ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

Anomaly Detection in IOT

Ammar Mehboob
Department of Information Technology, Superior University, Lahore, Pakistan
Syeda Aqsa Zahra
Department of Information Technology, Superior University, Lahore, Pakistan
Muhammad Nabeel Amin
Department of Information Technology, Superior University, Lahore, Pakistan
Hamza Shabbir
Department of Information Technology, Superior University, Lahore, Pakistan
Dr. Asad Naqvi
Department of Information Technology, Superior University, Lahore, Pakistan

Abstract
One of the newest and most popular technologies is the Internet of Things (IoT). IoT has an impact on a number of industries, such as smart cities, healthcare, logistics tracking, and the automobile sector. Networks containing IoT devices are being the target of an increasing number of cyberattacks and breaches. By employing machine learning to improve anomaly detection, this paper seeks to strengthen security in IoT networks. This revealed the difficulties and weaknesses in protecting Internet of Things networks. The complexity of IoT networks, the human element, the quantity of devices, and network size are the obstacles. The identified limitations include the dearth of modeling input parameters necessary for anomaly identification in IoT networks, as well as the paucity of research on signature-based intrusion detection systems utilized for anomaly detection. Additionally, the performance of machine learning algorithms on real and standard IoT datasets is not compared. In order to evaluate the anomaly binary classification capabilities of the machine learning methods Neural Networks, Gaussian Naive Bayes, Support Vector Machines, and Decision Trees, this paper generates a dataset and contrasts its outcomes with the KDDCUP99 dataset. The outcomes demonstrate that on the generated IoT dataset, Support Vector Machine and Gaussian Naive Bayes outperform the other models. In this paper, the average execution time was lowered by 58% by reducing the number of characteristics needed by machine learning algorithms for anomaly detection in IoT networks to just five features. This paper examines the ability of CNNwGFC, an improved Convolutional Neural Network model, to identify and categorize irregularities in Internet of Things networks. Compared to the traditional Convolutional Neural Network, our model achieves a 15.34% increase in accuracy for IoT anomaly classification in the UNSW-NB15. Compared to the best accuracy found in the literature, the CNNwGFC multi-classification accuracy (96.24%) is 7.16 percent higher.
Keywords: *IoT; Machine Learning; Security; Anomaly Detection*

INTRODUCTION

The use of machine learning to detect anomalies in Internet of Things (IoT) networks. Anomaly detection, also known as novelty or outlier detection, is a technique that detects unusual objects or occurrences that differ significantly from a large amount of data. It has various applications, such as preventing financial fraud and protecting computer networks from irregular traffic. The Internet of Things (IoT) is a network of devices, sensors, and software that connect and share data, impacting various industries such as smart cities, healthcare, automotive, and logistics tracking. The IoT is expected to permeate every aspect of human existence.

Machine learning, part of the artificial intelligence field, is a growing field that uses computational algorithms to emulate human intelligence by learning from their surroundings. It enables software applications to become more accurate in predicting events, using algorithms such as Bayesian networks, support vector machines (SVM), artificial neural networks (ANN), and decision trees. The integration of computer-based systems with the physical world is the vision of the IoT.

Motivation

The Internet of Things (IoT) is a rapidly growing technology that presents numerous security challenges. As it becomes more widespread, the variety and vastness of IoT devices are expected to exacerbate existing security weaknesses. The IoT's evasiveness can lead to hackers exploiting its evasiveness to disrupt communications, make financial gains, or physically injure people. For instance, IoT baby monitors contain significant vulnerabilities that hackers can use to commit criminal crimes. Internet-connected autos can also be remotely controlled, allowing drivers to open doors and turn off engines while in motion. IoT hacking has also been found to extend to medical equipment, causing disastrous effects on patients. Current security

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

countermeasures require a significant amount of computational overhead and memory, which is often limited due to budgetary constraints. Therefore, solving major IoT security vulnerabilities is crucial for a world where the IoT is ubiquitous and accessible from anywhere. Anomaly detection is a significant challenge for IoT devices, and the development of artificial intelligence has made machine learning useful for anomaly detection.

Aim and Objectives

This project aims to enhance security in Internet of Things networks by utilizing machine learning to improve anomaly detection methods. The study will identify obstacles and unmet research needs in this field, evaluate the performance of machine learning algorithms like Decision Trees, Support Vector Machine, Neural Networks, and Gaussian Naive Bayes on an actual IoT dataset, reduce the number of characteristics needed for anomaly detection, and develop a better deep learning model based on CNN for anomaly detection and classification in IoT networks.

LITERATURE REVIEW

The advent of smart devices that can connect with humans and with each other through the Internet is causing a significant change in the way we use the Internet. According to recently released study by the massive technology research firm Gartner, there will be 304 billion "Things" online by 2020 (Munirathinam, S., 2020). The same source also states that 20 billion will likely be added to this total in the near future.

The origin of such a figure would be one question that would be asked right away. The explanation is straightforward: a user may have multiple Internet-connected devices. The iPhone, iPad, iWatch, Smart TVs, and many more gadgets are on the list, but they're not the only ones. The Internet of Things (IoT) is made up of all the devices that can produce data and distribute it over the Internet.

The term "Internet of Things" (IoT) refers to a broad range of technologies, systems, and design concepts related to the rapidly expanding phenomena of Internet-connected devices, or "Things." The term "IoT" is not new, according to Vermesan, O. and Friess, P. eds. (2022). It was originally used in 1999 at the Massachusetts Institute of Technology (MIT) Auto-ID Center to describe the creation of an Internet-based global network that would enable automatic object identification through information exchange. Radio Frequency Identification (RFID) and other related technologies are used by the "Things" to facilitate communication and realization.

The UN organization that oversees information and communication technology, the International Telecommunication Union (ITU), expanded the idea of the Internet of Things (IoT) in 2005 and proposed four technologies to make it a reality: RFID technology, intelligent embedded technology, nanotechnology, and sensor technology (Askari, H.,et al., 2019).

The Internet of Things (IoT) is defined by the IERC and is a dynamic global network infrastructure that automatically provides and distributes resources to satisfy the needs of current and future devices. This can be achieved through software defined networks (SDN) and cloud computing (CC), which can self-configure based on standard communication protocols like TCP/IP and UDP.

The advent of Internet Protocol version 6 (IPv6) and the declining cost of semiconductors are the two factors driving the revaluation of the Internet of Things, according to Patel, S., et al. (2023). Advancements in semiconductor technology and reduced raw material costs have led to a decrease in manufacturing costs, resulting in widespread computing.

IPv4 has run out of resources, leading to the replacement of IPv4 with IPv6, which offers a larger IP address space. IPv6 addresses are 128 bits long, allowing for 2 to the power 128 (or 2128) addresses in the address space, supporting an infinite number of "Things" connected to the Internet. (Ashraf, S., Muhammad, D., and Aslam, Z., 2020).

Classification of "Things" in IoT

"Physical and virtual things" with identities, qualities, and characteristics are included in the IERC definition of the Internet of Things (IERC, 2014). According to Vermesan, O., et al. (2022), some academics have conceptualized the Internet of Things (IoT) as a network that addresses physical things in terms of connectivity, control, and ubiquity. According to Vermesan, O., et al. (2022), the mapping between "things" in the physical and digital realms is what makes the Internet of Things functional, and as such, "things" fall into two categories: First, physical items; second, digital items.

- Objects are tangible entities with quantifiable bodies. Birds, cars, people, and tablets are a few examples of items.
- Behaviors: the motions of objects brought about by certain causes. Running, driving, watching, and eating are a few instances of object behaviors.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

- Tendency: these are the patterns in things themselves or as a result of the outside world, such clogged roads or rainy weather.
- Physical events: they are the culmination of all the previous points working together to explain what is happening as a result of certain conditions in the physical world.
- Entities: they denote abstract objects like code and data are covered under the category of cyber things.
- Actions: like data transmission from one entity to another, these signify the processing of data.
- Events: these denote the sequence of events leading up to an entity's actions, including data being communicated or reports from an entity.
- Services: they represent the services that an object offers or that are provided to an object in order to achieve a particular objective.

Google's automobile is a representation of an object capable of driving, maintaining itself, and parking. It may initiate physical actions like turning on wiper blades when it detects rain. Cyber objects provide context and meaning to physical objects, while abstractions communicate and process data based on available services. The code delivered to the wiper blades represents the code that performs actions and reports events.

IoT Architecture

The "Device layer" or "perception layer" is the layer at the bottom of the stack, containing IoT devices like smart meters and sensors, as well as gateways, a crucial component of Internet of Things systems. (Vermesan and Friess, 2014).

Gateways are crucial in sensor-rich environments, where hundreds of sensing devices from various manufacturers perceive various stimuli. These devices transmit their observations to the gateway via various communication protocols like Bluetooth, Wi-Fi, Ethernet, ZigBee, USB, and serial interfaces. Effective communication and management in this environment would be impossible without the gateway's ability to comprehend data from sensing objects using various protocols. (Ketshabetswe, L.K., et al., 2019).

The ability to use the Internet to transfer data from sensing items and gateways to the cloud is provided by the network layer in Internet of Things architecture. According to Lamkimel et al. (2018), gateways and certain sensing objects are capable of supporting internet protocols like IPv4, IPv6, and 6LoWPAN.

In Internet of Things architecture, management capabilities refer to tools used for remotely monitoring and controlling objects. These capabilities involve remote connection protocols like SSH and VPNs, which enable remote operations like firmware updates, patching, and configuration. (Dissanayake, N., et al., 2022).

Asset management is a network management capability where every object has an ID linking its properties to it. It includes attributes such as object ID, manufacturing, date of purchase, date of installation, capabilities, and supported protocols.

Security capabilities refer to the measures and safeguards implemented to protect sensing objects and their data as they move between devices. These controls can be logical or physical, such as access control, identity management, and encryption, which prevent illegal access or data changes. Physical controls, such as IoT objects like CCTV, can also be implemented to ensure data security. (Razaque, A., et al., 2022).

IoT intrinsic Characters

Ning (2013) lists three distinctive qualities of IoT that make it the Internet of things to come. These qualities are:

1. The Internet of Things (IoT) can provide sensing capabilities beyond human senses, bridging the gap between humans and machines. The EU-funded Guardian Angels for a Smarter Life project aims to develop intelligent personal assistants that support and protect people from early childhood to old age. Zero-power sensing devices can track human physical statures, communicate this data to doctors, and monitor the environment. Environmentally sensing devices can warn people of dangerous situations, and emotional sensing devices can identify and respond to human emotions. This initiative demonstrates how IoT can create a world where humans can help themselves while remaining invisible and powerless.
2. Network of networks - The Internet of Things encompasses multiple diverse network types. Stated differently, it is an IP-based, GSM, CDMA, WCDMA, and heterogeneous network.
3. IoT devices are designed to be intelligent, handling vast amounts of sensor data faster than humans can. This allows for qualitative thinking instead of information exploitation, allowing people to focus on more important tasks. The time and pace of data handling differs..

IoT Challenges

Although IoT is a promising technology, it is still in its infancy and faces numerous obstacles. Some of the IoT's problems are identified by Jung, Cho, and Kang (2014). The Internet of Things (IoT) systems face several challenges due to their infancy, lack of common architecture, heterogeneity of networks, and difficulty in

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

connecting, operating, managing, and securing them. Vendors are reluctant to create products that compete with others, leading to vendor lock-in. The vast number and variety of IoT devices, combined with their hardware limitations, make it difficult to safeguard them using host-based security solutions. This necessitates anomaly detection, protection, and centralized and decentralized network-based defenses to ensure the security of IoT systems.

Machine Learning for Anomaly Detection

One of the top 10 developing technologies for 2022 is machine learning (ML) (Malekloo, A., et al., 2022). Figure 2.4 shows that machine learning is one of the top four technologies that consumers can benefit from, but most Chief Information Officers (CIOs) have not yet utilized machine learning, as indicated by the red circle.

Definition

Making a computer machine capable of learning without direct programming or human involvement is the focus of the computer science field of machine learning (ML). ML learning has existed as a field since 1959, and it may have existed before. Bell J. (2022)Machine learning (ML) was first defined in 1959 by Arthur Samuel, an IBM engineer. It enables computers to learn without explicit programming by identifying pictures with brief textual descriptions. The system constructs its knowledge base after labeling the dataset. When given unlabeled data, such as an image of a "Lion," the system can respond to inquiries.

Unsupervised learning

Unsupervised learning uses unlabeled data, where the teacher doesn't know the labels or classes of objects in the input data. Instead, the teacher relies on the system itself to learn about these labels or classes. The machine performs methodical processing and analysis to divide the data into groups based on pre-selected attributes. The instructor can then label these groups in the system, allowing for forecasting when new unlabeled data will be presented. For example, a dataset of millions of transactional records can be divided into groups based on IP addresses used in each transaction. The instructor can assign geographic labels to each group and analyze the data to identify trends in purchasing behaviors based on their location. This model can make recommendations to users based on their location, similar to how Amazon suggests products to users of its websites. (Karn, A.L., et al., 2023).

Reinforcement Learning

Reinforcement learning is a method used to develop foundational knowledge using both labeled and unlabeled data. It operates by rewarding the system for each right or wrong prediction, which serves as input for subsequent predictions. The system then accesses a knowledge base with tailored execution pathways, attempting to determine the optimal path or combine multiple paths to provide predictions. This approach becomes advantageous if the reward is better than prior payouts for the same input.

Online games like chess use reinforcement learning, where a robot records every move made by a human opponent and waits for a reward from the tutor. The machine rewards correct moves and uses them to create new strategies to defeat human opponents. The output of machine learning models is highly precise, with a probability scale ranging from 0 to 1. In reinforcement learning, the machine makes predictions with a certainty level of 1 and waits for feedback to determine if the predictions were correct.

Unlabeled data is easier to collect than labeled data, but requires more time and resources to interpret and generate predictions. Tagged data is processed to include descriptive tags, but obtaining labeled data is more difficult than unlabeled data, often not free due to human evaluation and classification.. (Yuan, H., et al., 2024).

Related Work

This section critically assesses previous research on intrusion detection systems in the Internet, particularly the Internet of Things. It examines methods used to achieve primary outcomes and highlights the unique challenges faced by intrusion detection due to the variety of devices, computational limitations, and the vast number of linked devices. Intrusion Detection Systems (IDS) are used to prevent unwanted data access and modification within an information system.. According to Martins, I., et al., (2022), an IDS is designed to distinguish between normal behaviors from abnormal behaviors based on effective classification model built inside that IDS.

Intrusion Detection Techniques

Martins et al. (2022) propose a hybrid intrusion detection system based on Support Vector Machine (SVM) and Decision Tree (C5.0) for efficient classification models. The combination of these algorithms increases the accuracy of intrusion detection. SVM is recommended for classifying data into groups and can map data into a high-dimensional features space using nonlinear mapping. Support vectors can be computed and used to forecast class or category in new data samples. Martins et al. (2022) recommend integrating SVM with C5.0,

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

a decision tree implementation, where one root node and several non-root nodes respond to a specific attribute or tuple in the input record.

Intrusion Detection for IoT

Nguyen, X.H., et al. (2022) propose a lightweight intrusion detection system (IDS) for the Internet of Things ecosystem. The model identifies nodes with multiple identities based on the strength of the wireless signal and secret key and ID. It is applicable to wireless networks but is more of a patrolling model than an anomaly detection model. The model cannot identify irregularities in payloads transmitted to the control center. Non-IoT detection strategies are ineffective in IoT due to its open deployment and limited resources, and heterogeneous network of devices without common protocols and standards.

Nguyen et al. (2021) refer to this data as "perception layer data," which is sent from IoT objects to a centralized collection point for analysis. This data, including temperature and humidity, can be used to identify anomalies by analyzing data patterns and reporting them to the stage 2 distributed intrusion scheme.

1. Pulse: a sudden abruption of data for a very short time.
2. *Sharp ascent*: a sudden abruption of data that stays for a long time.
3. *Sharp descent*: a sudden drop of data that stays for a long time.
4. *Constant reading*: constant reading over a period of time.
5. *Change in patter*: data over a certain period of time is similar to normal patterns. However, there could be a difference in mean and/or variance.

RESEARCH METHODOLOGY

Materials

Sample Dataset

This study utilized secondary data from the Kaggle website, which contained numerous distinct attributes. The data was processed to extract relevant information and identify selection criteria with the strongest correlation to the event.

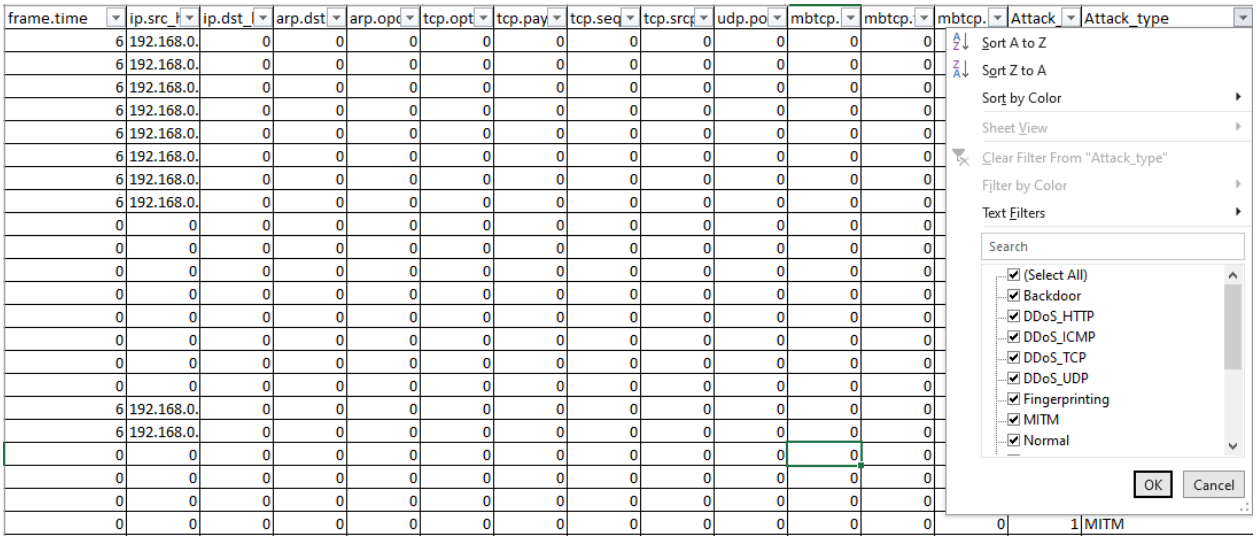


Figure 3.1: Dataset used in the study

Data Processing

Data was processed for the extraction of useful information and selection of features to be used in the study. Data processing involved following steps.

Data Cleaning

This step required creating an organization for the information model that was used to extract missing data, identify duplicate data, and purge superfluous information. Eventually, we obtain the most valuable data that underwent additional processing.

Data Selection

At this point, the data from the dataset that needs to be used for the upcoming research has been chosen and restored. Numerous features of the heart disease dataset are previously shown in figure 3.1. The dataset must go through this process because of the nature of the information.

Data Transformation

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

The data consolidation stage, also known as the normalization stage, is crucial as it converts previously selected data into a format suitable for data mining algorithms, minimizing the impact of scaling once the data file is stored.

Data Mining Stage

ANN, a machine learning algorithm, was utilized in a dataset analysis phase, utilizing four primary methodologies: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest, Artificial Neural Network (ANN), and Logistic Regression (LR), to identify intriguing illness patterns and hidden data within the dataset.

Methodology

The dataset was analyzed using various algorithms to predict cardiac illnesses based on provided attributes, with 13 distinct variables employed to ensure the best accuracy and demonstrate the predictive power of the data mining methods.

Machine Learning

Overview of Machine Learning

Machine learning is a branch of artificial intelligence that generates output from computers without explicit programming. It is popular for building prediction models through data analysis. Machine learning prioritizes tasks that computers can perform, followed by tasks that people cannot. Researchers argue that learning and prediction are the two phases of machine learning. The learning component provides training data, while the prediction section deals with machine-made predictions. Supervised and unsupervised learning are the two most popular types of machine learning.

Supervised Learning

Supervised machine learning is a method that uses labeled data to teach a computer to accurately classify input examples or forecast values, with regression and classification being the most commonly used outcomes.[11],

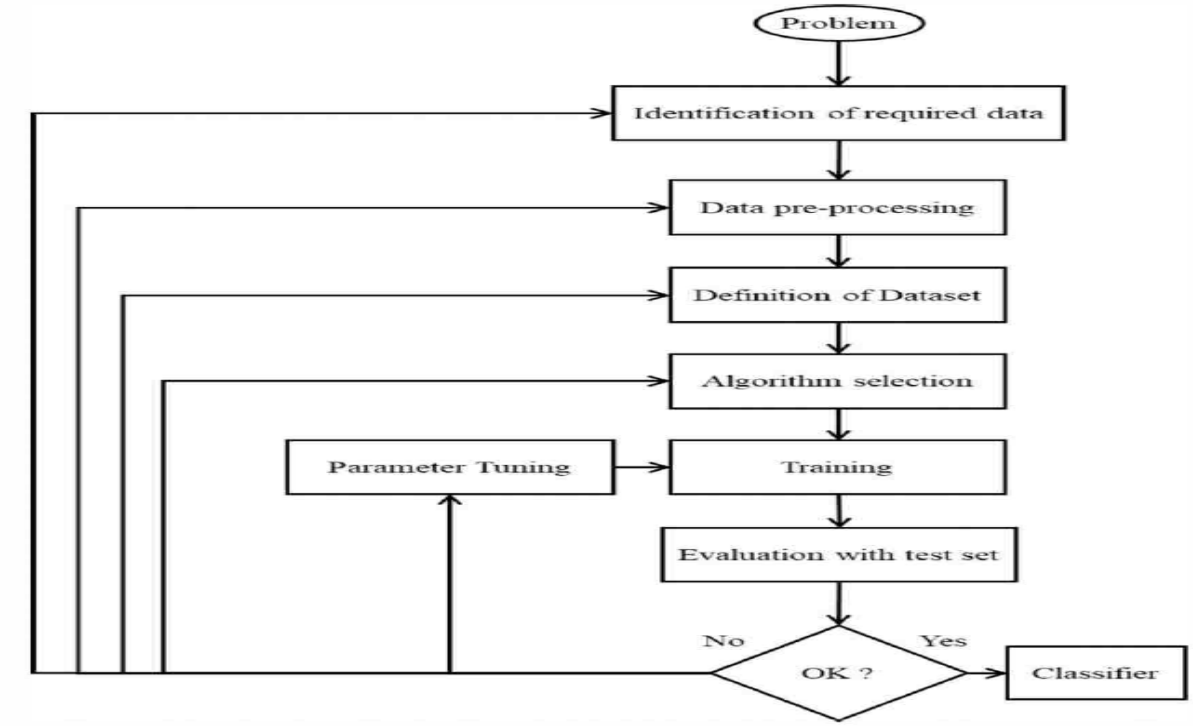


Figure 3-1: The process of supervised learning

Unsupervised Learning

Unsupervised learning aims to identify hidden patterns in input data, often used when a tagged dataset is unavailable. It can also classify unlabeled datasets, which are then used for supervised learning. Common forms include dimension reduction and grouping. This paper uses a labeled dataset, applying supervised learning to create a model of class label distribution, which can forecast class labels for test data. The process of categorization is crucial in selecting the appropriate classification model for a task.

Regression and classification are two kinds of learning algorithms, as illustrated in Figure (3-2). The overseer The research topic of Normal/Abnormal Traffic uses machine learning approaches for categorization, including supervised algorithms like Random Forest, Naive Bayes Classifier, Logistic Regression, Linear Classifier, Neural Networks, K-Means Clustering, Boosting, Perceptron, Decision Tree, Support Vector

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

Machines, Quadratic Classifiers, and Bayesian Networks. [14] [15]. [16] [17] The article provides an overview of classification-supervised machine learning techniques like neural networks, support vector machines, naive bayes, decision trees, and deep learning for anomaly detection.

Artificial Neural Networks

Artificial Neural Networks (ANN) are supervised machine learning techniques that mimic the structure and functions of the human brain. Initially introduced half a century ago, ANNs analyze data by labeling, searching for trends, and classifying raw input. Real-world input, such as text, images, and sounds, must be converted into numerical representation before being evaluated. ANNs consist of two main parts.

• Attachments (weights): These are the connections in a neural network that have values that are modified during training.

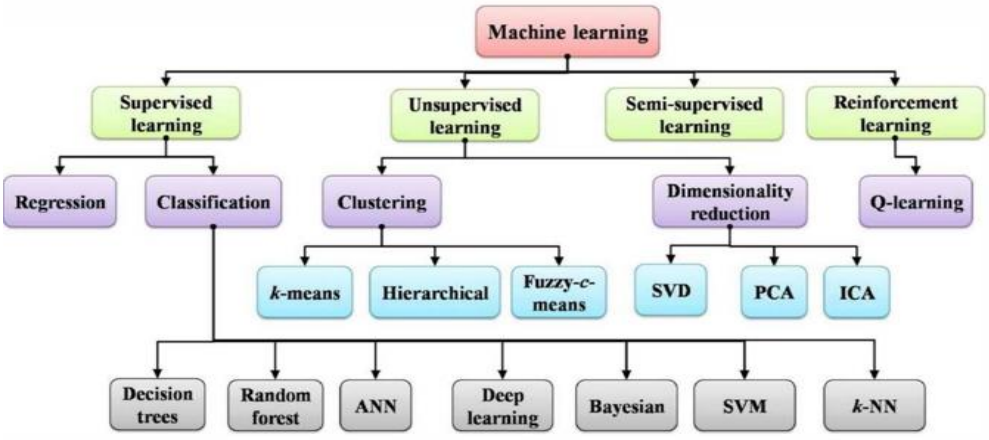


Figure 3-2: Taxonomy of ML techniques [13],

Neurons receive inputs from other neurons through weight-valued connections, multiply them by the proper weights, and add them together. An activation function is applied to determine if the neuron "fires," represented mathematically in equation (2-1). The process's output, Y, is processed by the activation function (2-2).

The term "fire" originates from the brain's basic activities, and one neuron must have an activation function to complete a binary classification task, such as the sigmoid function, which is an example of such an function.

f(z) = 1/ (1 + exp(-z)) **(3.1)**

Where:

- f(z) is the output value between 0 and 1,
- z is the input to the sigmoid function, which can be any real number,
- exp() is the exponential function, and
- 1 / (1 + exp(-z)) is the formula that maps z to the range (0, 1).

Figure (3.3) displays a graph of the sigmoid function.

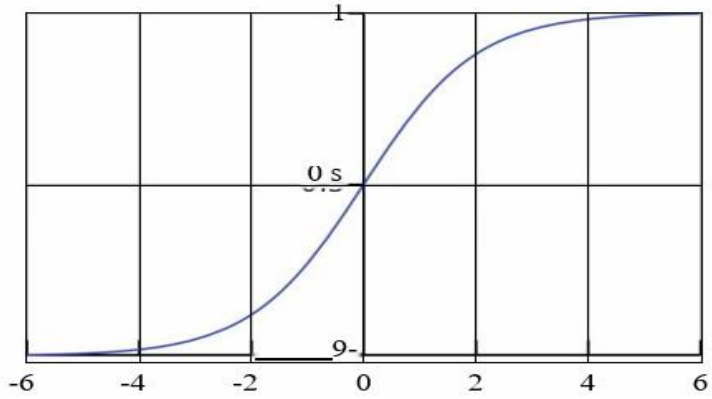


Figure 3-3: Sigmoid Function.

Three layers make up a multi-layer perceptron, also referred to as a basic neural network, as shown in Figure (2-4) [19]:

• Input Layer: This top layer is responsible for absorbing external data.

• Hidden Layer: Located in between the input and output layers, this layer is responsible for converting input data into output.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

• Output Layer: In a neural network, the output is generated last in this layer.

Equation (2-3) displays the backpropagation formula, which is used to modify the weights of the neurons inside each layer [20]. The gradient is the derivation inside fraction, alpha is the learning rate, and W is the change in the edge weight at time t (or t — 1 for the previous iteration).

Random Forest

Random forests, also known as random choice forests, are a popular ensemble learning method for classification and regression tasks. They build numerous decision trees during the training phase, with the output representing the class most trees select for classification tasks and the mean or average prediction for regression tasks.[21],

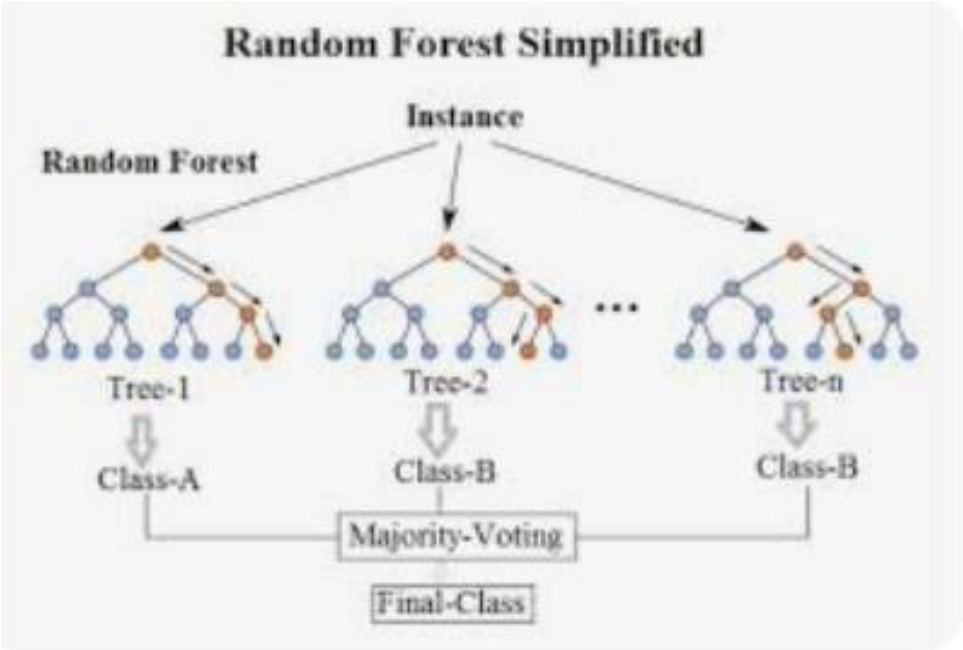
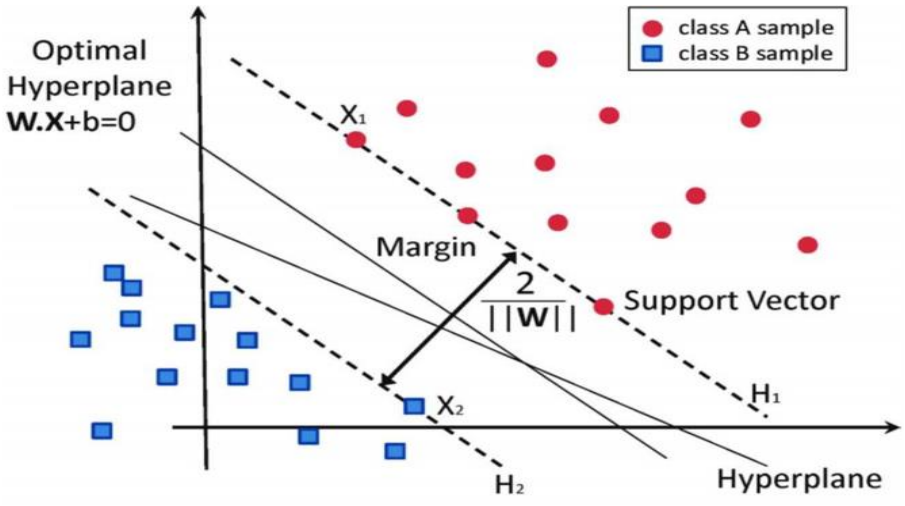


Figure 3-5: An example of a Random Forest

Trees often overfit their training sets, leading to low bias but high variation, resulting in chaotic patterns. Random forests, a strategy of averaging multiple deep decision trees trained on different parts of the same set, reduce variation and improve performance, but may increase bias and loss of interpretability. This approach is often used to reduce variation in models.

Support Vector Machines

Support Vector Machine (SVM) is a supervised machine learning technique that aids in regression and classification using n-dimensional space. It uses a hyperplane with dimensions of (n-1) to split two classes, assisted by support vectors. SVM has a maximum margin classifier with a soft margin, but certain observations can violate this limit.



K-Nearest neighbor (KNN)

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

Supervised machine learning (KNN) is a method that classifies new data points into the target class based on the attributes of nearby data points. This approach is simple and effective in various challenges. KNN compares each data point's properties with its neighboring data points to classify data points based on feature similarity. It avoids assumptions and works with available information. K Nearest Neighbor is the number of nearest neighbors, and it is used for regression and classification problems. For example, a new data point with seven nearest neighbors will be allocated to class B, as it has more adjacent points than the other data point.

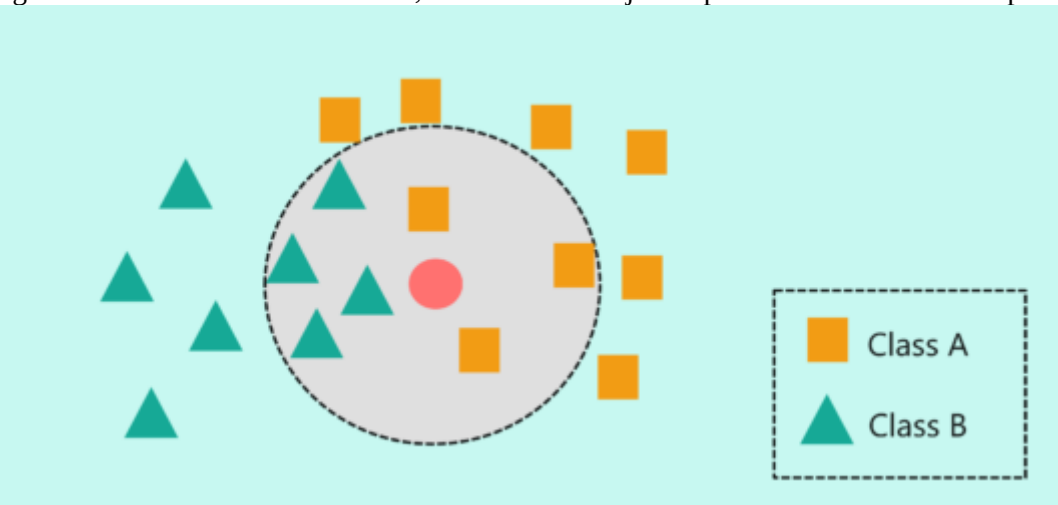


Figure 3-7: KNN showing number of nearest neighbors to new data point

Identifying Intrusions

An invasion is any action that compromises a resource's availability, security, or privacy. The National Institute of Standards and Technology (NIST) provides standards for intrusion detection systems. Addressing intrusion detection is challenging due to the complexity of computer systems, potential vulnerabilities, and attacker skill. An invasion occurs when a user gains unauthorized access or maliciously exploits information resources.

- Traditional intrusion detection focuses on anomaly or misuse detection, aiming to identify unusual individual or group actions. This involves building knowledge bases from observed behaviors profiles. A common method used to identify anomalies is using a method that uses a combination of these methods.:

i. • Threshold detection, which detects anomalous behavior on the network or in the server, like anomalous CPU usage on a single server or anomalous network congestion.
ii. • Statistical measures obtained through statistical analysis of historical values.
iii. • Law-based regulations backed by intelligent systems
iv. Non-linear algorithms include, for example, neural networks and genetic algorithms.
v. The second method, known as misuse detection, looks for patterns in user behavior that are connected to hostile actors trying to get access to a system [38]. While anomaly detection often uses threshold monitoring to discover occurrences, it uses a rule-based approach. One of the following techniques is frequently used to detect abuse [39]:

vi. Expert systems that provide an array of rules to characterize attacks.
vii. Signature verification, in which attack scenarios are used to generate audit event sequences.
viii. Petri nets: These tools show examples of known attacks.
ix.State-transition diagrams, which show the objectives and modifications of an attack.

Detection

Defense systems should be part of an effective attack detection stage to identify invasions before they cause significant damage. Intrusions are attempts to access, create, alter, or remove data, making a system untrustworthy. Researchers are developing intrusion detection systems (IDSs) to address the increasing number of attacks. IDSs can be classified into anomaly-based and signature-based detection. Signature-based detection is more successful but relies on prior knowledge of breaches, leaving a network vulnerable. Hybrid detection systems combine signature-based and anomaly-based detection techniques.

Reaction

Backup plans are crucial in case of an attack. Intelligent reply strategies use less bandwidth and separate attack flow packets from routine ones. Filtering only attack traffic and leaving routine traffic unfiltered is essential. Combining detection methods with reaction mechanisms can improve results. Techniques include filtering,

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

breaking off active network connections, establishing rules, and following tracebacks. Defenses can be active or passive, with active systems responding to attack traffic immediately, while passive systems examine the attack traffic record passively to identify attack origins.

Internet of Things

The Internet of Things (IoT) is a rapidly growing technology that is transforming our lives. With an estimated 27 billion IoT devices in 2017, it is projected to reach 75 billion by 2025. These devices, equipped with software, sensors, and electronics, collect and exchange data between devices. They are useful and practical due to their user-friendly features. IoT devices have various applications, including machine-to-machine and machine-to-human technology, wearable technology, security cameras, smart plugs, locks, smart meters, RFID, and wireless sensors. Their use is expected to impact every aspect of human life.

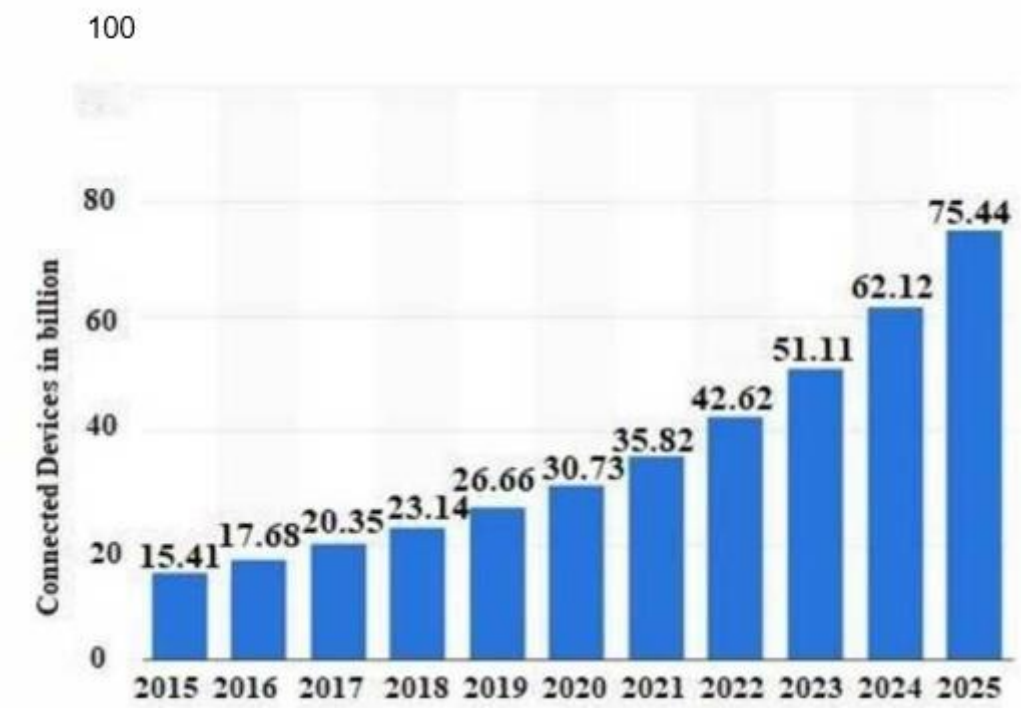


Figure 2-16: Growing number of IoT Devices [82],

Challenges in Securing IoT

A) Network Size and Number of Devices

The security of IoT networks is challenging due to limited cybersecurity techniques, budget constraints, and limited memory and processing power. New scalable security mechanisms are needed to improve energy efficiency and balance computing and storage capacities.

B) Human Factor

The Internet of Things (IoT) is revolutionizing human-machine interactions, enabling remote data collection and medication delivery. However, unauthorized access and user involvement challenges persist, necessitating innovative trust and reputation systems to support billions of users.

C) Complexity

The Internet of Things (IoT) connects people, machines, and objects, but many rely on outdated, proprietary technologies, creating the "Intranet of Things." Despite open protocols, accessibility issues persist due to changing device positions and communication capabilities.

RESULTS AND DISCUSSION

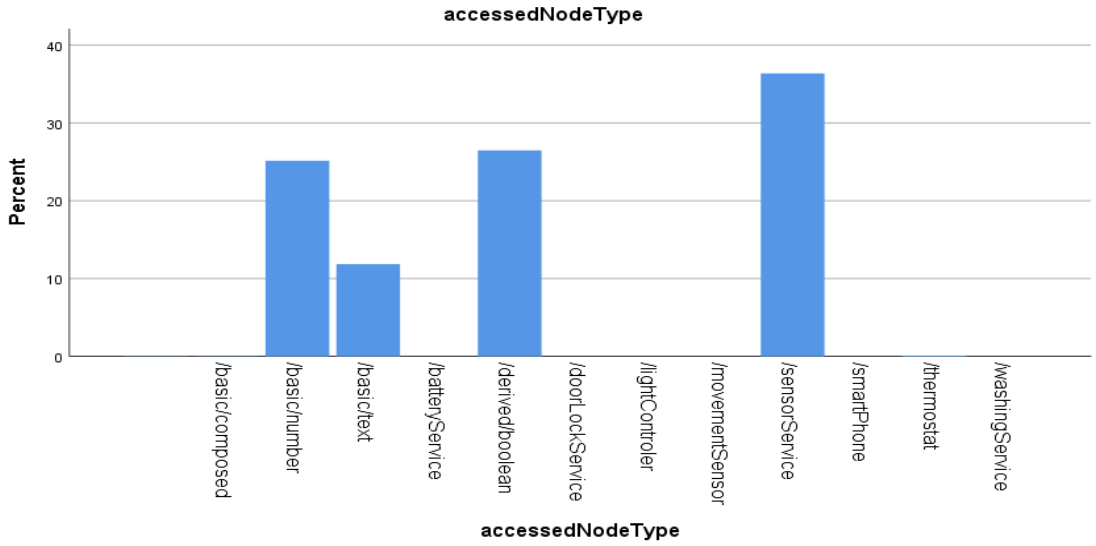
The study utilized smart home sensor readings to identify anomalies in IoT devices using a hybrid method combining statistical techniques and machine learning algorithms, identifying 95% of injected anomalies.

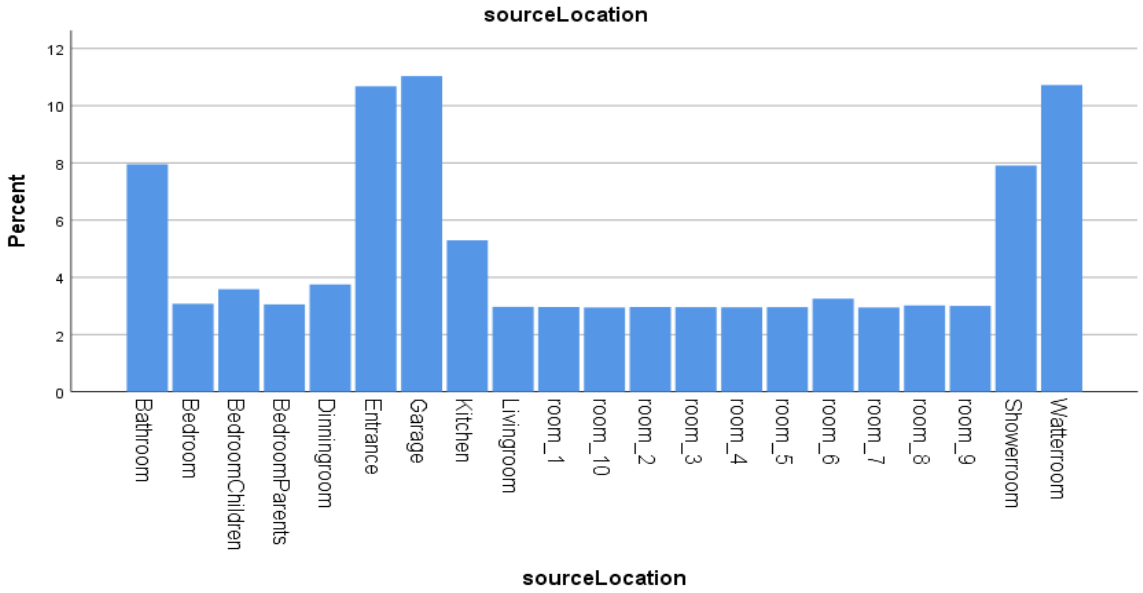
Univariate Analysis

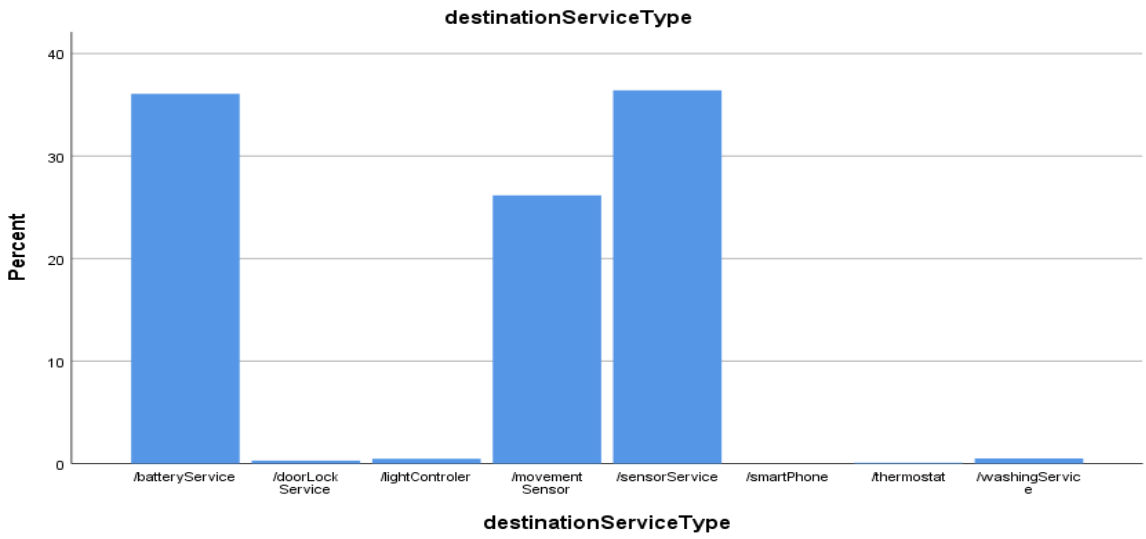
The report reveals a significant disparity in access node types in the Internet of Things, with sensor services, numerical data, and boolean data being the most common, emphasizing the importance of state-based data for anomaly detection.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

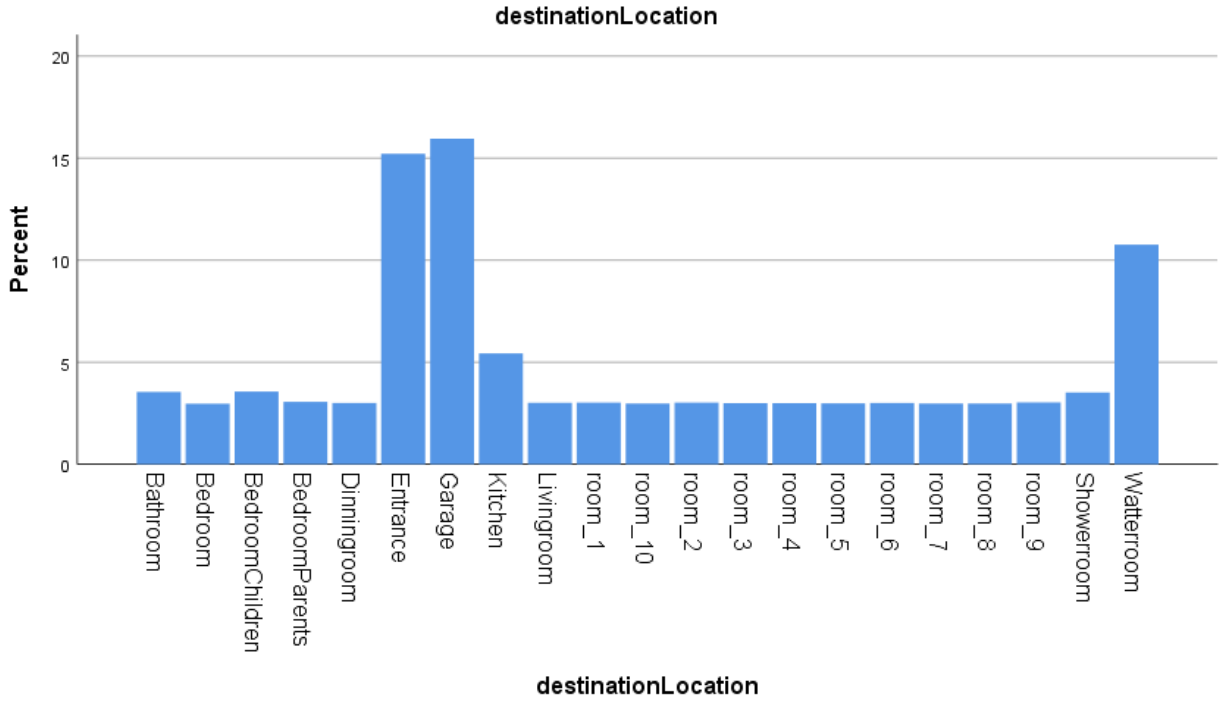
The dataset reveals a diverse range of IoT data sources, with garages, entrances, wards being the most common. Shower rooms and bathrooms contribute significantly, while designated bedrooms and numerical labels highlight the need for anomaly detection systems.



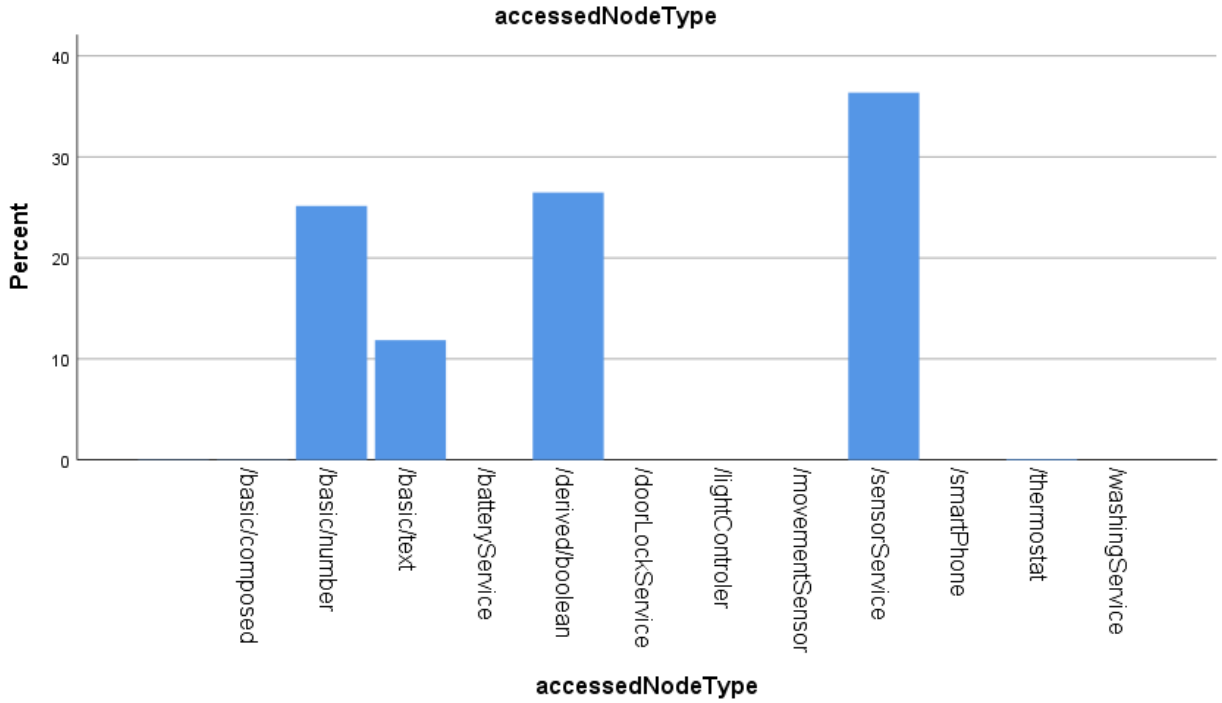




ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)



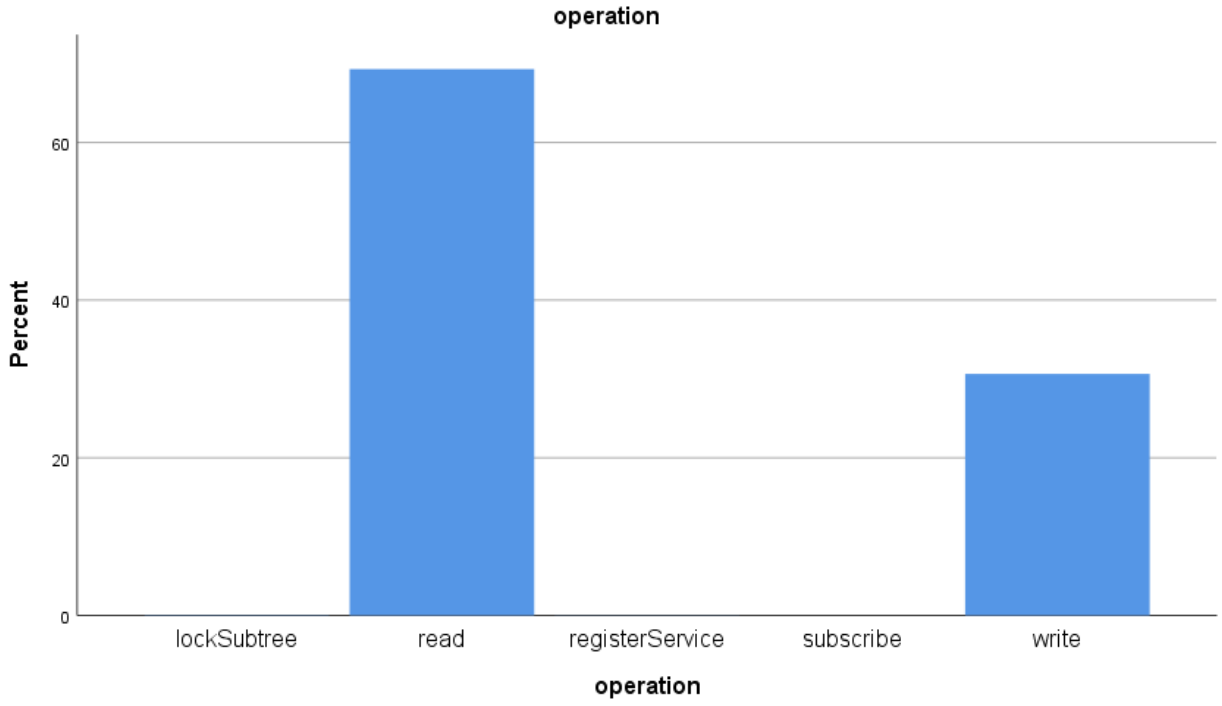
The dataset shows a diverse distribution of destination locations across IoT ecosystem domains, with the garage and entrance being the most visited areas. The Watterroom and Kitchen are also significant, while smaller but consistent data distributions are found in the Showerroom, Bathroom, and numerically labeled rooms. This suggests a consistent monitoring strategy across various locations.

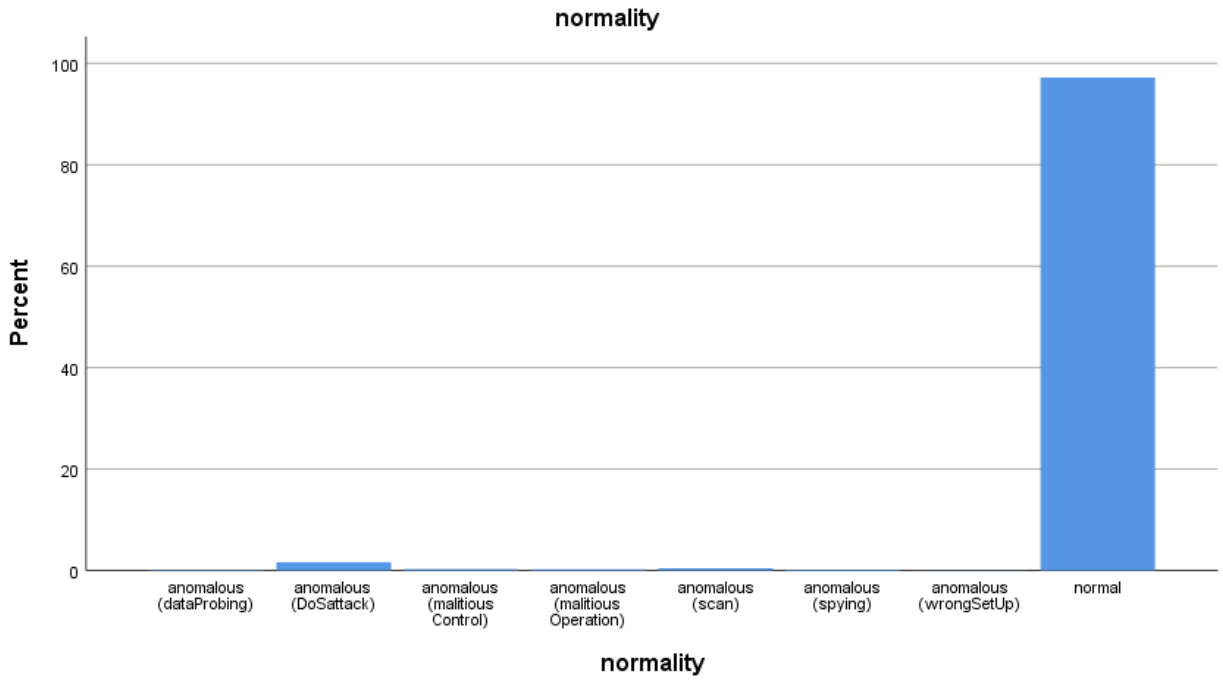


Multinomial Logistic Regression

Multinomial logistic regression is a statistical method that simulates relationships between multiple dependent and independent variables. It compares probabilities of each category to a reference category, determining the variables affecting each event's probability. This method is useful in fields like social sciences, marketing, and healthcare for understanding and predicting outcomes.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)





CONCLUSIONS

This paper describes various machine learning and deep learning strategies for anomaly detection in the Internet of Things during the last few years. Researchers studying anomaly detection in the Internet of Things can gain new knowledge from this review. A new unknown attack is found every day, necessitating the development of new algorithms to identify the attack. IoT requires a lot of compute because it involves a lot of devices and generates a lot of data. To increase computing speed, feature extraction is required. In the Internet of Things, anomaly detection at edge devices will lead to quicker responses and higher service quality. Real-time anomaly detection is necessary for IoT real-time data streaming. Finding anomalies with this method is difficult and expensive deep learning, which on IoT devices with limited resources requires high compute

SUMMARY AND RECOMMENDATIONS

This study presents a hybrid anomaly detection strategy for smart home IoT systems, combining machine learning tools and statistical methodologies. The method has a low false positive rate of 3% and a high success

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)

rate of 95% in recognizing abnormalities. The dataset's sensor and numerical data are crucial for efficient anomaly detection. Battery and temperature monitoring are essential in the IoT ecosystem. The multinomial logistic regression model achieves 100% accuracy in distinguishing normal and specialized anomaly types, demonstrating the utility of advanced statistical methods.

REFERENCES

Álvarez, I., Ballesteros, A., Barranco, M., Gessner, D., Djerasevic, S. and Proenza, J., 2019. Fault tolerance in highly reliable ethernet-based industrial systems. *Proceedings of the IEEE*, *107*(6), pp.977-1010.

Arivardhini, S., Alamelu, L.M. and Deepika, S., 2020, March. A Hybrid Classifier Approach for Network Intrusion Detection. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 824-827). IEEE.

Azeroual, O. and Nikiforova, A., 2022. Apache spark and mllib-based intrusion detection system or how the big data technologies can secure the data. *Information*, *13*(2), p.58.

Bell, J., 2020. *Machine learning: hands-on for developers and technical professionals*. John Wiley & Sons.

Butun, I., Almgren, M., Gulisano, V. and Papatriantafilou, M., 2020. Intrusion detection in industrial networks via data streaming. *Industrial IoT: Challenges, Design Principles, Applications, and Security*, pp.213-238.

Chaabouni, N., Mosbah, M., Zemmari, A., Sauvignac, C. and Faruki, P., 2019. Network intrusion detection for IoT security based on learning techniques. *IEEE Communications Surveys & Tutorials*, *21*(3), pp.2671-2701.

Dissanayake, N., Jayatilaka, A., Zahedi, M. and Babar, M.A., 2022. Software security patch management-A systematic literature review of challenges, approaches, tools and practices. *Information and Software Technology*, *144*, p.106771.

Karn, A.L., Karna, R.K., Kondamudi, B.R., Bagale, G., Pustokhin, D.A., Pustokhina, I.V. and Sengan, S., 2023. Customer centric hybrid recommendation system for e-commerce applications by integrating hybrid sentiment analysis. *Electronic Commerce Research*, *23*(1).

Ketshabetswe, L.K., Zungeru, A.M., Mangwala, M., Chuma, J.M. and Sigweni, B., 2019. Communication protocols for wireless sensor networks: A survey and comparison. *Heliyon*, *5*(5).

Lamkimel, M., Naja, N., Jamali, A. and Yahyaoui, A., 2018, November. The Internet of Things: Overview of the essential elements and the new enabling technology 6LoWPAN. In *2018 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)* (pp. 142-147). IEEE.

Lei, X., Mohamad, U.H., Sarlan, A., Shutaywi, M., Daradkeh, Y.I. and Mohammed, H.O., 2022. Development of an intelligent information system for financial analysis depend on supervised machine learning algorithms. *Information Processing & Management*, *59*(5), p.103036.

Malekloo, A., Ozer, E., AlHamaydeh, M. and Girolami, M., 2022. Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Structural Health Monitoring*, *21*(4), pp.1906-1955.

Martins, I., Resende, J.S., Sousa, P.R., Silva, S., Antunes, L. and Gama, J., 2022. Host-based IDS: A review and open issues of an anomaly detection system in IoT. *Future Generation Computer Systems*, *133*, pp.95-113.

Maurer, S., 2022. *Guardian angel: a driver-vehicle interaction for oversteering the driver in a highly automated vehicle* (Doctoral dissertation, Universität Ulm).

Mrabet, H., Belguith, S., Alhomoud, A. and Jemai, A., 2020. A survey of IoT security based on a layered architecture of sensing and data analysis. *Sensors*, *20*(13), p.3625.

Mugabo, E., Zhang, Q.Y., Ngaboyindekwe, A., Kwizera, V.D.P.N. and Lumorvie, V.E., 2021. Intrusion detection method based on mapreduce for evolutionary feature selection in mobile cloud computing. *International Journal of Network Security*, *23*(1), pp.106-115.

Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J. and Poor, H.V., 2021. Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *23*(3), pp.1622-1658.

Nguyen, X.H., Nguyen, X.D., Huynh, H.H. and Le, K.H., 2022. Realguard: A lightweight network intrusion detection system for IoT gateways. *Sensors*, *22*(2), p.432.

Raj, M., Gupta, S., Chamola, V., Elhence, A., Garg, T., Atiquzzaman, M. and Niyato, D., 2021. A survey on the role of Internet of Things for adopting and promoting Agriculture 4.0. *Journal of Network and Computer Applications*, *187*, p.103107.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.02 No.04 (2024)
Ray, S., Mishra, K.N. and Dutta, S., 2020. Big data security issues from the perspective of IoT and cloud computing: A review. *Recent Advances in Computer Science and Communications*, *12*(1), pp.1-22.

Razaque, A., Shaldanbayeva, N., Alotaibi, B., Alotaibi, M., Murat, A. and Alotaibi, A., 2022. Big data handling approach for unauthorized cloud computing access. *Electronics*, *11*(1), p.137.

Samara, M.A., Bennis, I., Abouaissa, A. and Lorenz, P., 2022. A survey of outlier detection techniques in IoT: Review and classification. *Journal of Sensor and Actuator Networks*, *11*(1), p.4.

Sen, P.C., Hajra, M. and Ghosh, M., 2020. Supervised classification algorithms in machine learning: A survey and review. In *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018* (pp. 99-111). Springer Singapore.

Taha, A. and Hadi, A.S., 2019. Anomaly detection methods for categorical data: A review. *ACM Computing Surveys (CSUR)*, *52*(2), pp.1-35.

Yuan, H., Chan, S., Creagh, A.P., Tong, C., Acquah, A., Clifton, D.A. and Doherty, A., 2024. Self-supervised learning for human activity recognition using 700,000 person-days of wearable data. *npj Digital Medicine*, *7*(1), p.91.

Zhou, X., Koltun, V. and Krähenbühl, P., 2020, August. Tracking objects as points. In *European conference on computer vision* (pp. 474-490). Cham: Springer International Publishing.