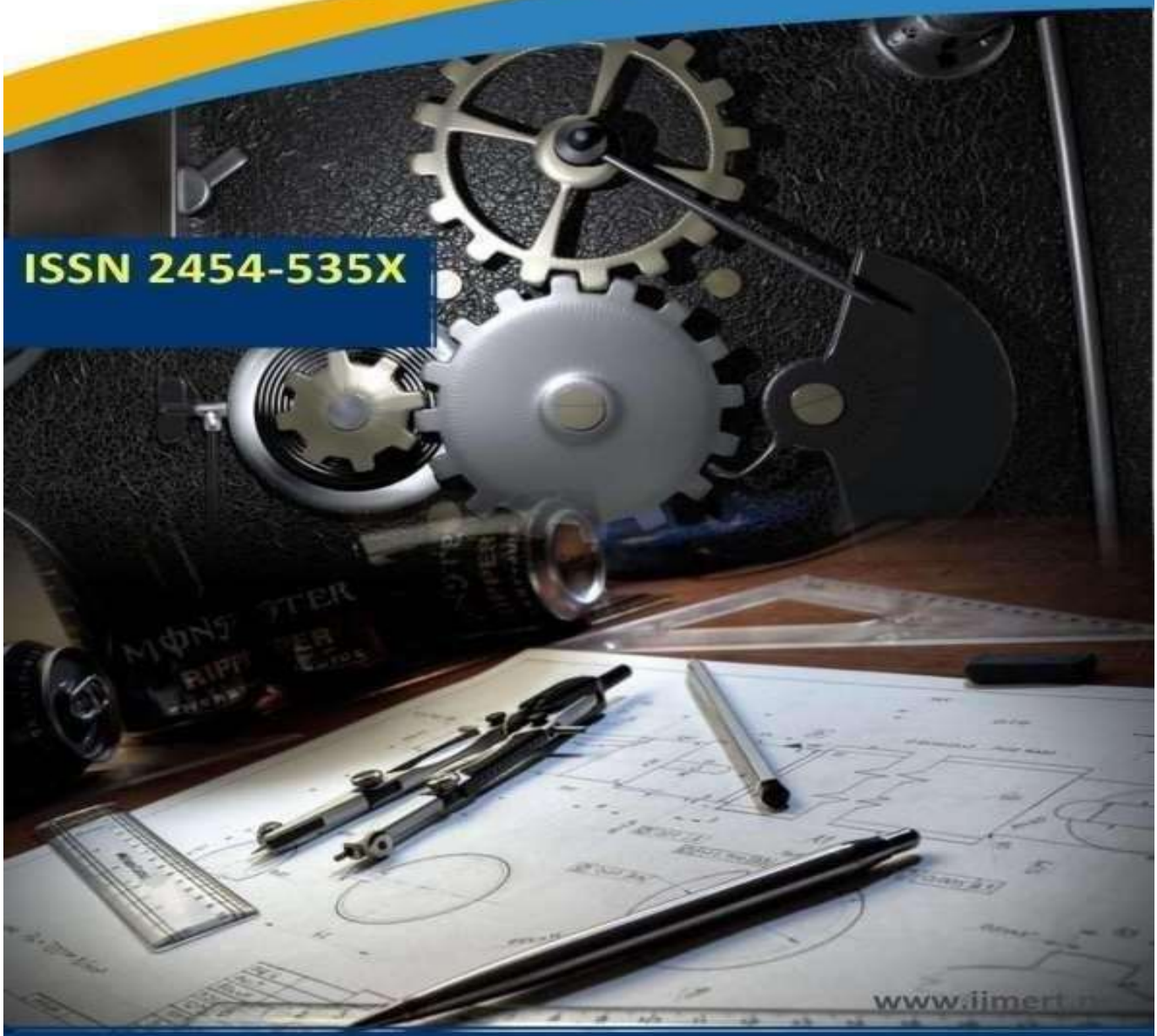
# International Journal of Mechanical Engineering Research and Technology

ISSN 2454-535X



Email ID: info.ijmert@gmail.com or editor@ijmert.net

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

# MACHINE LEARNING FOR IOT DEVICE ANOMALY DETECTION ATTACK CLASSIFICATION

*G VISWANATH[1], VELIGARAM MADHAVI[2], B AJITH KUMAR[3]*

*[1]Associate Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: viswag111@gmail.com, ORCID: https://orcid.org/0009-0001-7822-4739*
*[2]P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: madhaviveligaram@gmail.com*
*[3]Assistant Professor, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur, Email: ajithkumaryadav34@gmail.com*

**Abstract:** The theoretical underscores the security risks of Internet of Things (IoT) gadgets from programmers and assailants. IoT gadgets are defenseless against inconsistency assaults because of their interconnectedness. Support Vector Machine (SVM) and Random Forest (RF), stacking classifier, and voting classifier are utilized to recognize odd attacks in IoT gadgets in the Task. Every strategy is picked for its discovery and component determination abilities. The review explores different avenues regarding the arff NSL-KDD dataset. The suggested calculations RF and stacking classifier have great accuracy. Center around misleading positive rates shows a low rate in all cases. Featuring the technique's promising outcomes, quite Random Forests' preferable exactness over past writing. The stacking classifier and Random Forest exhibit promising exactness, review, and accuracy for recognizing and relieving IoT abnormal dangers. Troupe approaches like Voting Classifier (RF + AB) and Stacking Classifier (RF + MLP with LightGBM) consolidate various model expectations to make a more vigorous and precise forecast. Voting Classifier had 100 percent accuracy, Stacking Classifier 100 percent accuracy, and we made the front end involving flagon for client testing and IoT anomaly detection with client verification.[16]

***Index terms*** *– IOT devices, Support Vector Machine (SVM) and Random Forest (RF).*

## 1. INTRODUCTION

The Internet of Things (IoT) expands web association past computers, workstations, cell phones, and tablets to non-web-viable actual gadgets and regular things. Sensor information from IoT gadgets might be utilized by people, associations, and others. Three key gatherings are purchaser, endeavor, and modern. The market has a few IoT gadgets, and clients pick them in light of their specs, conditions, and expenses. IoT contraptions incorporate toaster ovens and coolers. Margaret Lee (2020) [1] anticipated 64 billion IoT gadgets by 2025.

Inconsistencies are information designs that make heads spin. It additionally alludes to IoT capability exceptions, special cases, irregularities, amazements, and anomalies. Peculiarity location assists IoT programming applications with identifying unusual way of behaving and examine patterns. Strange information could show mechanical deficiencies or

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

client conduct changes. The review [2] included interruption counteraction, extortion location, and information spillage as peculiarities. IoT applications including brilliant urban communities, network security, and businesses use abnormality recognition.

There is little exploration on ML irregularity identification in IoT gadgets [2, 7, 8, 9, 10, 11, 12]. Many examinations know nothing about different elements, for example, gadget security, which is urgent to picking the right hardware. Because of their web network, IoT gadgets are progressively powerless against programmers. The review report [3] featured different attacks, including Western Advanced's My Book Live assault. My Book Live gadgets are private mists. Because of a framework issue, programmers might reset gadgets without passwords and delete every one of their information.

In 2018, the quantity of hacked Internet of Things (IoT) gadgets and cryptographic money networks in Japan generally expanded contrasted with earlier years [4]. Along these lines, oddity recognition should be carried out when it can lessen interloper or programmer hurt. In information mining and ML, peculiarity examination is pivotal, as per Xu et al. (2019) [5]. It searches for information regions that leave from expectations.[18]

## 2. LITERATURE SURVEY

One of the quickest developing innovations is IoT. It lets billions of smart gadgets or "Things" use sensors to gather information about themselves and their current circumstance. They may then impart information to approved gatherings to direct and screen modern administrations or increment business capabilities. Today, the Web of Things goes up against more noteworthy security weaknesses than any time in recent memory. [2, 7, 8, 9, 10, 11, 12] A significant specialized advance in machine learning (ML) has opened up new examination choices to deal with IoT issues. Be that as it may, ML can recognize chances and dubious action in smart gadgets and organizations [2, 7, 8, 9, 10, 11, 12]. After a careful writing examination on [2, 7, 8, 9, 10, 11, 12] AI draws near and IoT security with regards to various potential dangers, this exploration [2] looks at ML calculations for assault and irregularity identification. Moreover, ML-based IoT [1, 2, 3] insurance techniques have been proposed.

Machine learning (ML) notices a framework to learn. Network occasions decide network conduct. Therefore ML is utilized in PC network security to recognize unlawful interruption. At the point when ML investigation goes amiss from anticipated regular organization action, dubious way of behaving is identified. Support vector machines (SVM) [6, 7] recognize ordinary and neurotic organization action utilizing ML. They figure out how to plan an ideal hyperplane that arranges obscure information vector values via plane area. [6] We recommend involving SVM models to distinguish vindictive action in low-power, low-rate, short-range networks like the Internet of Things. C-SVM and OC-SVM were tried. The previous requirements two classes of vector values (one for ordinary and one for distorted action), while the last option notices simply typical conduct action. The two techniques were utilized in an intrusion detection system (IDS) to screen and recognize brilliant hub gadget irregularities. We made and tried SVM location models utilizing real organization

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

traffic with our organization layer attacks. When tried with obscure information from similar organization geography as its preparation, the C-SVM accomplishes 100 percent arrangement accuracy and 81% accuracy in a new geography. Harmless OC-SVMs have a most extreme accuracy of 58%.

Irregularity identification has been utilized for a really long time to find and concentrate information peculiarities. Numerous techniques have been investigated to track down anomalies. ML is a developing methodology in this space [2, 7, 8, 9, 10, 11, 12]. An Systematic Literature Review (SLR) of ML models that recognize application abnormalities is introduced in this work [7]. We talk about the models according to four points of view: abnormality identification applications, ML draws near, execution measures for ML models, and characterization. In the wake of auditing 290 exploration distributions from 2000-2020, we tracked down ML peculiarity identification techniques. Our investigation of the chose research distributions yielded 43 abnormality recognition applications. Moreover, we find 29 ML models used to recognize irregularities. At long last, we give 22 oddity discovery datasets and numerous extra conventional datasets. Moreover, scientists favor unaided irregularity recognition over arrangement inconsistency identification. Scientists have applied numerous ML models [2, 7, 8, 9, 10, 11, 12] to distinguish irregularities, a promising field of study. We give scientists counsel and direction in light of this survey.[20]

Exceptions are information perceptions that digress from the standard. Dataset uprightness might be impacted by exceptions. Carrying out [2, 7, 8, 9, 10, 11, 12] AI methods in genuine applications and applying them to the medical services dataset will change the area. These applications might distinguish strange physiological information, provoking a fast reaction and uncovering more about the locale. The exhibition of irregularity discovery techniques on normal public datasets has been broadly examined [8]. Be that as it may, examine physiological data utilizing managed and unaided strategies negligibly. Bosom disease information is worldwide and mathematical. This article inspected four ML techniques for bosom disease dataset oddity recognition [2, 7, 8, 9, 10, 11, 12].

Scientists are as yet intrigued by interruption recognition, which has gathered broad consideration. The interruption discovery local area faces serious troubles following quite a while of study. Fixing the high misleading admonition rate while distinguishing obscure assault designs is as yet a test. A few late examinations give solutions for this quandary. Interruption recognition depends on oddity discovery, which identifies assaults, imperfections, abandons, and different anomalies. [10] This paper audits research on administered and unaided inconsistency discovery calculations. Major hypothetical subjects will be canvassed in the references, coordinating the scientist in captivating review roads.

## 3. METHODOLOGY

### i) Proposed Work:

Machine calculations are recommended for IoT anomaly detection. The calculations picked incorporate SVM, RF, stacking, and voting classifiers. Solid administered learning techniques SVM and RF

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

were utilized for discovery and element determination. Standard inconsistency dataset NSL-KDD [12] was utilized for tests. We look at the proposed model's exactness, review, accuracy, and f1-score to prior discoveries. The review introduced outfit calculations including the Voting Classifier (Random Forest and AdaBoost) and the Stacking Classifier (Random Forest, Multi-Layer Perceptron, LightGBM). Voting classifier had 100 percent exactness, stacking close to 100%, exhibiting their value in working on model prescience. Flask was utilized to give an easy to understand front-end interface for testing and execution. The recommended peculiarity recognition strategy might be utilized in IoT gadgets with client confirmation for safe access.

**ii) System Architecture:**

System for research is displayed in Fig. 1. This examination article analyzes the proposed execution of the two calculations used in Weka apparatus program to the most significant prior executions. These calculations are SVM and Random Forest. SVM is a complex Regulated Learning approach for order and relapse. Most of its utilization is in ML for Order [2, 7, 8, 9, 10, 11, 12]. Be that as it may, the random forest technique is not difficult to apply and adjust. It utilizes gathering learning for relapse and characterization.

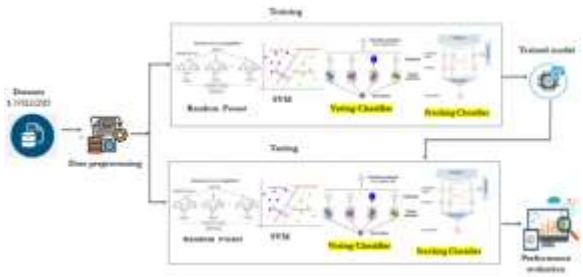


Fig 1 Proposed architecture

**iii) Dataset collection:**

Understanding the NSL KDD dataset's design and properties is the accentuation. To grasp qualities, information sorts, and examples, the dataset is stacked and analyzed. The review utilizes the standard anomaly dataset NSL-KDD [12] to look at intrusion detection systems. The proposed framework estimates model execution using accuracy, false positive rate, true positive rate, precision, recall, and F-measure. The public NSL-KDD dataset was made from the KDD cup99 dataset (Tavallaee et al., 2009). A factual examination of the cup99 dataset uncovered blemishes that significantly influence interruption location precision and misconstrue Helps (Tavallaee and al., 2009). The NSL_KDD dataset has 22 preparation interruption attacks and 41 highlights. This dataset has 21 association qualities and 19 host-explicit characteristics (Tavallaee et al., 2009).

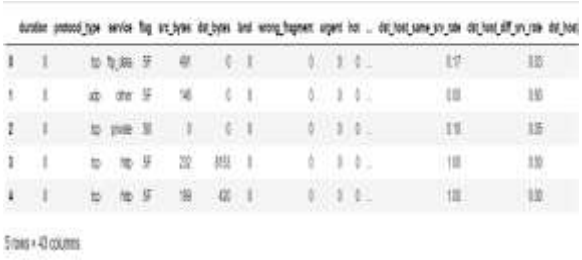


Fig 2 NSL KDD dataset

**iv) Data Processing:**

Data processing transforms crude information into business-helpful data. Information researchers accumulate, sort out, clean, confirm, break down, and organize information into diagrams or papers.

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

Information can be handled physically, precisely, or electronically. Data ought to be more important and decision-production simpler. Organizations might improve tasks and settle on basic decisions quicker. PC programming advancement and other robotized information handling innovations add to this. Huge information can be transformed into applicable experiences for quality administration and navigation.

**v) Feature selection:**

Feature selection chooses the most predictable, non-repetitive, and significant elements for model turn of events. As data sets extend in amount and assortment, deliberately bringing down their size is pivotal. The essential reason for highlight determination is to increment prescient model execution and limit processing cost.[22]

One of the vital pieces of feature engineering is picking the main qualities for ML calculations. Feature selection takes out repetitive or pointless attributes and limits the assortment of info factors to those generally vital to the [2, 7, 8, 9, 10, 11, 12] ML model. The significant benefits of pre-choosing attributes instead of letting the ML model choose.

**vi) Algorithms:**

**Random Forest** is a conspicuous regulated learning procedure for characterization and relapse. Ensemble learning works on anticipated exactness by consolidating a few choice trees on unmistakable dataset subsets. Random Forest is tried with 10 and 20 folds utilizing k-FOLD cross-validation on preparing and testing subsets to gauge speculation execution. Random Forest is fantastic for IoT applications and lessens overfitting worries because of its group nature and ability to deal with muddled datasets [14, 15].


Fig 3 Random forest

**Support Vector Machine** (SVM) is a typical characterization centered directed learning technique. It looks for an optimal choice limit, or hyperplane, to arrange n-layered space. K-Fold Cross-Validation with 10 and 20 folds separates information into subsets for preparing and testing in various rounds in SVM. SVM, which handles high-layered information really, is great for IoT anomaly detection, where choice cutoff points are intricate [14, 15].

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024


Fig 4 SVM

Stacking, a conspicuous ensemble demonstrating technique, parallelizes frail students and meta-students to improve future forecasts by figuring out how to total info expectations from various models. Stacking further develops IoT oddity distinguishing proof by social event shifted designs from sub-models and a meta-classifier to more readily sum up the dynamic and inconsistent nature of inconsistencies.


Fig 5 Stacking classifier

Voting classifiers use greater part casting a ballot to foresee a result in light of the best probability class from many models. In anomaly detection, the Voting Classifier totals expectations from many models to further develop execution and guarantee adjusted decision-production for a more reliable IoT framework.[25]


Fig 6 Voting classifier

## 4. EXPERIMENTAL RESULTS

**Precision:** Precision quantifies the percentage of certain events or tests that are well characterized. To attain accuracy, use the formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

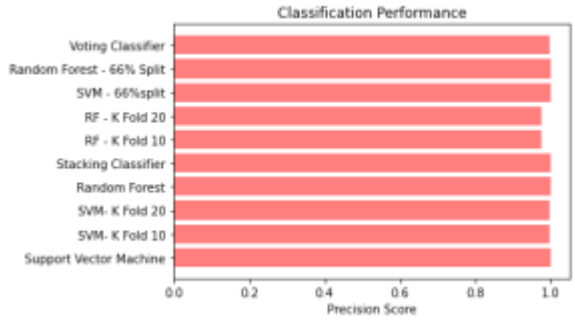



Fig 7 Precision comparison graph

**Recall:** ML recall measures a model's ability to catch all class occurrences [2, 7, 8, 9, 10, 11, 12]. The model's ability to recognize a certain type of event is measured by the percentage of precisely anticipated positive prospects that turn into real earnings.

$$Recall = \frac{TP}{TP + FN}$$

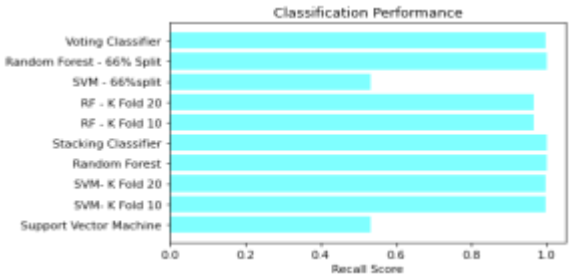



Fig 8 Recall comparison graph

**Accuracy:** The model's accuracy is the percentage of true predictions at a grouping position.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

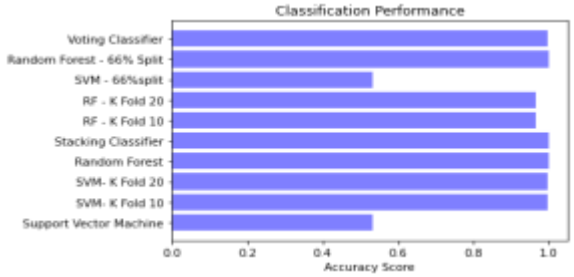



Fig 9 Accuracy graph

**F1 Score:** The F1 score captures both false positives and false negatives, making it a harmonized precision and validation technique for unbalanced data sets.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

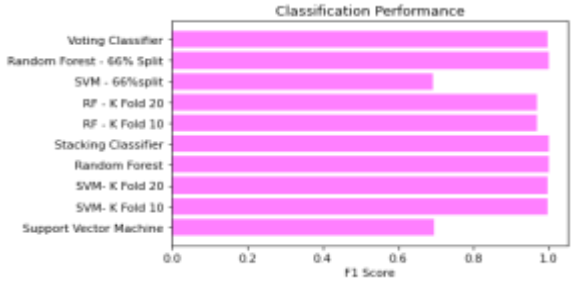



Fig 10 F1Score

| ML Model | Accuracy | Precision | Recall | F1 - score |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

Fig 11 Performance Evaluation

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

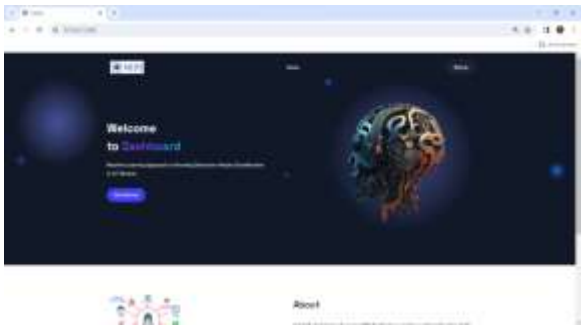



Fig 12 Home page

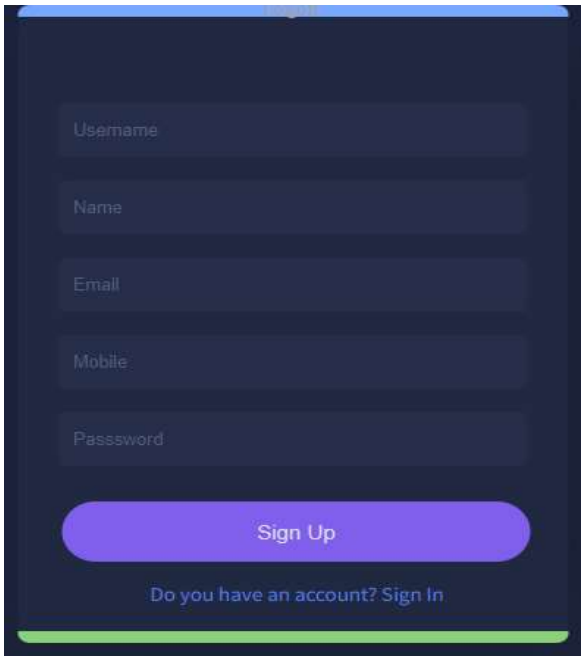



Fig 13 Signin page

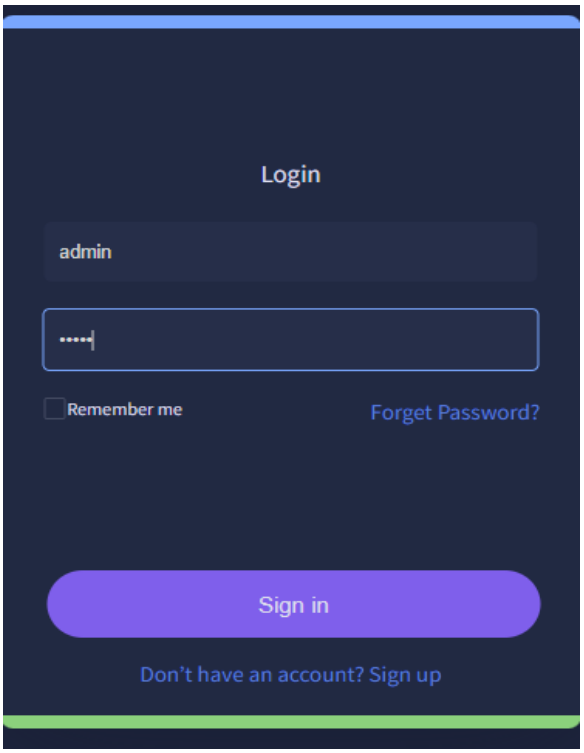



Fig 14 Login page


Fig 15 User input

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024


Fig 16 Predict result for given input

## 5. CONCLUSION

Support Vector Machine (SVM) and Random Forest (RF) calculations can distinguish and moderate atypical attacks in IoT gadgets [2, 7, 8, 9, 10, 11, 12]. Results show the proposed system outflanks writing. The methodology's accuracy in IoT peculiarity location demonstrates its reliability. The task keeps a low false positive rate across circumstances by underscoring method trustworthiness. For exact peculiar assault arrangement in the differed IoT gadget biological system, consistency is key [1, 2]. The NSL-KDD dataset is utilized to completely evaluate ML techniques [12]. This normalized dataset is a solid benchmark for IoT irregularity recognition viability. Another calculation utilizing gathering techniques like the Voting Classifier and Stacking Classifier accomplished 100 percent exactness. A thorough front-end test involving highlight values showed the calculation's security and constancy in certifiable conditions, demonstrating its convenience in further developing IoT irregularity discovery. This examination presents an answer that is compelling and tackles peculiar attacks, further developing IoT gadget protection from potential dangers.

## 6. FUTURE SCOPE

What's to come involves creating and carrying out modern ML strategies or ways to deal with further develop irregularity discovery. This could include involving profound learning models or attempting new procedures for better execution. Building the venture to distinguish IoT peculiarities continuously is a promising course. Remaining in front of rapidly arising digital dangers calls for constant information handling and examination procedures and strategies. Because of the powerful idea of IoT biological systems, the venture will construct versatile models to oversee new gadgets, information examples, and peculiarities [6, 11, 14]. Anomaly detection system improvement and updating are required. The undertaking might add encryption and irregular reaction in later adjusts. As the undertaking develops to battle progressively complex cyber threats, IoT gadget security should be thorough.

## REFERENCES

[1] M. Lee. “Anomaly Detection: Glimpse into the Future of IoT Data.” The New Stack. https://thenewstack.io/anomaly-detection-glimpse-into-thefuture-of-iot-data/ 2022, January 24.

[2] S. H. Haji, & S. Y. Ameen, “Attack and Anomaly Detection in IoT Networks using[2, 7, 8, 9, 10, 11, 12] Machine Learning Techniques: A Review.” In (p. 46). 2021.

[3] Firedome (2021). Top Cyber Attacks on IoT Devices in 2021. https://firedome.io/blog/top-cyber-attacks-on-iot-devices-in-2021/. 2021, November 30.

[4] A. ZMUDZINSKI, “Japan: Hacked IoT Devices and Cryptocurrency Networks Doubled in 2018.”. Cointelegraph. https://cointelegraph.com/news/japan-

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

hacked-iot-devices-andcryptocurrency-networks-doubled-in-2018. 2019, March 7.

[5] X. Xu, H. Liu, & M. Yao, Recent Progress of Anomaly Detection. Complexity, 2019, 1–11. https://doi.org/10.1155/2019/2686378. 2019.

[6] C. Ioannou, & V. Vassiliou, "Network Attack Classification in IoT Using Support Vector Machines." https://www.mdpi.com/2224-2708/10/3/58/pdf . 2021.

[7] B. Nassif, A. Abu Talib, M., Nasir, & F. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review." Ieee Access 9 (2021): 78658-78700. 2021 May 24.

[8] C. Das, A. Rasool, A. Dubey, & N. Khare,. "Analyzing the Performance of Anomaly Detection Algorithms." International Journal of Advanced Computer Science and Applications Vol. 12, no. 6 2021.

[9] Y. Gavrilova "Anomaly Detection in Machine Learning." Software Development Company. https://serokell.io/blog/anomaly-detection-inmachine-learning. 2021 December 10.

[10] S. Benqdara, & M. A. Ngadi,. "Machine Learning Techniques for Anomaly Detection: An Overview." International Journal of Computer Applications. Vol. 79, no. 2. 2013.

[11] M. Hasan, M. Islam, M. Md., I. Zarif, & M. M. A. Hashem. "Attack and Anomaly Detection in IoT Sensors in IoT sites using[2, 7, 8, 9, 10, 11, 12] Machine Learning Approaches." Internet of Things, Vol. 7, p.100059. 2019.

[12] Mathworks, "Machine Learning.". Www.mathworks.com. https://www.mathworks.com/discovery/machinelearning.html#:~:text=Machine%20learning%20uses%20two%20types. n. d,

[13] T. Crunch. "The evolution of machine learning." TechCrunch. 2017 Aug 8. https://techcrunch.com/2017/08/08/the-evolution-of-machinelearning/ (16 January 2023).

[14] B. Posey, S. Shea "What are IoT Devices?' TechTarget.com. IoT Agenda. 2022 https://www.techtarget.com/iotagenda/definition/IoT device (Accessed 16 January 2023).

[15] A.W. S. Amazon, "What is IoT? - Internet of Things Beginner's Guide - AWS.". Amazon Web Services, Inc. 2022 https://aws.amazon.com/what-is/iot/ (Accessed 16 January 2023).

[16] G.Viswanath, "Hybrid encryption framework for securing big data storage in multi-cloud environment", Evolutionary intelligence, vol.14, 2021, pp.691-698.

[17] Viswanath Gudditi, "Adaptive Light Weight Encryption Algorithm for Securing Multi-Cloud Storage", Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol.12, 2021, pp.545-552.

[18] Viswanath Gudditi, "A Smart Recommendation System for Medicine using Intelligent NLP Techniques", 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), 2022, pp.1081-1084.

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 535X
http://www.ijmert.com
Vol.16 Issue 3, 2024

[19] G.Viswanath, "Enhancing power unbiased cooperative media access control protocol in manets", International Journal of Engineering Inventions, 2014, vol.4, pp.8-12.

[20] Viswanath G, "A Hybrid Particle Swarm Optimization and C4.5 for Network Intrusion Detection and Prevention System", 2024, International Journal of Computing, DOI: https://doi.org/10.47839/ijc.23.1.3442, vol.23, 2024, pp.109-115.

[21] G.Viswanath, "A Real Time online Food Ording application based DJANGO Restfull Framework", Juni Khyat, vol.13, 2023, pp.154-162.

[22] Gudditi Viswanath, "Distributed Utility-Based Energy Efficient Cooperative Medium Access Control in MANETS", 2014, International Journal of Engineering Inventions, vol.4, pp.08-12.

[23] G.Viswanath," A Real-Time Video Based Vehicle Classification, Detection And Counting System", 2023, Industrial Engineering Journal, vol.52, pp.474-480.

[24] G.Viswanath, "A Real- Time Case Scenario Based On Url Phishing Detection Through Login Urls ", 2023, Material Science Technology, vol.22, pp.103-108.

[25] Manmohan Singh,Susheel Kumar Tiwari, G. Swapna, Kirti Verma, Vikas Prasad, Vinod Patidar, Dharmendra Sharma and Hemant Mewada, "A Drug-Target Interaction Prediction Based on Supervised Probabilistic Classification" published in Journal of Computer Science, Available at: https://pdfs.semanticscholar.org/69ac/f07f2e756b79181e4f1e75f9e0f275a56b8e.pdf