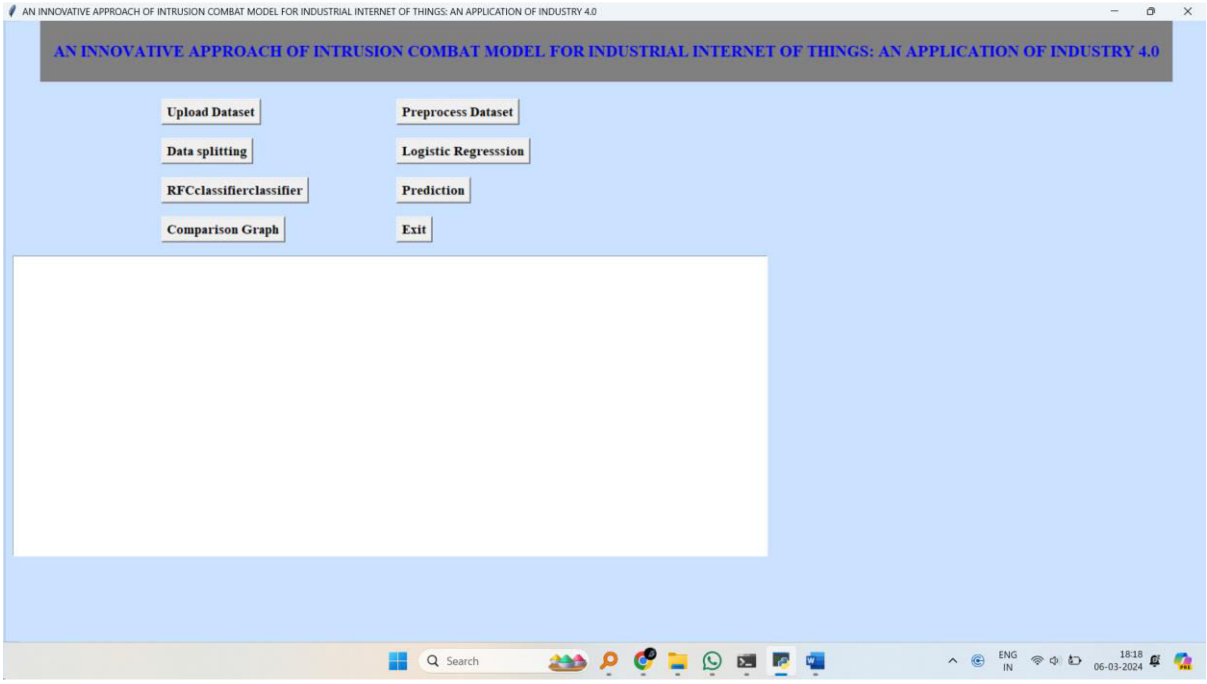
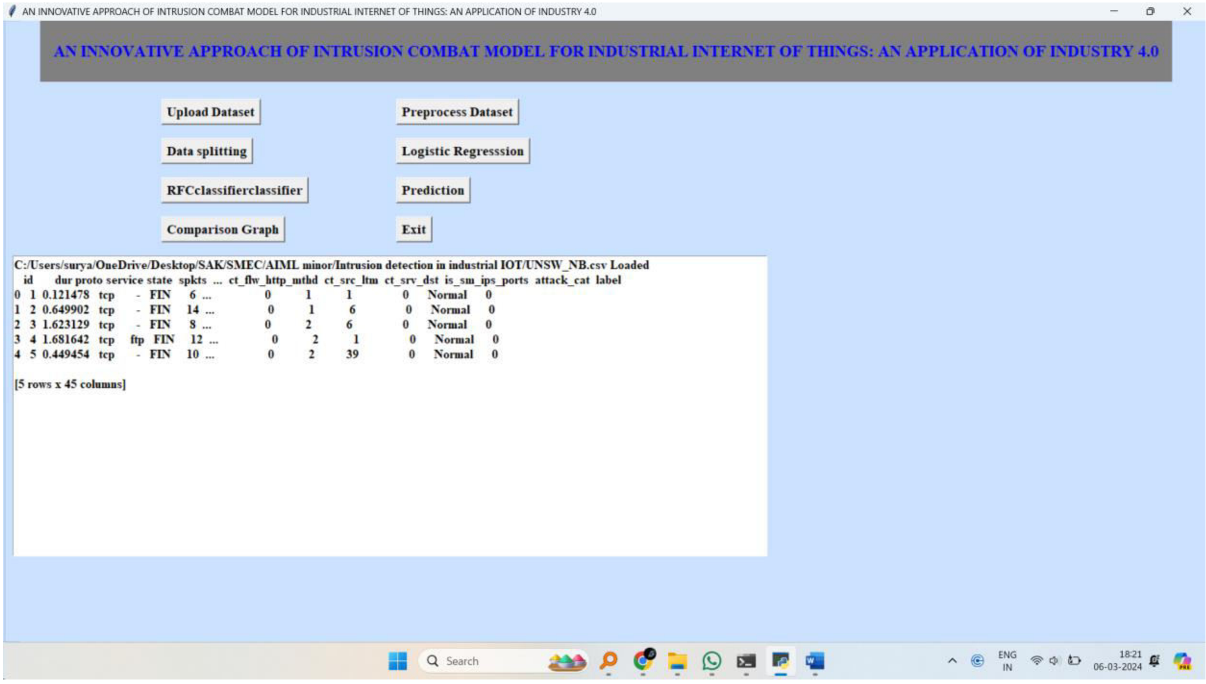

AN INNOVATIVE APPROACH OF INTRUSION COMBAT MODEL FOR INDUSTRIAL INTERNET OF THINGS: AN APPLICATION OF INDUSTRY 4.0
AN INNOVATIVE APPROACH OF INTRUSION COMBAT MODEL FOR INDUSTRIAL INTERNET OF THINGS: AN APPLICATION OF INDUSTRY 4.0
Upload Dataset
Preprocess Dataset
Data splitting
Logistic Regresssion
RFCclassifierclassifier
Prediction
Comparison Graph
Exit
Search
ENG
IN
18:18
06-03-2024

Figure 4: Presents The Intrusion Combat Model For Industrial IOT.


AN INNOVATIVE APPROACH OF INTRUSION COMBAT MODEL FOR INDUSTRIAL INTERNET OF THINGS: AN APPLICATION OF INDUSTRY 4.0
AN INNOVATIVE APPROACH OF INTRUSION COMBAT MODEL FOR INDUSTRIAL INTERNET OF THINGS: AN APPLICATION OF INDUSTRY 4.0
Upload Dataset
Preprocess Dataset
Data splitting
Logistic Regresssion
RFCclassifierclassifier
Prediction
Comparison Graph
Exit
Search
ENG
IN
18:21
06-03-2024

```
C:/Users/surya/OneDrive/Desktop/SAK/SMEC/AIML minor/Intrusion detection in industrial IOT/UNSW_NB.csv Loaded
  id      dur proto service state  spkts  ...  ct_flw_http_mthd  ct_src_ltm  ct_srv_dst  is_sm_ips_ports  attack_cat  label
0  1  0.121478   tcp       -   FIN      6  ...                 0           1           1                0      Normal      0
1  2  0.649902   tcp       -   FIN     14  ...                 0           1           6                0      Normal      0
2  3  1.623129   tcp       -   FIN      8  ...                 0           2           6                0      Normal      0
3  4  1.681642   tcp     ftp   FIN     12  ...                 0           2           1                0      Normal      0
4  5  0.449454   tcp       -   FIN     10  ...                 0           2          39                0      Normal      0

[5 rows x 45 columns]
```

Figure 5: Presents the uploaded dataset and its preprocessing.

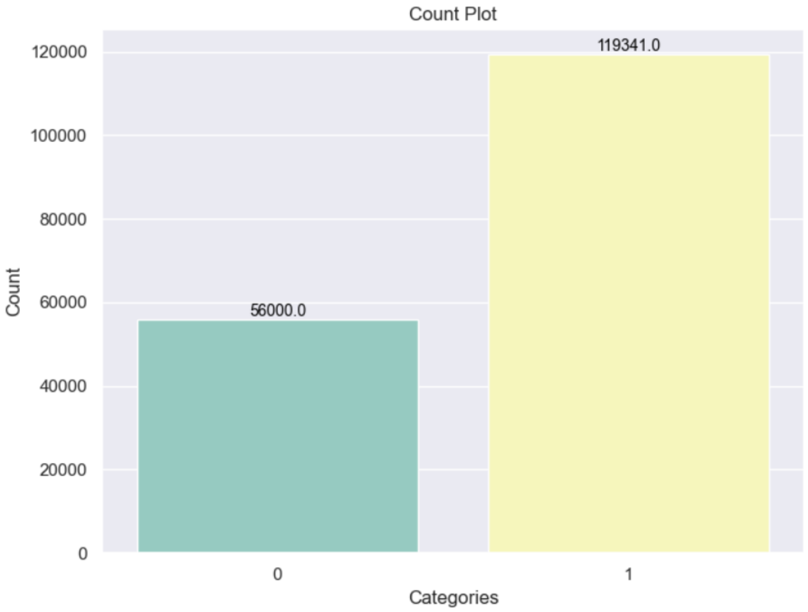


Figure 6: Presents the count plot of each Label in Categories.

Figure 6: This plot visualizes the count of each label or category in the dataset. It provides insights into the distribution of different classes within the data, which is crucial for understanding the class imbalance and selecting appropriate evaluation metrics for the classification models. Figure 7: This figure presents the confusion matrices of two different models: logistic regression and random forest. A confusion matrix is a table that summarizes the performance of a classification model by comparing predicted labels with actual labels. It provides information about true positives, true negatives, false positives, and false negatives, allowing for the assessment of model accuracy and error rates.

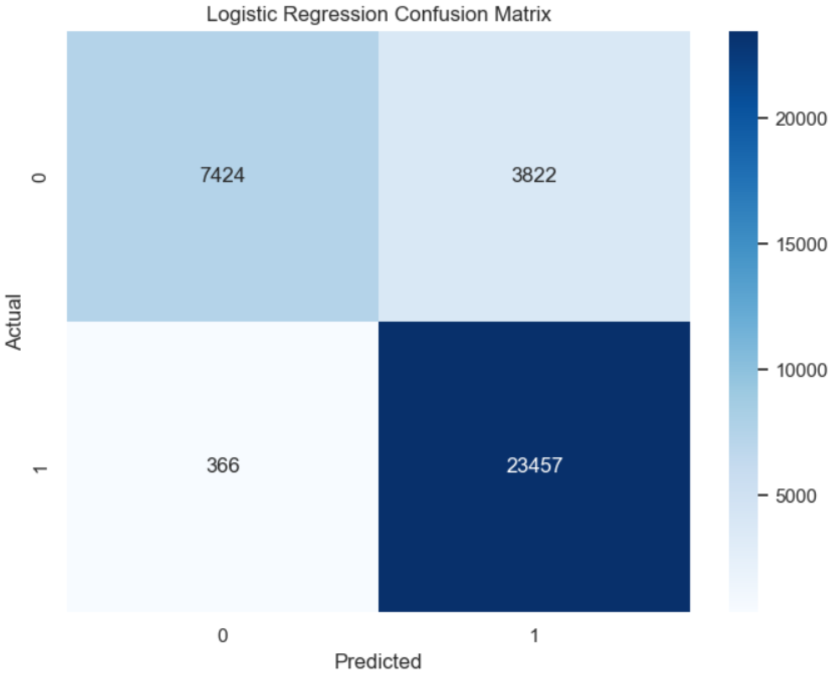


Figure 7: Presents the Confusion Matrix of Logistic Regression model.

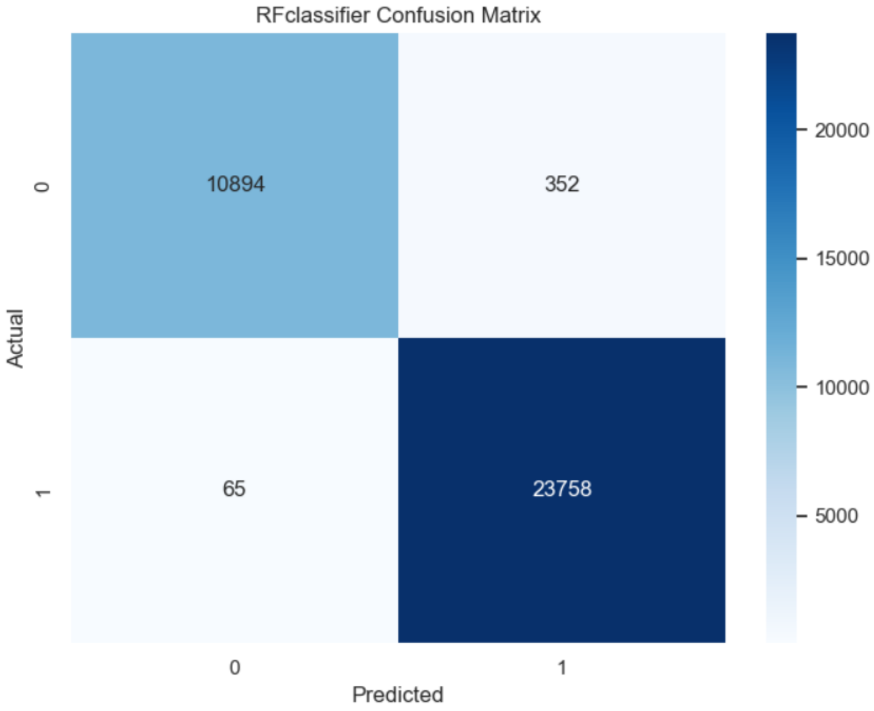
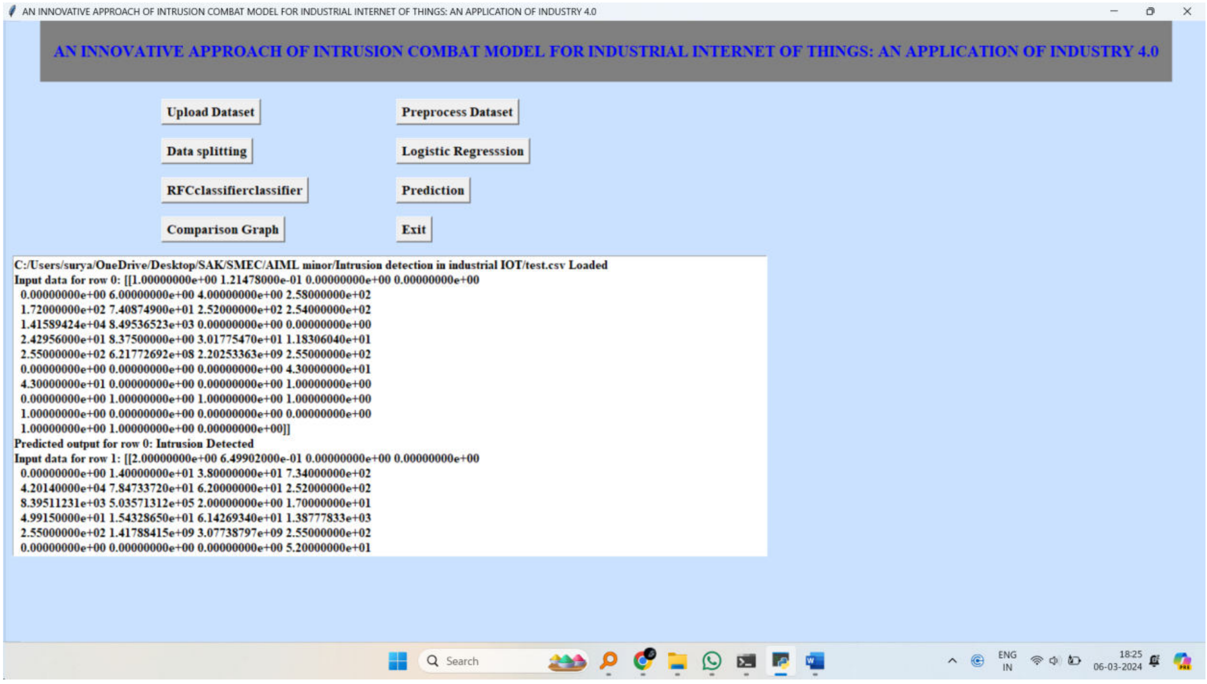


Figure 8: Presents the Confusion Matrix of Random Forest model.


AN INNOVATIVE APPROACH OF INTRUSION COMBAT MODEL FOR INDUSTRIAL INTERNET OF THINGS: AN APPLICATION OF INDUSTRY 4.0
AN INNOVATIVE APPROACH OF INTRUSION COMBAT MODEL FOR INDUSTRIAL INTERNET OF THINGS: AN APPLICATION OF INDUSTRY 4.0
Upload Dataset
Preprocess Dataset
Data splitting
Logistic Regresssion
RFCclassifierclassifier
Prediction
Comparison Graph
Exit
Search
ENG
IN
18:25
06-03-2024

```
C:/Users/surya/OneDrive/Desktop/SAK/SMEC/AIML minor/Intrusion detection in industrial IOT/test.csv Loaded
Input data for row 0: [[1.00000000e+00 1.21478000e-01 0.00000000e+00 0.00000000e+00
  0.00000000e+00 6.00000000e+00 4.00000000e+00 2.58000000e+02
  1.72000000e+02 7.40874900e+01 2.52000000e+02 2.54000000e+02
  1.41589424e+04 8.49536523e+03 0.00000000e+00 0.00000000e+00
  2.42956000e+01 8.37500000e+00 3.01775470e+01 1.18306040e+01
  2.55000000e+02 6.21772692e+08 2.20253363e+09 2.55000000e+02
  0.00000000e+00 0.00000000e+00 0.00000000e+00 4.30000000e+01
  4.30000000e+01 0.00000000e+00 0.00000000e+00 1.00000000e+00
  0.00000000e+00 1.00000000e+00 1.00000000e+00 1.00000000e+00
  1.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
  1.00000000e+00 1.00000000e+00 0.00000000e+00]]
Predicted output for row 0: Intrusion Detected
Input data for row 1: [[2.00000000e+00 6.49902000e-01 0.00000000e+00 0.00000000e+00
  0.00000000e+00 1.40000000e+01 3.80000000e+01 7.34000000e+02
  4.20140000e+04 7.84733720e+01 6.20000000e+01 2.52000000e+02
  8.39511231e+03 5.03571312e+05 2.00000000e+00 1.70000000e+01
  4.99150000e+01 1.54328650e+01 6.14269340e+01 1.38777833e+03
  2.55000000e+02 1.41788415e+09 3.07738797e+09 2.55000000e+02
  0.00000000e+00 0.00000000e+00 0.00000000e+00 5.20000000e+01
```

Figure 9: Presets Proposed Model Prediction on test data.

Figure 9: The image illustrates the predictions made by the proposed model on test data. It includes a comparison between the actual labels and the predicted labels for each instance in the test dataset. This visualization helps evaluate the model's performance in accurately classifying data points. Figure 10: This figure displays a comparison of the performance metrics of different models. It include

metrics such as accuracy, precision, recall, and F1-score for each model, allowing for a comprehensive assessment of their predictive capabilities. This comparison aids in identifying the most effective model for the given dataset and task.

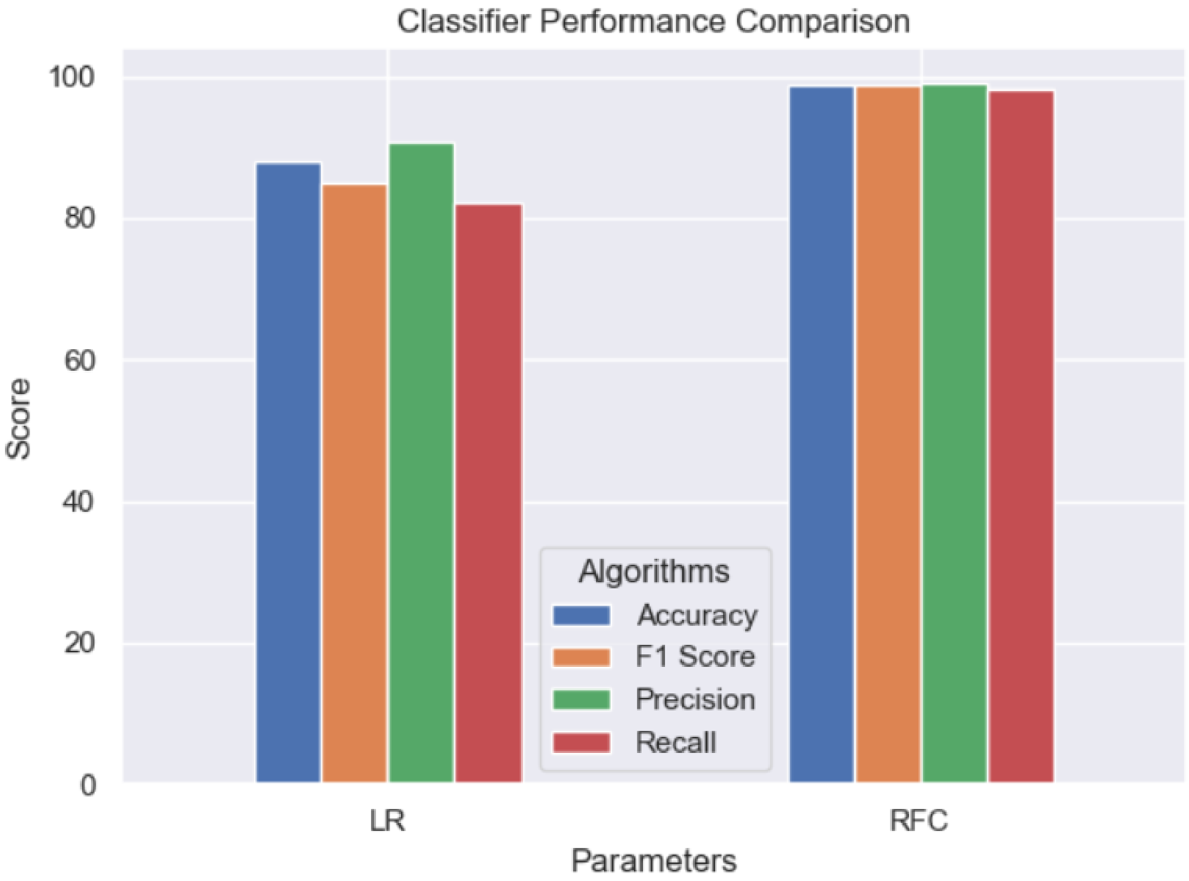


Figure 10: Displays the Each Model Performance Comparison.

Table 1: Performance Metrics of all models.

Algorithm	Precision	Recall	F1-Score	Accuracy
Logistic Regression	90.65	82.24	84.90	88.06
Random Forest	98.97	98.30	98.63	98.81

Precision: Precision measures the accuracy of positive predictions made by the model. For logistic regression, the precision is 90.65%, indicating that out of all the instances predicted as positive, 90.65% were actually positive. Conversely, random forest achieved a higher precision of 98.97%, indicating a higher accuracy in positive predictions.

Recall: Recall measures the ability of the model to identify all positive instances. Logistic regression achieved a recall of 82.24%, meaning that it correctly identified 82.24% of all actual positive instances. On the other hand, random forest had a slightly higher recall of 98.30%, indicating its superior ability to capture positive instances.F1-Score: The F1-Score is the harmonic mean of precision and recall, providing a balance between the two metrics. For logistic regression, the F1-

Score is 84.90%, reflecting the balance between precision and recall in its predictions. Random forest achieved a higher F1-Score of 98.63%, indicating a stronger balance between precision and recall.Accuracy: Accuracy measures the overall correctness of the model's predictions. Logistic regression achieved an accuracy of 88.06%, meaning that it correctly classified 88.06% of all instances. In comparison, random forest achieved a higher accuracy of 98.81%, indicating its overall superior performance in classification tasks.

5. CONCLUSION

In conclusion, the development of an innovative intrusion combat model tailored for Industrial Internet of Things (IIoT) systems within the framework of Industry 4.0 represents a significant step towards enhancing cybersecurity in industrial environments. This research addresses the critical need to safeguard industrial processes, data, and infrastructure from cyber threats, which have become increasingly prevalent and sophisticated in today's interconnected world.

By leveraging cutting-edge technologies such as artificial intelligence, machine learning, and anomaly detection, the proposed model offers a proactive and adaptive approach to combatting intrusions in real-time. It moves beyond traditional security paradigms, which often rely on static rules and signatures, and embraces dynamic and intelligent methodologies to detect, prevent, and mitigate cyber-attacks. The significance of this research extends beyond individual industries to the broader implications for the adoption and successful implementation of Industry 4.0 technologies. A secure IIoT ecosystem is fundamental to realizing the full benefits of the fourth industrial revolution, including improved efficiency, productivity, and flexibility. By enhancing the security posture of IIoT environments, the intrusion combat model enables organizations to safeguard critical assets, protect sensitive data, and ensure the continuity of operations.

REFERENCES

[1]. N. Moustafa and J. Slay, “The evaluation of Network Anomaly Detection Systems: statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set,” Information Security Journal: A Global Perspective, vol. 25, no. 1-3, pp. 18–31, 2016.

[2]. W. Shang, P. Zeng, M. Wan, L. Li, and P. An, “Intrusion detection algorithm based on OCSVM in industrial control system,” Security and Communication Networks, vol. 9, no. 10, p. 1049, 2016.

[3]. L. A. Maglaras and J. Jiang, “Intrusion detection in SCADA systems using machine learning techniques,” in 2014 Science and Information Conference, pp. 626–631, London, UK, August 2014.

[4]. P. Silva and M. Schukat, “On the use of k-nn in intrusion detection for industrial control systems,” in Proceedings of The IT&T 13th International Conference on Information Technology and Telecommunication, pp. 103–106, Dublin, Ireland, August 2014.

[5]. B. Stewart, L. Rosa, L. A. Maglaras et al., “A novel intrusion detection mechanism for scada systems which automatically adapts to network topology changes,” EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, vol. 4, no. 10, 2017.

[6]. W. Shang, J. Cui, M. Wan, P. An, and P. Zeng, “Modbus communication behavior modeling and SVM intrusion detection method,” in Proceedings of the 6th International Conference on Communication and Network Security, pp. 80–85, November 2016.

[7]. L. A. Maglaras and J. Jiang, “Ocsvm model combined with k-means recursive clustering for intrusion detection in scada systems,” in 10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, pp. 133-134, Rhodes, Greece, August 2014.

[8]. O. Linda, T. Vollmer, and M. Manic, “Neural network-based intrusion detection system for critical infrastructures,” in 2009 International Joint Conference on Neural Networks, pp. 1827–1834, Atlanta, GA, USA, June 2009.

[9]. E. Hodo, X. Bellekens, A. Hamilton et al., “Threat analysis of IoT networks using artificial neural network intrusion detection system,” in 2016 International Symposium on Networks, Computers and Communications (ISNCC), pp. 1–6, Yasmine Hammamet, Tunisia, May 2016.

[10]. R. Chen, C. M. Liu, and C. Chen, “An artificial immune-based distributed intrusion detection model for the internet of things,” in Advanced materials research, pp. 165–168, Trans Tech Publications Ltd, 2012.

[11]. T. Marsden, N. Moustafa, E. Sitnikova, and G. Creech, “Probability risk identification based intrusion detection system for SCADA systems,” in Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pp. 353–363, Springer, Cham, Switzerland, 2017.

[12]. M. S. Haghighi, F. Farivar, and A. Jolfaei, “A machine learning-based approach to build zero false-positive IPSs for industrial IoT and CPS with a case study on power grids security,” IEEE Transactions on Industry Applications, pp. 1–9, 2020.