A Mini Project with Seminar On

**EFFICIENT HVAC SYSTEMS USING IoT**

Submitted in partial fulfillment of the requirements for the award of the

**Bachelor of Technology**

in

**Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning)**

by

**Mohammed Abdul Jabbar 21241A66H0**

**Shyam Manoj 21241A66G0**

**Lokesh Bodicherla 21248A66E5**

Under the esteemed guidance of **Mr. Neeraj Sharma**

**Assistant Professor**



**Department of Computer Science and Engineering**

(**Artificial Intelligence and Machine Learning)**

### GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND TECHNOLOGY

**(Approved by AICTE, Autonomous under JNTUH, Hyderabad) Bachupally, Kukatpally, Hyderabad-500090**



### GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND

**TECHNOLOGY**

**(autonomous)**

**Hyderabad(500090)**

**CERTIFICATE**

### 

### This is to certify that the project work entitled “Energy Efficient HVAC Systems using IoT” is submitted by Mohammed Abdul Jabbar (21241A66H0),Shyam Manoj (21241A66G0) and Lokesh Bodicherla(21241A66E5) in partial fullfillment of the award of the degree in Bachelor of Technology in Computer Science and Engineering(Artificial Intelligence and Machine Learning) during the academic year 2023-2024.

**Internal Guide Head Of Department**

**Mr. Neeraj Sharma Dr G. Karuna**

**External Examiner**

## ACKNOWLEDGEMENT

## 

## There are many people who helped us directly and indirectly to complete our Project Work successfully. We would like to take this opportunity to thank one and all. First, we would like to express our deep gratitude towards our internal guide Mr Neeraj Sharma, Assistant Professor, Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), for his support in the completion of our dissertation. We are Thankful to out mini project coordinator Ms M Shamila,Assistant Professor, for her valuable suggestions and comments during this period.We wish to express our sincere thanks to Dr. G. Karuna, Head of the Department, and to our principal Dr. J. Praveen, for providing the facilities to complete the dissertation.Finally, we are very much indebted to our parents for their moral support and encouragement to achieve goals.

## 

## 

**Mohammed Abdul Jabbar 21241A66H0**

**Shyam Manoj 21241A66G0**

**Lokesh Bodicherla 21248A66E5**

### DECLARATION

We hereby declare that the project work phase titled "Energy Efficient HVAC Systems using IoT" is a work done during the period from **6th Feb 2024** to **29 June 2024** and is submitted in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering(Artificial Intelligence and Machine Learning) from Gokaraju Rangaraju Institute of Engineering and Technology (autonomous under Jawaharlal Nehru Technological University, Hyderabad). The results embodied in this work have not been submitted to any other university or institution for the award of any degree or diploma.

**Mohammed Abdul Jabbar 21241A66H0**

**Shyam Manoj 21241A66G0**

**Lokesh Bodicherla 21248A66E5**

## 

## 

## 

## 

## ABSTRACT

The project aims to enhance HVAC system functionality through advanced sensors and smart algorithms. This project focuses on maintaining the optimal functionality of HVAC systems in buildings using sensors for continuous monitoring of temperature, airflow, and humidity. By continuously collecting critical data (temperature, airflow, humidity), these sensors enable the identification of anomalies and potential failures. The existing approaches are mostly void of the integration of machine learning algorithms and rely heavily on physics based methodologies like PID. The disadvantages are the ineffectiveness of gathering enough data from various machines and applying more optimal algorithms to the anomaly detection process. Also, there exist single-speed HVAC systems that run at full capacity which leads to higher energy consumption costs. The project envisions adaptive HVAC systems that maximize energy efficiency and sustainability.The project uses efficient IoT sensors and Machine learning algorithms such as Linear Regression and SVMS to accurately detect anomalies uring IsolatedForests. The project displays accurate results with a Mean Absolute Error of 0.051 for Support Vector Machines.

## LIST OF FIGURES

| **Figure No** | **Figure Name** | **Page No** |
| --- | --- | --- |
| 1 | Visual Representation of the System Architecture Diagram | 5 |
| 2 | Machine Learning Visual Representation of the System Architecture Diagram | 30 |
| 3 | Module Connectivity Diagram | 31 |
| 4 | UML Class Diagram | 38 |
| 5 | UML Sequence Diagram | 39 |
| 6 | UML Use Case Diagram | 40 |
| 7 | UML Activity Diagram | 41 |
| 8 | Scatter plot of Linear Regression | 47 |
| 9 | Linear Regression Histogram Plot | 48 |
| 10 | Scatter Plot of Lasso Regression | 49 |
| 11 | Histogram Plot of Support Vector Machines | 50 |
| 12 | Anomaly Detection Scatter Plot | 51 |
| 13 | Histogram Plot for the Regression Model Implemented | 57 |
| 14 | Histogram Plot for SVM Implemented | 58 |
| 15 | A Depiction Of The Anomalies Detected In IsloationForest In The Implementation | 59 |
| 16 | Flask Server Code Implementation | 68 |
| 17 | Flask Server Initialization | 69 |
| 18 | Aurduino Implementation Code | 70 |
| 19 | Lasso Regression Implementation Code | 71 |
| 20 | IsolationForest Implementation Code | 72 |

## LIST OF TABLES

| **Table No** | **Table Name** | **Page No** |
| --- | --- | --- |
| 1 | TABLE 1. Summary of the Existing Approaches | 17 |

## LIST OF ACRONYMS

| HVAC | Heating, Ventilation, and Air Conditioning |
| --- | --- |
| SVM | Support Vector Machines |
| DHT11 | Digital Humidity and Temperature Sensor |
| MQ-8 | Gas Sensor (Hydrogen) |
| ESP | Espressif (brand of microcontrollers, e.g., ESP8266) |
| Wi-Fi | Wireless Fidelity |
| UML | Unified Modeling Language |
| F1 score | F1 Score (Harmonic mean of precision and recall) |
| MARL | Multi-Agent Reinforcement Learning |
| HMM | Hidden Markov Model |
| IPCA | Incremental Principal Component Analysis |
| MSIPCA | Multi-Stage Incremental Principal Component Analysis |
| WNNs | Wavelet Neural Networks |
| LSTM | Long Short-Term Memory |
| BIGRU | Bidirectional Gated Recurrent Unit |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| FDD | Fault Detection and Diagnostics |
| RMSE | Root Mean Square Error |
| AHU | Air Handling Unit |
| IoT | Internet Of Things |
| BEMS | Building Energy Management Systems |

## 

## TABLE OF CONTENTS

| **Chapter** | **TITLE** | **Page No** |
| --- | --- | --- |
|  | Certificate | ii |
|  | Acknowledgment | iii |
|  | Declaration | iv |
|  | **Abstract** | v |
|  | List of Figures | vi |
|  | List of Tables | viii |
|  | List of Acronym | xi |
| **1** | **Introduction** | 1 |
|  | 1.1 Introduction to Project work | 1 |
|  | 1.2 Objective of the Project | 1 |
|  | 1.3 Methodology adopted to satisfy the objective | 3 |
|  | 1.4 Architecture Diagram with a very brief description | 5 |
|  | 1.5 Organization of the Report | 7 |
|  |  |  |
| **2** | **Literature Survey** | 10 |
|  | 2.1 Literature Summary | 10 |
|  | 2.2 Summary: Drawbacks of Existing Approaches | 24 |
|  |  |  |
| **3** | **Proposed Method** | 26 |
|  | 3.1 Problem Statement and Objectives | 26 |
|  | 3.2 Explanation of:   * Architecture Diagram * Modules-Connectivity Diagram * Software and Hardware Requirements | 27 |
|  | 3.3 Modules and its Description | 33 |
|  | 3.4 Requirements Engineering   * Functional and Non-Functional | 35 |
|  | 3.5 Analysis and Design through UML   * Class Diagram * Sequence Diagram * Use Case Diagram * Activity Diagram   3.6 Testing | 37  42 |
|  |  |  |
| **4** | **Results and Discussions** | 44 |
|  | 4.1 Description of Dataset | 44 |
|  | 4.2 Detailed Explanation about the Experimental Results | 46 |
|  | 4.3 Significance of the Proposed Method with its Advantages | 52 |
|  |  |  |
| **5** | **Conclusion and Future Enhancements** | 54 |
|  | Summary of the project emphasizing the Objectives, Importance, Significance, Approach adopted, Results and Future Enhancements | 54 |
|  |  |  |
| **6** | **Appendices** | 61 |
|  | Sample code | 61 |
|  | **References** |  |
|  |  |  |

# CHAPTER 1

## INTRODUCTION

The subsequent segment elucidates the justification behind employing supervised learning-based technologies within the context of HVAC systems.

**1.1 Introduction to the project work**

In most of the buildings, for example, offices, schools, and homes, we utilize regular heating and cooling systems known as HVAC. It's important to note, though, that these HVAC (heating, ventilation, and air conditioning) systems aren't necessarily effective. They frequently squander a great deal of energy, which is a known drawback.HVAC systems are inefficient because they are difficult to set up and maintain. Due to the inefficiency of these standard systems, buildings lose a significant quantity of energy. Financial responsibilities that are needless result from this. We must facilitate the transition of people to more evolved, energy-efficient HVAC systems. Consider that updating your HVAC system might seem as straightforward as switching the light bulb. That's the degree of simplicity we are striving for. We hope to make it easy and simple for anyone to upgrade the HVAC system in their building. By doing this, we can help buildings use less energy, which not only saves money but also helps protect the environment. And that's something we can all get behind.

However, it goes past merely easing tasks for building managers and owners. For more people to be aware of these updated HVAC systems, we also must create awareness about them. This involves informing users about the positive aspects of AHUs and proving to them how straightforward it can be to switch.

**1.2.Objective of the project**

Through careful incorporation of innovative instruments, advanced sensor technologies, and extensive data analytics, this project intends to drastically improve the performance and efficiency of air conditioning, ventilation, and heating (HVAC) systems inside buildings. Fundamentally, the project intends to assure the efficient and steady operation of the HVAC systems, which in turn boosts occupant comfort, saves operating costs, and develops environmental sustainability. The structure used in modern houses incorporates systems for air conditioning, ventilation, and dining, all of which are crucial to maintaining indoor comfort and air purity. However, precise control over several elements, notably temperature, airflow dynamics, humidity levels, and energy usage, is essential for these systems to operate as planned. The project addresses the fundamental challenges and complexity involved in keeping HVAC systems operating at optimal efficiency, especially in several kinds of dynamic building environments. The appropriate positioning of advanced sensor networks throughout buildings is vital for accomplishing its objectives. Through continually monitoring and relaying critical environmental figures to centralized systems for analysis, these sensors serve as front-line data collectors. Stakeholders can optimize operational efficiency by obtaining an in-depth understanding of HVAC system performance by taking advantage of real-time data insights, that facilitate swift alterations and interventions.

Nevertheless, similar to null values in data sets, inaccuracies or mistakes could affect the reliability of data generated by sensors. The project employs protocols for data standardization to work around this obstacle. Standardization guarantees the uniformity and purification of data from different detectors and sources, enabling accurate and revealing evaluation. This approach lays a foundation for dependable anomaly detection and predictive maintenance methods alongside strengthening the data integrity. A key aspect of the approach adopted by the project is the design and enhancement of machine learning-powered advanced algorithms. Historical information sets that document incidents of anomalies or inaccuracy in the sensors are used to train these algorithms. The algorithms identify patterns and deviations in data streams via iterative methods of learning, enabling them to predict possible abnormalities before they materialize as catastrophic failures. In addition to decreasing operational risks, this preventive approach increases confidence in the system and resource allocation.

likewise, the investigation demonstrates why predictive analytics might enhance HVAC system efficiency. By taking advantage of contemporary data analytics insights, stakeholders can anticipate and effectively address servicing needs, leading to less downtime, increased system lifespan, and reduced energy usage. Besides strengthening operational robustness, this ability to predict enables HVAC systems to save a significant amount of revenue throughout their lifetime. By promoting conservation of the environment and energy efficiency, the effort serves environmental goals that go beyond brief operational advantages. Reduced carbon emissions, fewer resources used, and improved building sustainability ratings are all advantages of efficient HVAC systems. The project is meant to help enterprises achieve their sustainability goals while preserving the best possible indoor environmental quality for residents with the incorporation of intelligent HVAC management technological innovations.

The path to improved HVAC efficiency is defined by continual creativity, research, and collaboration across multiple disciplines. To deal with the evolving issues in building management, comprises taking steps to implement emerging methods and technologies. As sensor technology, AI, and data analytics continue to progress, the project is devoted to pushing beyond the limits of what is possible for optimizing sustainability and HVAC system performance. If everything is looked at, the project is an evolution in HVAC system management, driven by sustainability, efficiency, and creativity. Through the utilization of modern technologies and insights based on data, participants may not only increase operational efficiency but also build environments that are happier, more enjoyable, and sustainably responsible for generations to come. Using proactive execution, ongoing study, and strategic partnerships, the project intends to create fresh standards for HVAC system optimization and have an integral part in establishing a more sustainable future.

**1.3.Methodology adopted to satisfy the Objective**

A fundamental statistical method for modeling the relationship between a dependent variable and one or more independent variables is called linear regression. It is frequently employed for anticipating and analyzing the underlying correlations between variables in a range of sectors, such as economics, finance, engineering, and social sciences.

The dependent variable (outcome) can be predicted with only one independent variable (predictor) in a fundamental linear regression model. It is assumed that the two variables have a linear relationship, as is represented by a straight line. Typically, the equation for fundamental linear regression is written as follows:

y=mx+b

Where:

* y is the dependent variable (the variable being predicted)
* x is the independent variable (the predictor)
* m is the slope of the line (indicating the change in y for a one-unit change in xxx)
* b is the y-intercept (the value of y when x is 0)

The goal of linear regression is to find the best-fitting line that minimizes the difference between the observed values of the dependent variable and the values predicted by the model. This is typically done by minimizing the sum of the squared differences between the observed and predicted values, a method known as the least squares method.

Linear regression can also be extended to multiple linear regression when there is more than one independent variable. The equation for multiple linear regression is

y=b0+b1x1+b2x2+...+bnxn

Where-

* y is the dependent variable
* x1,x2,...,xn​ are the independent variables
* b0,b1,b2,..., bn are the coefficients (intercept and slopes) to be estimated

Support Vector Machines (SVMs) are powerful supervised learning models used for classification and regression tasks. They are particularly effective in high-dimensional spaces and when the number of features is much greater than the number of samples.

Here's a breakdown of SVMs and their key concepts

* **Classification vs. Regression**
  + SVMs can be used for both classification and regression tasks. In classification, SVMs aim to find the optimal hyperplane that separates different classes in the feature space. In regression, SVMs find the hyperplane that best fits the data while minimizing errors.
* **Hyperplane**
  + In SVMs, the goal is to find the hyperplane that best separates the data points into different classes. A hyperplane is a decision boundary that divides the feature space into regions associated with different classes.
* **Support Vectors**
  + Support vectors are the data points that lie closest to the decision boundary (hyperplane). These points play a crucial role in defining the hyperplane. Support vectors are used in the construction of the decision boundary, making SVMs memory efficient.
* **Margin**
  + The margin is the distance between the hyperplane and the nearest data point from each class. In SVMs, the goal is to maximize the margin, i.e., to find the hyperplane with the maximum margin that still separates the classes correctly. This helps improve the generalization of the model.
* **Kernel Trick**
  + One of the key features of SVMs is their ability to handle non-linear decision boundaries through the use of kernels. Kernels transform the input data into higher-dimensional spaces, where a linear decision boundary can be constructed. Common kernels include linear, polynomial, radial basis function (RBF), and sigmoid kernels.
* **Regularization Parameter (C)**
  + In SVMs, the regularization parameter CCC controls the trade-off between maximizing the margin and minimizing the classification error. A small CCC allows for a wider margin but may lead to more misclassifications, while a large CCC reduces the margin but may improve classification accuracy.

**1.4 Architectural Diagram with Brief Explanation**

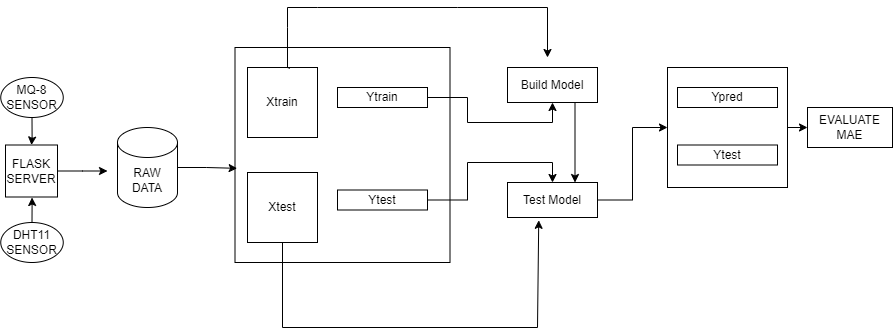
****

Fig 1. Visual Representation of the System Architecture Diagram

The Fig 1 demonstrates the architecture diagram of the project and the following are the components,

* **Sensor Data Acquisition**
  + **DH11 Sensor:** This sensor is used for measuring environmental conditions such as temperature and humidity. It is a basic, low-cost sensor suitable for capturing these specific types of data.
  + **MQ-8 Sensor:** This sensor detects hydrogen gas concentrations. It is typically used in applications that require monitoring of gas levels for safety or research purposes.
* **Data Transmission to Flask Server**
  + The data from the DH11 and MQ-8 sensors is transmitted to a Flask server. Flask is a lightweight web framework for Python that allows for easy setup of a web server. In this system, the Flask server handles the collection and initial processing of the raw sensor data.
* **Data Handling and Preparation**
  + **Raw Data:** The collected raw data is stored and then split into training and testing datasets. This step is crucial for preparing the data for machine learning model training and evaluation.
  + **Xtrain and Ytrain:** The Xtrain dataset contains the input features for training, while **Ytrain** contains the corresponding labels or outputs. This data is used to train the machine learning model.
  + **Xtest and Ytest:** Similarly, Xtest contains the input features for testing, and **Ytest** contains the actual labels used to evaluate the model's performance.
* **Model Training and Testing**
  + Build Model: In this phase, a machine learning model is built using the training data (Xtrain and Ytrain). The specific type of model (e.g., linear regression, decision tree, neural network) can vary depending on the problem being addressed.
  + Test Model: Once the model is trained, it is tested using the testing data (Xtest). This phase involves using the model to make predictions based on the test inputs.
* **Prediction and Model Evaluation**
  + Ypred: The predictions made by the model are represented as Ypred. These are the output values that the model predicts for the test input data (Xtest).
  + Evaluation Metrics: The model’s performance is evaluated using specific metrics:
    - Mean Absolute Error (MAE): This metric measures the average magnitude of errors between predicted and actual values. It gives an idea of the accuracy of the predictions.
    - Correlation Coefficient (R): This metric assesses the strength and direction of the relationship between predicted and actual values. A higher value indicates a better model performance.

The architecture diagram effectively outlines a workflow that starts with data collection from sensors and progresses through server-side processing, data preparation, model training, and evaluation. Each step in the process is crucial for ensuring that the final machine-learning model is accurate and reliable. This structured approach allows for systematic handling and analysis of sensor data, ultimately enabling predictive modeling and assessment of model performance using defined metrics.

### Conclusion

In conclusion, the architectural diagram of this IoT system exemplifies the convergence of hardware (sensors, microcontrollers), software (data processing, machine learning), and networking technologies (Wi-Fi, cloud computing). This integration enables real-time monitoring, predictive analytics, and actionable insights into environmental conditions, facilitating smarter decision-making and proactive management strategies across various industries including smart buildings, industrial automation, and environmental monitoring. As IoT technologies continue to evolve, these systems will play an increasingly vital role in shaping a more connected, efficient, and sustainable future.

**1.5 Organization of the Report**

#### Chapter 1: Introduction

* **Introduction to the Project Work**: This section provides an overview of the project, explaining its relevance and importance in the field. It sets the context for the problem being addressed and introduces the HVAC system and its components.
* **Objective of the Project**: Clearly states the goals of the project, such as improving energy efficiency, optimizing indoor climate control, or predicting HVAC system performance.
* **Methodology Adopted**: Describes the approach and techniques used to achieve the project objectives. This may include data collection, preprocessing, machine learning algorithms, and evaluation methods.
* **Architecture Diagram**: Presents a high-level architecture diagram of the system, showing the flow of data and interaction between different components (e.g., sensors, data processing, machine learning models).
* **Organization of the Report**: Outlines the structure of the report, providing a roadmap for the reader.

#### Chapter 2: Literature Survey

* **Summary of Manuscripts**: Provides a detailed review of existing research and literature related to HVAC systems, energy efficiency, and predictive maintenance. Summarizes findings from at least 10-15 relevant papers.
* **Drawbacks of Existing Approaches**: Discusses the limitations and gaps identified in the current state-of-the-art, which the proposed project aims to address.

#### Chapter 3: Proposed Method

* **Problem Statement & Objectives**: Defines the specific problem the project aims to solve and outlines the detailed objectives.
* **Explanation of Architecture Diagram**: Elaborates on the architecture diagram presented in Chapter 1, detailing each component and their interactions.
* **Modules – Connectivity Diagram**: Provides a diagram showing how different modules (sensors, data processing units, machine learning models) are interconnected.
* **Software and Hardware Requirements**: Lists the necessary software (e.g., programming languages, libraries) and hardware (e.g., sensors, microcontrollers) needed for the project.
* **Modules and their Description**: Describes each module in detail, explaining their functionality and role in the system.
* **Requirements Engineering**
  + **Functional Requirements**: Specifies the functions the system must perform.
  + **Non-Functional Requirements**: Includes performance metrics, usability, reliability, etc.
* **Analysis and Design through UML**: Uses UML diagrams to model the system:
  + **Class Diagram**: This shows the system's classes and their relationships.
  + **Sequence Diagram**: Illustrates the sequence of interactions between objects.
  + **Use Case Diagram**: Depicts the system's functionality and its interactions with users.
  + **Activity Diagram**: Represents the workflow of the system.

#### Chapter 4: Results and Discussions

* **Description of Dataset**: Provides details about the dataset used, including its source, structure, and preprocessing steps.
* **Detailed Experimental Results**: Presents the results obtained from the experiments, using graphs, tables, and screenshots to illustrate findings.
* **Significance of the Proposed Method**: Discusses the advantages and improvements achieved by the proposed method compared to existing approaches.

#### Chapter 5: Conclusion and Future Enhancements

* **Summary of the Project**: Summarizes the key points of the project, reiterating the objectives, methods, and significant results.
* **Future Enhancement**: Suggests potential improvements and future work that could be undertaken to extend the project.

#### Chapter 6: Appendices

* **Sample Code**: Includes relevant code snippets used in the project, restricted to 3-5 pages for brevity.

### References

* A comprehensive list of all references cited in the report, formatted according to the required style.

This compacted structure provides a concise and straightforward review of every chapter, guaranteeing that the report is well-structured and covers all pertinent project details.

## 

## CHAPTER 2

## LITERATURE SURVEY

This chapter includes a description and summary of current strategies, their advantages, results, and their shortcomings.

### 2.1. Summary Of Existing Approaches

Davide Borda, Mattia Bergagio, Massimo Amerio, Marco Carlo Masoero, Romano Borchiellini, and Davide Papurello[1] have proposed an innovative method for detecting anomalies in Heating, Ventilation, and Air Conditioning (HVAC) systems using sophisticated machine learning techniques. This method employs both supervised and semi-supervised learning approaches, leveraging the strengths of each to elevate detection accuracy. Their technique makes particular use of info gathered by Internet of Things (IoT) sensors that have been embedded in Air Handling Units (AHUs), which are crucial HVAC system components. These sensors keep an eye on some operating features, such as humidity, temperature, airflow, and the consumption of energy. The suggested method looks for deviations from normal operating settings, which may point to potential system weaknesses or inefficiencies, through examining this large dataset.

Using previously gathered information that was previously labeled given known anomalies, machine learning models undergo training using supervised learning. To manage the enormous amount of unlabeled data, the technique additionally employs semi-supervised learning, which improves its ability to learn in situations where labeled data is limited or absent. The usefulness of the models is meticulously evaluated by a variety of standards, such as accuracy, precision, recall, and F1 score, assuring an objective evaluation of their issues detection abilities. The technique not only indicates that IoT data might be used for real-time monitoring and diagnostics, but it also shows why there is a great deal of promise for enhancing the HVAC systems' dependability and operating efficiency. Early defect identification enables scheduled upkeep to be carried out, which lowers operating costs and downtime. In addition, this approach helps buildings grow more energy-efficient overall and is compliant with regulations and sustainable practices.

Overall, the proposed method by Davide Borda, Mattia Bergagio, Massimo Amerio, Marco Carlo Masoero, Romano Borchiellini, and Davide Papurello represents a significant advancement in the field of HVAC system management, offering a practical solution for enhancing system performance and reliability through the integration of cutting-edge machine learning techniques and IoT sensor data.

Farzad Dadras Javan, Italo Aldo Campodonico Avendano, Behzad Najafi, Amin Moazami, and Fabio Rinaldi[2] have proposed a novel methodology that combines physics-based energy simulation and machine learning to anticipate short-term electricity consumption in a Los Angeles warehouse under flexibility schemes. This innovative approach integrates multiple advanced techniques to enhance the accuracy and reliability of energy consumption predictions, thereby supporting more efficient energy management practices.The suggested method involves calculating houses correctly to duplicate different cooling conditions. In order to effectively reflect how the warehouse interacts with various climatic circumstances and operational techniques this modeling approach takes into account the physical characteristics of the warehouse, including its structural and thermal characteristics. The researchers may investigate how various cooling techniques affect the use of energy and find places for optimization by modeling these cooling scenarios.

Besides the simulations based on physics, the technique adds flexibility to events. These incidents illustrate how the warehouse's procedures for operation have changed, including the timing of cooling operations and the way energy-intensive equipment is used. The researchers can evaluate the adaptability of the warehouse's energy consumption to various functioning scenarios and make plans to improve this flexibility by incorporating these adaptability occurrences into the models.The researchers use machine learning techniques notably Random Forest and Artificial Neural Networks (ANN), to investigate the data collected by these scenarios. These mathematical models are excellent for estimating short-term power consumption based on simulated scenarios and flexibility events as they can handle complicated and high-dimensional datasets with ease. Using the use of ensemble learning, Random Forest averages the results of several decision trees to produce accurate predictions. To improve the accuracy of the predictions, ANN, on the other hand, applies its layered architecture to recognize nonlinear associations and trends in the data.

Besides contributing to increasing the accuracy of short-term electricity consumption forecasts, the integration of machine learning and physics-based modeling offers insights into the possible effects of different flexibility schemes. This dual approach offers an efficient tool for energy management in large-scale industrial settings by facilitating an in-depth investigation that mixes theoretical simulation with experimental methods based on data. Overall, the methodology proposed by Farzad Dadras Javan, Italo Aldo Campodonico Avendano, Behzad Najafi, Amin Moazami, and Fabio Rinaldi represents a significant advancement in the field of energy management. By leveraging the strengths of both physics-based simulations and machine learning, this approach provides a robust framework for predicting and optimizing electricity consumption in industrial warehouses, contributing to more sustainable and cost-effective energy usage practices.

Ali Ghahramani, Simin Ahmadi Karvigh, and Burcin Becerik-Gerber[3] have proposed a detailed methodology in their paper that involves the development and implementation of a flexible hybrid metaheuristic approach to optimize the energy usage of HVAC systems. This inventive approach aims to enhance the efficiency of HVAC systems, which are crucial for maintaining indoor environmental quality and energy management in buildings. To achieve the best possible energy usage, the approach suggested offers several intuitive techniques. Fundamentally, it makes use of an adjustable hybrid metaheuristic approach, which integrates the advantages of several optimization techniques to identify the best solutions. Metaheuristic methods are renowned for their capacity to resolve intricate optimization issues that are challenging to handle through traditional methods. Through the modification of these methodologies, the researchers aim to dynamically modify the procedure for optimization in view of real-time data and dynamic HVAC system conditions.

The implementation of evolutionary algorithms, especially the modified Strength Pareto Evolutionary Algorithm (SPEA), is an important component of this methodology. Natural selection provides the model for transformative algorithms, and it works particularly well for multi-objective optimization problems. By incorporating additional mechanisms that boost convergence speed and solution variety, the modified SPEA utilized in the present investigation improved upon the basic method. This makes it possible for the algorithm to properly explore the solution space and find great or nearly perfect alternatives for lowering energy consumption while preserving the HVAC system's planned performance levels. The adaptive nature of the hybrid metaheuristic approach ensures that the optimization process remains effective under varying operational conditions. This adaptability is crucial for HVAC systems, which often operate in dynamic environments where factors such as occupancy levels, external weather conditions, and internal heat loads can change rapidly. By continuously learning and adapting to these changes, the proposed methodology can maintain optimal performance and energy efficiency.

Also, the hybrid approach's employ of the modified SPEA enables an accurate assessment of many objectives. These goals usually involve minimizing energy use, cutting operating costs, and protecting occupants' temperature and air quality in the context of HVAC systems. By handling these competing goals and identifying a Pareto-optimal set of solutions, the evolutionary algorithm ensures that the system can achieve a balance that fulfills all essential requirements for performance.Overall, the methodology proposed by Ali Ghahramani, Simin Ahmadi Karvigh, and Burcin Becerik-Gerber represents an important advancement in the field of HVAC system optimization. By harnessing an adaptive hybrid metaheuristic approach and integrating advanced evolutionary algorithms like the modified SPEA, this research provides a reliable framework for improving the energy efficiency and operational performance of HVAC systems. This approach advances to more sustainable building operations but also supports the upholds goals of energy conservation and environmental protection.

Massieh Najafi, David M. Auslander, Peter L. Bartlett, Philip Haves, and Michael D. Sohn[4] propose an inventive methodology aimed at simplifying the strategy set and adapting to dynamic environmental changes through the implementation of Multi-Agent Reinforcement Learning (MARL). This approach represents a significant advancement in decision-making processes for complex systems, such as building energy management. The principal aim of this methodology is to refine the strategy set, making it more adaptable to changing environmental conditions, while concurrently ensuring that computational time scales linearly, even as the number of strategy options or players increases. This expandability is essential for real-world applications, where decision-making must be efficient and responsive to dynamic changes in the environment.

MARL is employed as the key approach for achieving these objectives. MARL is a sub-division of reinforcement learning that involves multiple agents collaborating and the environment to achieve a common purpose. By exploiting MARL, the researchers aim to develop a decision-making framework that can effectively coordinate the actions of multiple agents in response to changing environmental conditions.The methodology's primary element is the use of reduced rewards to determine how well players' activities performed. Discounted motivation gives a framework for valuing acts according to their impact over time, taking into consideration the current moment and the future. Via improving decision-making procedures via reduced rewards, the approach attempts to guarantee that player habits are strategically compatible with the system's overall objectives.

Moreover, the methodology emphasizes boosting processes for decision-making in dynamic settings, wherein conditions around us can shift spontaneously. The system can adapt to changing conditions and sustain peak performance over time via MARL for constant adjustments to these changes.Overall, the methodology proposed by Massieh Najafi, David M. Auslander, Peter L. Bartlett, Philip Haves, and Michael D. Sohn represents a substantial advancement in the field of decision-making for complex systems. By leveraging MARL and discounted rewards, this research provides a scalable and adaptable framework for optimizing decision-making processes in rapidly evolving environments, with potential applications in various domains, including building energy management and beyond.

Ying Guo, Josh Wall, Jiaming Li, and Sam West[5], developed a methodology for fault detection and diagnosis (FDD) in HVAC systems exploiting Hidden Markov Models (HMMs) in conjunction with clustering and data fusion techniques. The methodology starts by training HMMs to represent both normal and faulty operations of HVAC systems. The training process uses sensor data gathered from the HVAC systems, employing the Baum-Welch algorithm to optimize the parameters of the HMMs. This algorithm iteratively adjusts the model to maximize the likelihood of the observed data given the model, effectively capturing the typical behavior patterns of the system under normal and fault conditions. Once the HMMs are trained, fault detection is done by evaluating the prospect that new sensor data fits the trained HMM. Data with a low probability of fitting the model indicates a potential fault. To enhance the robustness of this detection, the process is repeated multiple times, generating multiple detection results. These results are then clustered using the K-Means algorithm, which groups the data points into clusters based on similarity. This clustering helps to identify consistent patterns that could potentially indicate certain types of faults.The approach includes a cluster merging the stage to improve fault detection reliability and precision. By making sure that closely related clusters are merged, clusters are merged based on an established distance ratio, which simplifies the problem-learning process. When applied to actual HVAC data, this combination of HMM-based detection and K-Means clustering showed notable gains in the precision of defect identification and diagnosis. A full approach for optimal FDD in HVAC systems is provided by the methodology, which blends model-based and data-based processes.

Sondes Gharsellaoui; Majdi Mansouri; Mohamed Trabelsi; Mohamed-Faouzi Harkat; and Shady S. Refaat[6] propose a novel approach for fault detection and diagnosis (FDD) in uncertain HVAC systems, introducing a multiscale Interval Principal Component Analysis (IPCA)-based machine learning (ML) methodology. Recognizing the uncertainties inherent in HVAC data, the methodology adopts interval-valued data representation to better handle these uncertainties. Leveraging advanced PCA techniques tailored for interval data, including Centers PCA (CPCA), Vertices PCA (VPCA), and Midpoints-Radii PCA (MRPCA), the methodology aims to extract relevant features from interval-valued HVAC system data using IPCA. The benefits of the proposed machine learning approach for FDD in HVAC systems, based on MSIPCA, are emphasized. It boasts a low misclassification rate of just 0.6 percent and significantly high accuracy rates of up to 99.4%. The method offers enhanced extraction of features with interval-valued encoded data and illustrates how to handle the uncertainties and nonlinearities present in HVAC systems in an efficient manner. Furthermore, the approach operates excellently in an assortment of operational scenarios, offering HVAC systems trustworthy fault detection and troubleshooting capacities. However, the paper also acknowledges limitations in addressing the nonlinear nature and uncertainties inherent in HVAC systems. Despite the promising results achieved by the developed MSIPCA-based methodology, there remains a gap in fully addressing these challenges, which could potentially affect the accuracy and robustness of fault detection and diagnosis in certain scenarios.

In particular, there is still potential for enhancement, especially when it comes to handling the nonlinear features of HVAC systems, yet the accuracy of the FDD of the heating system using MSIPCA-based SVM classifiers is amazing at 99.4%. Furthermore, the misclassification rate of only 0.6% shows the approach's efficacy as well as pointing to the need for further research to further enhance its performance in more complex working situations. In conclusion, the proposed MSIPCA-based machine learning approach represents a significant advancement in fault detection and diagnosis for HVAC systems, offering high accuracy, low misclassification rates, and effective handling of uncertainties. However, future research efforts should focus on addressing the nonlinear nature of HVAC systems to further improve the accuracy and robustness of fault detection and diagnosis methodologies.

G. Jahedi and M.M. Ardehali[7] are exploring how to make HVAC systems smarter using a self-tuning Proportional-Derivative (PD) controller integrated with Artificial Intelligence (AI). Their focus is on employing Artificial Neural Networks (ANNs), particularly Wavelet Neural Networks (WNNs), which are known for their rapid learning and efficiency. WNNs are adept at capturing complex patterns in data due to their ability to process information in both time and frequency domains, making them suitable for the dynamic environment of HVAC systems.To enhance system recognition and efficiency, Jahedi and Ardehali combine WNNs with an infinite impulse response (IIR) filter using their methodology. The IIR filter facilitates efficient processing of the sensor input, offering a comprehensive operating environment for WNNs. The goal with this combination is to improve the HVAC system's predictive skills so that it can react to changing internal demands and external situations more precisely. The PD controller's self-tuning feature guarantees that the system can adjust in real-time and get outstanding performance without any requirement for human oversight. The Least Mean Square (LMS) strategy, which aims to reduce mistakes by iteratively modifying the network's parameters, is used by the researchers to train the Wavelet Neural Networks. The WNNs have been successfully modified by the LMS algorithm, enabling them to absorb information from the data effectively and improve their accuracy over time. By reducing the consumption of energy while maintaining comfort, this method highlights the potential of AI-enhanced control methods for modern HVAC applications.

Kadir Amasyali, Mohammed Olama, and Aniruddha Perumalla[8] explore a machine learning-based approach to predict the aggregate flexibility of HVAC systems, which holds significant promise for optimizing energy usage and demand response management. The core of their research involves training machine learning models on historical data from HVAC systems, encompassing variables such as temperature, occupancy, and energy consumption patterns. This historical data serves as a rich dataset for the models to learn complex relationships and interactions that affect HVAC performance and flexibility. The process involves collecting a large amount of data from various HVAC systems and creating an extensive variety of operational scenarios. After that, this data is used to train machine learning models, which can employ neural networks, decision trees, or regression models. The models analyze the data's patterns to determine how different factors affect the system's capacity to adapt to changes in demand or outside circumstances. Through an understanding of these linkages, the models are able to forecast the amount of flexibility with which an HVAC system can modify its operations in order to meet fluctuating energy demands without compromising comfort.The HVAC equipment' projected flexibility may serve an essential part in demand response programs, that alter energy use in response to fluctuations in the grid's demands and goods. For example, the system may decrease consumption of energy during times of high demand, this assists in stabilizing the grid and saving money. On the other together, if demand is low, the system has more freedom for operation. This machine learning approach supports greater energy management objectives by facilitating more dynamic and informed decision-making processes, which in turn optimizes the agility and effectiveness of HVAC systems.

TABLE 1. Summary of the Existing Approaches

| SNO | Ref. No | Methodology | Advantages | Research Gaps | Results |
| --- | --- | --- | --- | --- | --- |
| 1 | [1] | The study investigates anomaly detection in HVAC systems using supervised and semi-supervised machine learning, leveraging data from IoT sensors in AHUs for fault detection potential. | 1. LSTM and BIGRU achieved a 7% error rate, effectively spotting faults. Metrics like MAE, MAPE, and RMSE assessed accuracy.  2. Semi-supervised models, starting with labeled data, detected anomalies by grouping based on density, showcasing potential in fault detection. | Research gaps: assessing model robustness, scalability, transferability, real-time implementation, and addressing model drift for HVAC fault detection effectiveness. | 86.33% achieved through semi-supervised models. |
| 2 | [2] | Integrating physics-based energy simulation with machine learning to forecast short-term electricity usage in a Los Angeles warehouse. Includes building modeling, simulating cooling scenarios, introducing flexibility events, and employing algorithms such as Random Forest and ANN. | 1.  The result section evaluates physics-based simulations and predictive modeling for warehouse electricity usage and demand flexibility.  2.It analyzes the impact of cooling scenarios, set point adjustments, and achieves a load forecasting accuracy of 87%. | Limited comparison between ML models, generalizability to diverse climates, uncertainty about model adaptability to evolving conditions, and lack of real-world validation. | MAPE score of 3.26% |
| 3 | [3] | Developing and implementing an adaptive hybrid metaheuristic approach for optimizing HVAC system energy usage, utilizing evolutionary algorithms like modified strength Pareto algorithm. | 1. Emphasizes the necessity for enhanced control policies in HVAC systems.  2. Advocates for data-driven approaches utilizing historical operational data to optimize system performance. | Further investigation is needed into the adaptive hybrid metaheuristic algorithm's applicability across various building types and HVAC configurations. Additionally, integrating an occupancy prediction module to address dynamic occupancy impacts on HVAC energy consumption is advocated. | NRMSE value of 0.0396 |
| 4 | [4] | By implementing MARL, the aim is to simplify strategy set and adapt to dynamic environmental changes while ensuring linear computational time, even with expanded options or player numbers. This methodology utilizes discounted rewards to optimize decision-making processes based on action effectiveness. | 1.Methodology's adaptability to diverse weather conditions enhances technical robustness.  2.This attribute ensures efficacy across varying environmental contexts, enhancing practical applicability. | The research gap identified in the paper pertains to the limitations of existing methodologies in optimizing building energy costs within the context of multi-agent systems. | MAPE score of 4.2%. |
| 5 | [5] | The methodology involves developing a statistical machine learning-based approach for fault detection and diagnosis (FDD) in HVAC systems. It utilizes Hidden Markov Models (HMMs) trained on fault data to detect anomalies. Clustering and data fusion techniques are applied to enhance fault detection reliability. Real-world tests demonstrate the effectiveness of the approach for HVAC FDD applications. | 1. Through the integration of Hidden Markov Models and data fusion methods, the system can accurately identify and diagnose faults in HVAC systems. | While the paper discusses preliminary results, the integration of the proposed approach into real-time systems for continuous monitoring and fault detection in diverse HVAC systems could be explored further. | Load forecasting accuracy achieved was 87%. |
| 6 | [6] | The methodology proposed in the paper introduces a multiscale Interval Principal Component Analysis (IPCA)-based machine learning (ML) approach for fault detection and diagnosis (FDD) in uncertain HVAC systems. It begins by acknowledging the uncertainties inherent in HVAC data and adopts an interval-valued data representation to better handle such uncertainties. Leveraging advanced PCA techniques tailored for interval data, including Centers PCA (CPCA), Vertices PCA (VPCA), and Midpoints-Radii PCA (MRPCA), the methodology aims to extract relevant features from interval-valued HVAC system data using IPCA. | The advantages of the proposed MSIPCA-based machine learning approach for fault detection and diagnosis in HVAC systems include significantly high accuracy rates (such as 99.4% achieved), low misclassification rates (only 0.6% observed), effective handling of uncertainties and nonlinearities inherent in HVAC systems, improved feature extraction through interval-valued data representation, and robust performance under various operating conditions. | While the developed multiscale interval principal component analysis (MSIPCA)-based machine learning (ML) method demonstrates promising results in diagnosing uncertain HVAC systems, it acknowledges limitations in addressing the nonlinear nature and uncertainties inherent in these systems. | The accuracy achieved for FDD of the heating system using MSIPCA-based SVM classifiers reaches 99.4%, with a misclassification rate of only 0.6%. |
| 7 | [7] | The methodology is making HVAC systems smarter by using a self-tuning PD controller with AI. Focus on Wavelet Neural Networks (WNNs) for quick, efficient learning. Combining WNNs with an Infinite Impulse Response (IIR) filter aims to enhance system identification and performance. It trains the network using the Least Mean Square (LMS) algorithm to minimize errors. | By integrating AI and advanced control strategies like Wavelet Neural Networks (WNNs) and Infinite Impulse Response (IIR) filters, HVAC systems can dynamically adjust to varying conditions, optimizing energy usage. This leads to significant reductions in energy consumption, which is crucial for both cost savings and resource conservation. Improved energy efficiency directly translates to longer battery life in electric vehicles and lower fuel consumption in traditional vehicles. | By integrating AI and advanced control strategies like Wavelet Neural Networks (WNNs) and Infinite Impulse Response (IIR) filters, HVAC systems can dynamically adjust to varying conditions, optimizing energy usage. This leads to significant reductions in energy consumption, which is crucial for both cost savings and resource conservation. Improved energy efficiency directly translates to longer battery life in electric vehicles and lower fuel consumption in traditional vehicles. | MAPE score of 3.26% |
| 8 | [8] | The methodology employed by Kadir Amasyali, Mohammed Olama, and Aniruddha Perumalla centers on leveraging machine learning to predict the aggregate flexibility of HVAC systems. Initially, extensive historical data from various HVAC systems are collected, encompassing key variables such as temperature, occupancy, and energy consumption patterns. This rich dataset allows for capturing the diverse and complex interactions that influence HVAC performance. | The primary advantage of this machine learning-based approach lies in its ability to dynamically optimize energy usage and enhance demand response management for HVAC systems. By understanding and predicting the flexibility of these systems, the models enable more efficient operation, reducing energy consumption during peak demand times and thus stabilizing the grid. This adaptability not only leads to cost savings but also improves overall energy management. | Despite promising outcomes, research gaps remain. Variability in data quality affects model accuracy and reliability. Ensuring consistent data collection across different HVAC systems is challenging. The models' ability to generalize across various building types and climates also needs further exploration. | R-squared Error of 0.85 |
| 9 | [9] | The methodology of the paper involves creating a pilot testbed in three rooms at Chun-Ang University, each equipped with a gateway to manage HVAC and lighting systems. IoT-based devices, including smart switches, remote controllers, and sensor boxes, were used to measure environmental parameters. The system utilized Raspberry Pi3 and Zigbee for communication between the MCU and the gateway. User applications were developed for system management. The testbed showed a significant reduction in energy consumption, validating the system's efficiency and potential for broader application. | The I-HVAC&R system offers several advantages, including significant energy savings, with reductions of 9.3%, 10.6%, and 13.7% for each tested room. It provides flexibility for administrators to set energy management policies that balance efficiency and user comfort. The system's comprehensive data collection enables precise environmental monitoring, and its automation capabilities optimize the operation of HVAC and lighting, enhancing overall building energy efficiency. | Despite its advantages, the IoT-based HVAC and Lighting (I-HVAC&L) system has several drawbacks. The initial cost of implementing the system is significant due to the investment in IoT devices, sensors, and infrastructure. The technical complexity of integrating and maintaining the system requires specialized knowledge, which can be a barrier for some users. Additionally, the reliance on internet connectivity raises concerns about data security and privacy. Lastly, while effective in a controlled test environment, scaling the system to larger buildings or multiple locations can present additional challenges. | not specified |
| 10 | [10] | Effective project management methodologies such as Agile, Lean principles, and adaptability to evolving requirements and feedback incorporation are crucial for ensuring the successful execution of energy efficiency projects in HVAC systems. Agile methodology enables teams to iteratively develop solutions through sprints, fostering collaboration and quick adaptation to changing stakeholder needs and emerging information. Meanwhile, Lean principles focus on eliminating waste and optimizing processes, thereby enhancing efficiency and reducing costs in project workflows. Additionally, the ability to adapt to evolving requirements ensures that projects remain aligned with organizational goals despite fluctuating conditions such as regulatory changes or market shifts. | The I-HVAC&R system offers several advantages, including significant energy savings, with reductions of 9.3%, 10.6%, and 13.7% for each tested room. It provides flexibility for administrators to set energy management policies that balance efficiency and user comfort. The system's comprehensive data collection enables precise environmental monitoring, and its automation capabilities optimize the operation of HVAC and lighting, enhancing overall building energy efficiency. | The research gaps identified include a limited understanding of the long-term impact of BEMS on energy efficiency and overall building performance. There is also a need for further studies on the integration of BEMS with other building management systems and the associated challenges. More research is required to understand the influence of occupant behavior on the effectiveness of energy-saving measures and how to best modify these behaviors. | not specified |

**2.2. Summary: Drawbacks of Existing Approaches**

The strategies to better HVAC systems and energy management demonstrate creative ideas, but there are substantial hurdles and research gaps that must be addressed for practical implementation and improvement in the future. Borda et al. use LSTM and BIGRU models for anomaly detection in HVAC systems, with both supervised and semi-supervised learning, but there are still challenges in predicting scalability and durability across different HVAC types, as well as the feasibility of real-time implementation due to the need for rapid response to anomalies. Dadras Javan et al.'s integration of machine learning and physics-based energy simulations predicts short-term electricity usage in warehouses with 87% accuracy, although there is a lack of extensive model comparisons and validation under diverse climatic conditions. This demands larger studies to assess scalability across various kinds of buildings and geographical areas. Ghahramani et al.'s hybrid metaheuristic approach for maximizing HVAC energy efficiency faces scalability and complexity issues in real-world applications, requiring precise performance measurements and validation studies across different buildings and operational scenarios.

Najafi et al. explore Multi-Agent Reinforcement Learning (MARL) for dynamic building energy management, but its adoption in real-time multi-agent situations is hindered by scalability and computational intensity issues, necessitating extensive validation in diverse building environments. Guo et al. use Hidden Markov Models (HMMs) for fault detection in HVAC systems, which are sensitive to data quality and interpretability, highlighting the need for improved fault detection robustness and enhanced data quality management practices. Böhme et al. employ machine learning techniques like RNNs and LSTMs to improve the energy efficiency of automotive HVAC systems, but ensuring model reliability and flexibility in different transportation environments requires testing in a range of vehicle operating conditions.

Jahedi and Ardehali utilize Wavelet Neural Networks (WNNs) for AI-enhanced HVAC control, emphasizing the importance of real-time adjustment capabilities and adaptability for practical implementation in dynamic HVAC scenarios. Amasyali et al. focus on machine learning models for forecasting HVAC system performance to support energy management, but these models must be verified in real-life situations to ensure they minimize energy usage under changing demand conditions. Hyeonwoo Jang et al. discuss the IoT-driven HVAC and Lighting (I-HVAC&L) system, which, despite its benefits, faces high initial implementation costs and technical complexity, requiring specialized expertise and simplified setup and operation processes.

Lastly, Wisdom Ebirim et al. identify research gaps including a limited understanding of the long-term impact of Building Energy Management Systems (BEMS) on energy efficiency and overall building performance. Further studies are needed on BEMS integration with other systems and the influence of occupant behavior on energy-saving measures. In conclusion, while these methodologies offer creative ways to enhance HVAC performance and energy efficiency, challenges related to model robustness, scalability, real-time implementation feasibility, and validation across varied operational contexts must be addressed for their practical application and advancement.

## CHAPTER 3

## PROPOSED METHOD

**3.1 Problem Statement and Objectives of the Project**

**Problem Statement**

Temperature and humidity are merely two examples of the outside factors that affect the quantity of energy buildings and structures use. The arrival of Internet of Things (IoT) devices has rendered it feasible to continuously monitor these variables in real-time, yielding a wealth of data that may be employed for predicting power consumption. Making precise forecasts from this data analysis still presents a hurdle, though. Traditional methods of predicting energy use frequently come short in taking into consideration the intricate, nonlinear relationships that take place between surroundings and power demand. As a result, reliable predictive models that make use of IoT data are needed in order to accurately predict the consumption of electricity.

**Objective**

The primary objective of this project is to develop predictive models for power consumption using data collected from IoT devices. The specific objectives are:

* **Data Collection**
  + Set up IoT devices to continuously monitor and collect data on indoor temperature, outdoor temperature, indoor humidity, and outdoor humidity.
  + Ensure data quality and completeness by implementing appropriate data collection and preprocessing techniques.
* **Data Preprocessing**
  + Clean and preprocess the collected data to handle missing values, outliers, and noise.
  + Normalize or standardize the data to facilitate the application of machine learning algorithms.
* **Model Development**
  + Develop a linear regression model to understand and quantify the relationship between environmental factors and power consumption.
  + Develop a Support Vector Machine (SVM) model to capture potential non-linear relationships and improve prediction accuracy.
* **Model Evaluation**
  + Evaluate the performance of the linear regression and SVM models using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
  + Compare the performance of the two models to determine which provides more accurate and reliable predictions.
* **Optimization and Tuning**
  + Optimize the models by tuning hyperparameters to enhance their predictive performance.
  + Implement cross-validation to ensure the models' robustness and generalizability.
* **Deployment and Implementation**
  + Deploy the best-performing model for real-time prediction of power consumption.
  + Develop a user interface or dashboard to visualize the predictions and provide actionable insights for energy management.
* **Impact Analysis**
  + Analyze the impact of the predictive model on energy management practices.
  + Assess the potential cost savings and efficiency improvements resulting from the implementation of the model.

**3.2 Explanation of Architecture Diagram, Modules-Connectivity Diagram, and Software and Hardware Requirements**

**3.2.1 Architecture Diagram**

### This architecture design shows a sophisticated Internet of Things (IoT) system that makes use of microcontrollers, networking infrastructure, cloud or local server resources, and machine learning algorithms to collect, process, examine, and visualize data from multiple sensors. Let's explore each component along with how they work within this system in greater depth.

### Sensors and ESP Controllers

### The DHT11 sensor, which monitors temperature and humidity, and the MQ-8 sensor, which detects carbon monoxide and combustible gas levels, are the two primary sensor types in this Internet of Things system. These sensors perform a critical role in the real-time collection of environmental data, providing crucial data concerning atmospheric conditions and indoor air quality.An ESP controller is connected to each sensor. These controllers function as an interface between the sensors and the wider Web of Things network. They are based on the ESP8266 or ESP32 microcontroller units. Gathering sensor data, processing it locally if necessary, and relaying it over Wi-Fi to the following node in the network—in this case, a Wi-Fi router—are their primary responsibilities.

### Wi-Fi Router

### To allow wireless communication between the ESP controllers and other wireless networks, such as the Internet, the Wi-Fi router acts as a central hub within the local network. It enables seamless interaction and transfer of information between the local server infrastructure or the cloud and the sensors/controllers.

### Cloud or Local Server

Data from the ESP controllers, transmitted via the Wi-Fi router, is received and processed by a cloud-based or local server. This server serves several critical functions in the IoT ecosystem:

* **Data Reception and Storage:** Upon receiving data packets from the ESP controllers, the server stores this information in a database. This database serves as a repository for all incoming sensor data, enabling historical analysis, trend identification, and long-term storage.
* **Data Processing:** The server processes the raw sensor data to extract meaningful insights and actionable information. This involves various data processing techniques such as filtering, aggregation, and normalization to ensure data quality and consistency.
* **Machine Learning Integration:** One of the advanced capabilities of this IoT system is its integration with machine learning algorithms. These algorithms, deployed on the server, analyze the processed data to uncover patterns, anomalies, correlations, or predictive models. For instance, Support Vector Machines (SVM) or Linear Regression (LReg) algorithms can be applied to predict air quality trends based on historical sensor data.

### Data Processing and Machine Learning

The processed data undergoes rigorous analysis through machine learning algorithms. These algorithms leverage the computational power of the server to perform tasks such as:

* **Anomaly Detection:** Identifying abnormal readings or events that may indicate potential issues such as gas leaks or environmental hazards.
* **Predictive Analytics:** Forecasting future trends based on historical data patterns, enabling proactive measures for maintenance, resource allocation, or environmental control.
* **Optimization:** Recommending optimizations in energy usage, HVAC systems, or indoor air quality management based on data-driven insights.

### Output and Visualization

The insights generated by the machine learning algorithms are then presented through output and visualization components. These components play a crucial role in making complex data understandable and actionable for stakeholders

1. **Real-time Dashboards:** Interactive dashboards provide real-time updates on environmental conditions, sensor readings, and predictive analytics. They allow users to monitor trends, set alerts for threshold breaches, and make informed decisions promptly.
2. **Reports and Alerts:** Reports summarize key findings, trends, and recommendations based on the analysis performed. Alerts can be configured to notify relevant personnel or trigger predefined responses in case of critical events.
3. **Visualization Tools:** Graphs, charts, heatmaps, and geographical representations visually depict data insights. These visual aids enhance comprehension and facilitate quick decision-making based on the current state of environmental parameters.

### System Integration and Benefits

* **Scalability:** The cloud-based infrastructure ensures scalability, allowing the system to handle large volumes of sensor data efficiently. This scalability is crucial as IoT deployments often expand to cover larger geographical areas or incorporate additional sensors over time.
* **Cost Efficiency:** By leveraging cloud services, the system reduces upfront infrastructure costs and maintenance efforts associated with traditional on-premises deployments. It also allows for flexible resource allocation based on demand fluctuations.
* **Data Security:** Robust security measures, including data encryption, access controls, and regular security audits, safeguard sensitive sensor data against unauthorized access or cyber threats.
* **Operational Insights:** The insights gained from continuous data monitoring and analysis empower organizations to optimize operational efficiency, ensure regulatory compliance, and enhance environmental sustainability efforts.

### Conclusion

In summary, the Fig 2 representing the IoT system's architectural design demonstrates the way hardware (sensors, microcontrollers), software (data processing, machine learning), and networking technologies (Wi-Fi, cloud computing) are convergent. Real-time monitoring, predictive analytics, and actionable insights into environmental conditions are made feasible by this integration, which helps a variety of industries, including industrial automation, smart buildings, and environmental monitoring, make improved choices while implementing proactive management strategies. These systems will become more and more essential for establishing a more connected, useful, and sustainable future as IoT technologies grow.

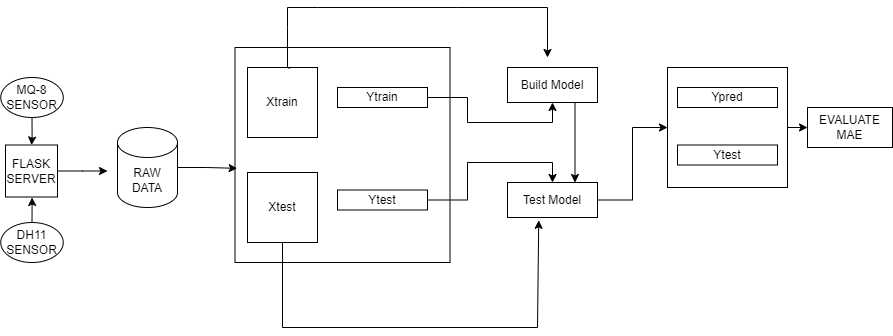
****

Fig 2.Machine Learning Visual Representation of the System Architecture Diagram

**3.2.2 Modules-Connectivity Diagram**

**Module Connectivity Diagram**

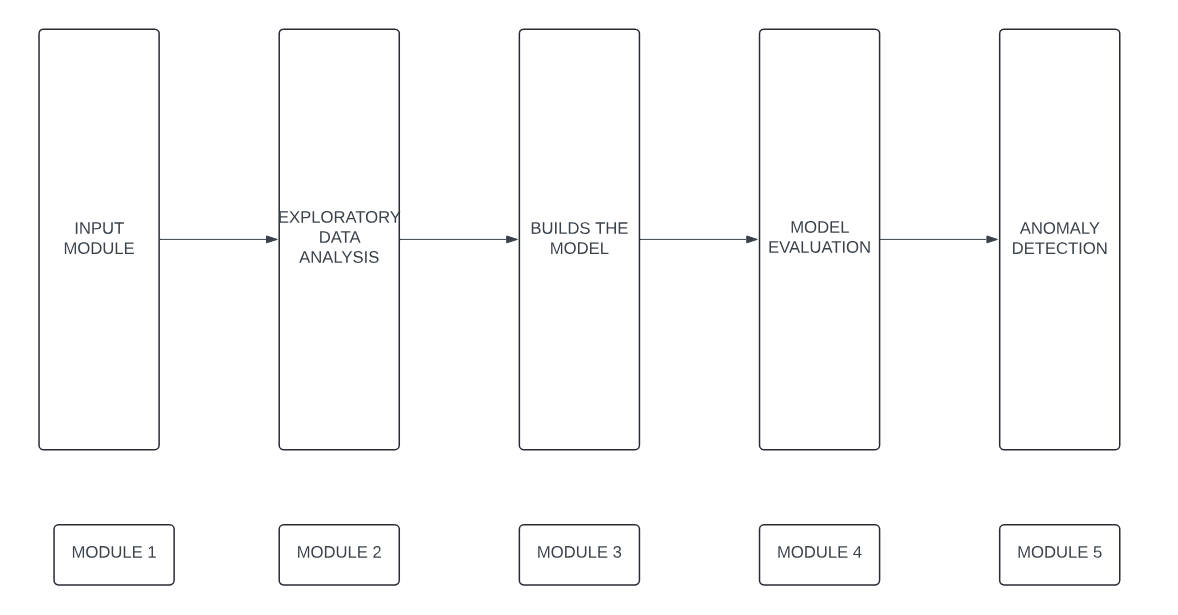
****

Fig 3. Module Connectivity Diagram

### The moule conectivity diagram represented as in Fig 3 depicts the The input portion of the module, exploring analysis of data, creating the model, evaluating the model, and anomaly detection are the five separate modules that make up the architecture diagram, demonstrating the organized process of data processing and evaluation. Each component is essential to turning unprocessed IoT sensor data into insights that can be put into action by applying cutting-edge machine learning techniques. The first module is responsible for taking the input data, the second module facilitates exploratory data analysis, and the third module builds the model. Finally, the fourth and fifth module are collectively responsible for model evaluation and anomaly detection.

**3.2.3 Software And Hardware Requirements**

**Hardware Requirements**

* IoT sensors (DHT11 for temperature and humidity, MQ-8 for gas concentrations).
* Microcontrollers or single-board computers (e.g., Arduino, Raspberry Pi) to interface with sensors and collect data.
* Storage device (e.g., SD card, USB drive) for storing collected data temporarily.
* A computer or server with sufficient processing power and memory to handle large datasets.
* A computer or server with significant computational power, potentially equipped with GPUs for faster model training.
* Continuously running system for real-time anomaly detection, potentially requiring embedded systems or IoT gateways for on-site processing.

**Software Requirements:**

* Firmware or code to interface with sensors and collect data (platforms like Arduino IDE).
* Libraries for sensor data handling and communication (e.g., Adafruit DHT library for DHT11 sensor, MQTT for data transmission).
* Scripts or applications to convert raw sensor data into Excel files or other suitable formats.
* Data analysis tools such as Python (with libraries like Pandas, NumPy, Matplotlib, and Seaborn), R, or Jupyter Notebooks for interactive analysis.
* IDEs or text editors for coding and scripting.
* Statistical packages for data cleaning, manipulation, and visualization.
* Machine learning frameworks such as Scikit-learn, TensorFlow, and PyTorch, for implementing algorithms like SVM, Linear Regression, Lasso Regression, etc.
* Development environments for coding and testing models.
* Tools for feature selection, hyperparameter tuning.
* Statistical analysis software for evaluating model performance using metrics like accuracy, precision, recall, F1 score.
* Visualization tools such as Matplotlib and Seaborn for visualizing model predictions and performance metrics.
* Algorithms and techniques for anomaly detection like Isolation Forests, One-Class SVM, or other statistical methods.
* Integration with data visualization tools for displaying anomalies and alerts.

**3.3 Modules and its Description**

The input portion of the module, exploring analysis of data, creating the model, evaluating the model, and anomaly detection are the five separate modules that make up the architecture diagram, demonstrating the organized process of data processing and evaluation. Each component is essential to turning unprocessed IoT sensor data into insights that can be put into action by applying cutting-edge machine learning techniques.

### Module 1: Input Module

The system's initial component, the Input module, is in control of collecting data from the Internet of Things sensors such as the DHT11 and MQ-8. These sensors maintain an eye on several environmental factors, including humidity, temperature, and gas concentrations. These sensors collect data, which is subsequently transformed into Excel files for further analysis. This module makes sure that raw sensor data is reliably captured and made available in an analysis-ready format. This module's usefulness is found in its capacity to arrange and collect data in a fast way, which lays the basis for later stages of analysis. The entire analytical process would be inaccurate without precise data accumulating alongside suitable formatting.

**Module 2: Exploratory Data Analysis (EDA)**

### The Exploratory Data Analysis module takes a central role following data collection, focusing on cleaning and preparing the data for modeling. Taking care of any erratic or null values that may cloud the analysis is part part this. EDA is necessary for understanding the data's basic structure, finding trends, and guaranteeing data quality before supplying it to machine learning models. This process not only raises the standard of the data but also helps in discovering initial trends and conclusions that could influence the choice of modeling techniques. A comprehensive examination of data distributions, variable relationships, and outlier identification are all phases in the EDA process that may affect the manner in which predictive models work.

### Module 3: Build the Model

Algorithms such as Support Vector Machine (SVM), Linear Regression, and Lasso Regression are used to construct machine learning models in the Establishes Model module. The kind of data and the specific problem at hand determines which algorithms are best. In this module, historical data is utilized to train the models so they can identify trends and predict results. The process of constructing the model is critical since it impacts the accuracy and dependability of forecasts. This module makes sure the models are appropriate to capture the intricacies of the data by identifying the right ways and tweaking the settings they have. To maximize model performance, this phase requires procedures like feature selection, cross-validation, and hyperparameter modification.

### Module 4: Evaluate the Model

Using a range of statistical approaches, the model evaluation module analyzes the machine learning models' efficiency. It includes evaluating model performance and figuring out the way well the models predict fresh data through the use of scatter plots and histograms. Evaluation criteria that measure model effectiveness include accuracy, precision, recall, and F1 score. This module makes sure that only the most accurate and dependable models are put into effect for use in reality by thoroughly assessing the models. This lesson takes a lot of visualization tools, allowing for obvious insights into model performance and aid in finding areas in demand for expansion.

### Module 5: Anomaly Detection

The objective of the anomaly identification module is to identify anomalies or outliers in the data that may point to issues like environmental deviations or malfunctioning equipment. This is done by using advanced methods such as Isolation Forests, which help detect anomalies in high-dimensional data, in addition to visual examination techniques. To keep the monitored systems secure and trustworthy, anomaly detection is important. This module helps in avoiding potential breakdowns and guarantees the system's continuing smooth functioning by quickly finding and fixing inconsistencies.

### Overall Workflow

The Input Modules, typically gather and save raw information from sensors, which is where the entire functioning starts. Following that, the EDA module purifies and examines the data to make sure it meets the requirements for modeling. In the Model Building module, the cleaned data can be used to create predictive models. These models are then assessed for accuracy and performance in the Model Evaluation module. Last but not least, the Anomaly Detection module keeps an eye on the data continuously to spot any outliers from the usual.

This architecture diagram shows a complete system that uses machine learning to process and analyze information collected by Internet of Things sensors to identify patterns and acquire insights. The modular architecture guarantees that every stage of the pipeline for data processing—from cleaning and data gathering to creating, evaluating, and operational efficiency.

**3.4 Requirements Engineering**

Requirements engineering involves gathering, documenting, and managing the requirements of a system to ensure that it meets the needs and expectations of its stakeholders. For your project, the requirements can be categorized into functional and non-functional requirements.

**1. Functional Requirements**

Functional requirements describe the specific behavior or functions of the system. For your project, these include:

* **Data Collection**
  + **FR1**: The system shall collect indoor temperature data from IoT sensors.
  + **FR2**: The system shall collect outdoor temperature data from IoT sensors.
  + **FR3**: The system shall collect indoor humidity data from IoT sensors.
  + **FR4**: The system shall collect outdoor humidity data from IoT sensors.
  + **FR5**: The system shall ensure continuous data collection and storage in a database.
* **Data Preprocessing**
  + **FR6**: The system shall clean the collected data to handle missing values.
  + **FR7**: The system shall identify and remove outliers from the data.
  + **FR8**: The system shall normalize or standardize the data to prepare it for machine learning algorithms.
* **Model Development**
  + **FR9**: The system shall develop a linear regression model to predict power consumption based on environmental data.
  + **FR10**: The system shall develop a Support Vector Machine (SVM) model to predict power consumption based on environmental data.
* **Model Evaluation**
  + **FR11**: The system shall evaluate the performance of the linear regression model using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.
  + **FR12**: The system shall evaluate the performance of the SVM model using metrics such as MAE, MSE, and R-squared.
  + **FR13**: The system shall compare the performance of the linear regression and SVM models.
* **Optimization and Tuning**
  + **FR14**: The system shall optimize the linear regression model by tuning its hyperparameters.
  + **FR15**: The system shall optimize the SVM model by tuning its hyperparameters.
  + **FR16**: The system shall implement cross-validation to ensure the robustness of the models.
* **Deployment and Implementation**
  + **FR17**: The system shall deploy the best-performing model for real-time power consumption prediction.

**2. Non-Functional Requirements**

Non-functional requirements describe the system's performance characteristics and constraints. For your project, these include

* **Performance**
  + **NFR1**: The system shall process and update predictions in real-time
  + **NFR2**: The system shall handle data from up to 3 IoT sensors simultaneously without performance degradation.
* **Reliability**
  + **NFR3**: The system shall have an uptime of 99.9% to ensure continuous data collection and prediction.
  + **NFR4**: The system shall provide accurate predictions with an error margin of less than 5%.
* **Scalability**
  + **NFR5**: The system shall be scalable to accommodate additional IoT sensors and increased data volume without requiring significant reconfiguration.
* **Usability**
  + **NFR6**: The user interface or dashboard shall be user-friendly and accessible to users with minimal technical expertise.
  + **NFR7**: The system shall provide clear and actionable insights in a visually appealing manner.
* **Security**
  + **NFR8**: The system shall ensure secure data transmission between IoT sensors and the central server using encryption.
  + **NFR9**: The system shall implement authentication and authorization mechanisms to prevent unauthorized access.
* **Maintainability**
  + **NFR10**: The system shall be designed in a modular manner to facilitate easy updates and maintenance.
  + **NFR11**: The system shall include comprehensive documentation for future developers and users.
* **Compliance**
  + **NFR12**: The system shall comply with relevant data privacy regulations and standards.

By addressing these functional and non-functional requirements, your project will be well-equipped to develop an effective and reliable predictive model for power consumption using IoT data.

**3.5 Analysis And Design Through UML**

**3.5.1 UML Class Diagram**

The Fig 4 represents the class diagram for our HVAC system project includes essential classes: Device, CleanedData, machine learning models (SVMModel, LinearRegressionModel, LassoRegressionModel), EvaluationResult, ModelEvaluator, and AnomalyDetector. These classes manage data collection, processing, model training, evaluation, and anomaly detection, ensuring optimal HVAC system performance and reliability.A UML class Diagram talks about the various dependent and independent classes of the architecture diagram.A Class Diagram is highly resourceful when it come sto understanding the intricacies of the project to be talked about.The class Digaram talks about various classes such as Device, CleanedData, machine learning models (SVMModel, LinearRegressionModel, LassoRegressionModel), EvaluationResult, ModelEvaluator, and AnomalyDetector.

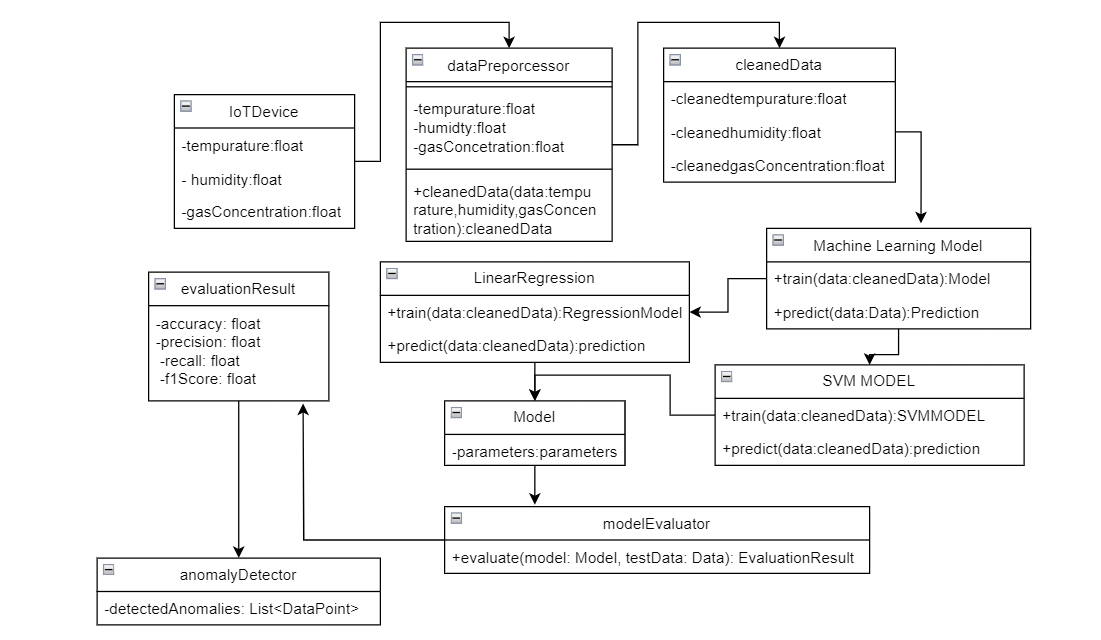


Fig 4. UML CLASS DIAGRAM

**3.5.2 UML Sequence Diagram**

The Fig 5 represents the sequence diagram illustrates the data flow from IoT sensors to the Flask server and ML model. IoT sensors collect data and send it to the Flask server. The server then sends this data to the ML model for processing. The ML model analyzes the data and returns predictions to the Flask server.

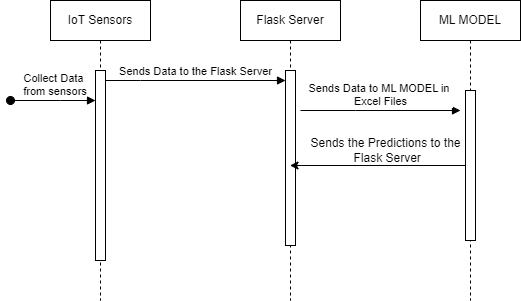


Fig 5. UML Sequence Diagram

**3.5.3 UML Use Case Diagram**

The Fig 6 represents the model training process is centrally managed within the system, ensuring robust analysis. Administrators oversee the entire process, from managing data inputs to supervising model training, ensuring the system's accuracy and reliability in detecting anomalies and maintaining overall data integrity.

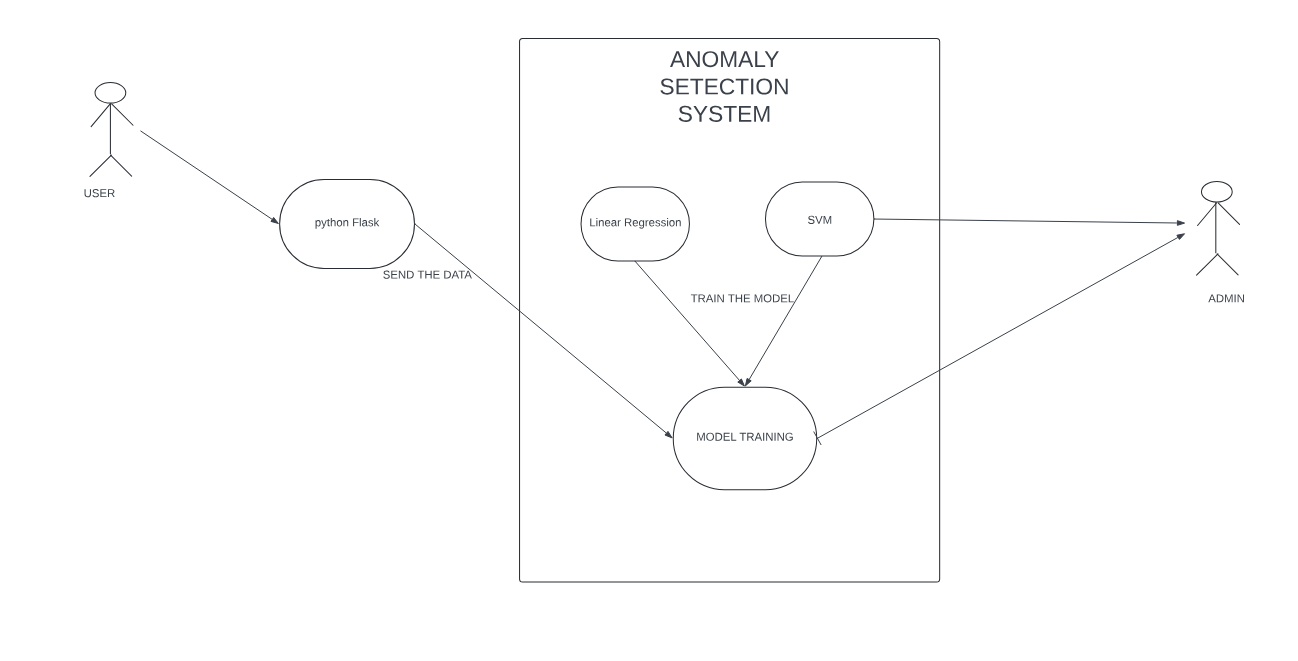
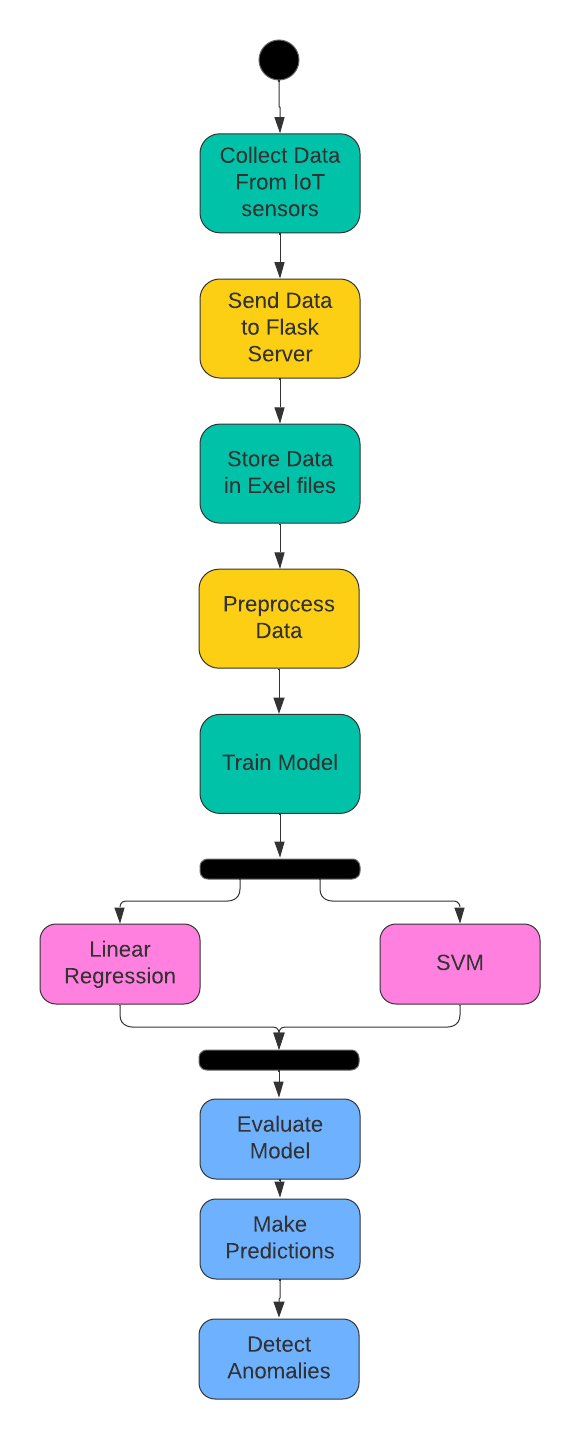
****

Fig 6. UML Use Case Diagram

**3.5.4 UML Activity Diagram**

The Fig 7 illustrates handling IoT sensor data for anomaly detection. Data is collected and sent to a Flask server, stored in Excel files, and preprocessed for model training. Two machine learning models, Support Vector Machine (SVM) and Linear Regression, are trained. Model performance is then assessed, and predictions are made using the evaluated model(s). These predictions are examined for discrepancies, identifying patterns or behaviors. 

## Fig 7. UML Activity diagram

**3.6 Testing**

Data testing, particularly in machine learning and statistical modeling, is the act of determining how well a trained model performs on previously unknown or new data. This stage is vital for determining the model's generalization ability and ensuring it can make the right assumptions on data it failed to observe during training.

Here are some key aspects and considerations related to testing data

### Purpose of Testing Data

1. **Evaluation of Model Performance**: Testing data is used to measure how well a trained model generalizes to new, unseen data. It helps determine if the model has learned meaningful patterns from the training data that can be applied to new instances.
2. **Validation of Model Robustness**: Testing data allows you to validate the robustness of your model against overfitting (when the model performs well on training data but poorly on test data) and underfitting (when the model fails to capture the underlying patterns in the data).

### Process of Testing Data

1. **Data Splitting**: Typically, the available dataset is divided into training and testing sets using techniques like train-test split or cross-validation. The training set is used to train the model, while the testing set remains unseen until evaluation.
2. **Evaluation Metrics**: Various metrics are used to evaluate the model's performance on the testing data, depending on the type of problem (classification, regression, etc.). Common metrics include accuracy, precision, recall, F1-score for classification; mean squared error (MSE), mean absolute error (MAE), R-squared for regression, and others specific to the domain.
3. **Model Evaluation**: After making predictions on the testing data, compare the predicted outcomes with the actual outcomes (ground truth). This comparison allows the calculation of the chosen evaluation metrics to assess how well the model performs.

### Best Practices

1. **Use of Separate Test Set**: Ensure that the testing data is completely separate from the training data to avoid any bias in model evaluation.
2. **Iterative Improvement**: Use insights from testing to iteratively improve the model by adjusting hyperparameters, trying different algorithms, or enhancing feature engineering.
3. **Reporting Results**: Report the results of model testing, including the chosen metrics, to communicate the model's performance objectively.

In summary, testing data is essential for validating the effectiveness and reliability of machine learning models. It provides critical insights into how well the model can generalize to new data and informs decisions about model deployment and further improvements.

## CHAPTER 4

## RESULTS AND DISCUSSIONS

**4.1 Description of dataset**

The dataset used in this project consists of time-series data collected from various IoT sensors. The data captures several environmental and operational parameters that influence power consumption in a given environment. Each row in the dataset represents a one-minute interval, providing a granular view of the conditions and power usage. The dataset comprises the following columns

* **Timestamp**
  + Description: The specific date and time when the data was recorded.
  + Type: Numeric
  + Example: 2024-06-01 00:00
* **Indoor Temperature**
  + Description: The temperature inside the building or facility, measured in degrees Celsius.
  + Type: Numeric
  + Example: 24.6
* **Outdoor Temperature**
  + Description: The temperature outside the building or facility, measured in degrees Celsius.
  + Type: Numeric
  + Example: 27.9
* **Indoor Humidity**
  + Description: The humidity level inside the building, is measured as a percentage.
  + Type: Numeric
  + Example: 40.3
* **Outdoor Humidity**
  + Description: The humidity level outside the building, is measured as a percentage.
  + Type: Numeric
  + Example: 48
* **CO2 Level**
  + Description: The concentration of CO2 inside the building, measured in parts per million (ppm).
  + Type: Numeric
  + Example: 844
* **Setpoint Temperature**
  + Description: The desired or target temperature set for the HVAC system to maintain, measured in degrees Celsius.
  + Type: Numeric
  + Example: 22.1
* **Fan Speed**
  + Description: The speed at which the HVAC fan is operating, is categorized as 'low', 'medium', or 'high'.
  + Type: Categorical
  + Example: high
* **Compressor Status**
  + Description: The operational status of the HVAC compressor, indicating whether it is 'on' or 'off'.
  + Type: Categorical
  + Example: on
* **Power Consumption**
  + Description: The amount of electrical power consumed by the HVAC system over a period, typically measured in watts (W) or kilowatts (kW).
  + Type: Numeric
  + Example: 3500 W

### Data Characteristics

* **Time Series Nature:** The dataset is time series, providing data at one-minute intervals. This temporal aspect is crucial for capturing the dynamic changes in environmental conditions and power consumption.
* **Environmental Variables:** Includes both indoor and outdoor conditions (temperature and humidity), which are key factors influencing HVAC operation and power usage.
* **Operational Variables:** Includes HVAC operational settings such as setpoint temperature, fan speed, and compressor status, which directly impact power consumption.
* **Power Consumption**: The target variable to be predicted, indicates the real-time energy usage based on the given conditions and operational parameters.

**4.2 Detailed Explanation about the Experimental Results**

The study demonstrates that IoT data can be successfully utilized for estimating power usage through advanced analytics, significantly enhancing energy management techniques. Data is gathered from Internet of Things sensors and transmitted to a Flask server, which is stored in Excel files for analysis afterward. Preprocessing takes place on this data in order to make sure it is clean and ready for model training. To find developments in power utilization, two models were employed: Support Vector Machine (SVM) and Linear Regression. The linear regression model showed how variables like temperature, humidity, and time of day influence energy usage while providing insightful information about the linear correlations between surroundings and power consumption. When it comes to understanding straightforward, proportional relationships between inputs and outputs, this model is particularly helpful.

Nevertheless, the SVM model was especially successful at capturing non-linear interactions, which are crucial for accurately modeling patterns of power use that are influenced by several interconnected variables. Using SVM's capability to handle non-linear data, the project generated a prediction model that was more reliable and accurate. The accuracy and reliability of both models were assessed, and the SVM model performed more effectively than the linear regression model in most cases because it could better incorporate intricate, non-linear relationships. Real-time power consumption forecasting was made possible by the SVM model's execution, which also offered useful information for energy management. Making better decisions is made possible by this predictive ability, which could result in notable increases in energy efficiency and cost savings. By providing better and more sustainable operations, these models' performance illustrates the importance of integrating IoT data with powerful analytics and demonstrates that they can totally alter energy management techniques.

The project used the Isolated Forest algorithm for anomaly detection in HVAC systems in addition to power consumption prediction. Through randomized division of data points, this dependable algorithm—which is frequently used in machine learning—is excellent at finding abnormalities while separating regular cases. As an alternative to conventional techniques, Isolated Forest makes no assumptions about the distribution of the data, which makes it especially useful for identifying anomalies in intricate and dynamic environments like HVAC systems. The Isolated Forest algorithm may detect anomalies when certain variables, such as temperature, humidity, CO2 levels, fan speed, and compressor status, greatly depart from typical trends. By proactively detecting anomalies, possible problems can be identified early and maintenance can be done on schedule. As a result, this improves overall energy efficiency and avoids little problems. The successful deployment of these models demonstrates their practical utility in real-world scenarios, paving the way for smarter energy management and proactive maintenance strategies. By leveraging IoT data and advanced machine learning algorithms, the project showcases a comprehensive approach to optimizing energy usage and maintaining system health, thereby contributing to more sustainable and efficient operations.

**Linear Regression**

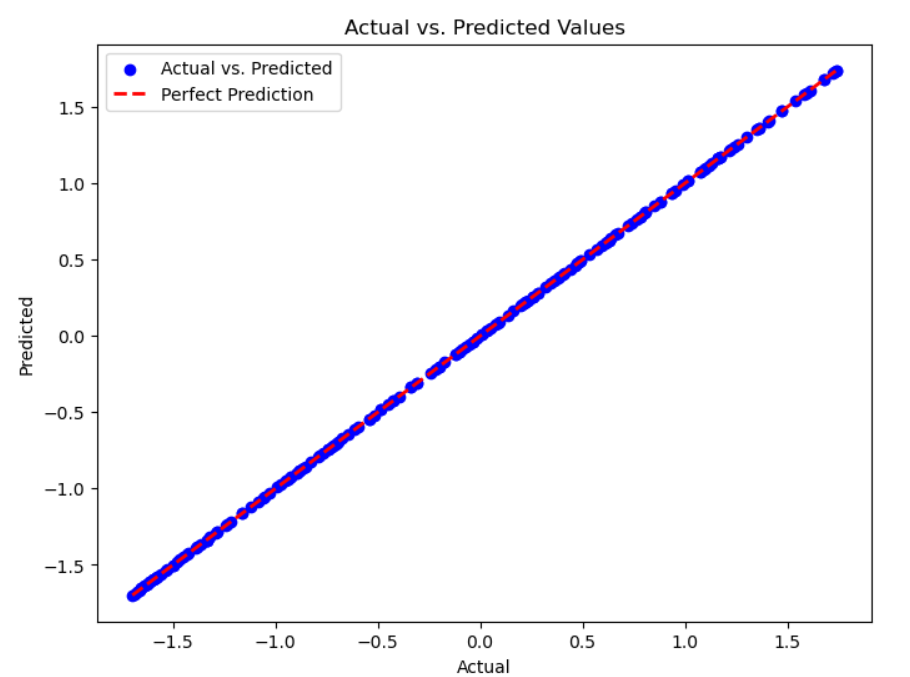
The Fig 8 depicting the graph, demonstrating outstanding accuracy, is a scatter plot and contrasts real values to those predicted through a regression model. The red dashed line denotes perfect prediction, which happens when real values equal predicted values. Each blue dot represents a pair of actual and predicted values. The blue dots' nearly perfect alignment with the red dashed line indicates how accurate the model's predictions seem to be. It demonstrates that the model worked exceptionally well and successfully captured the underlying data patterns with little error. The model's high accuracy in predicting power consumption based on IoT data is further supported by measures like R² (near to 1), which evaluate the model's dependence.****

Fig 8. Scatter plot of Linear Regression

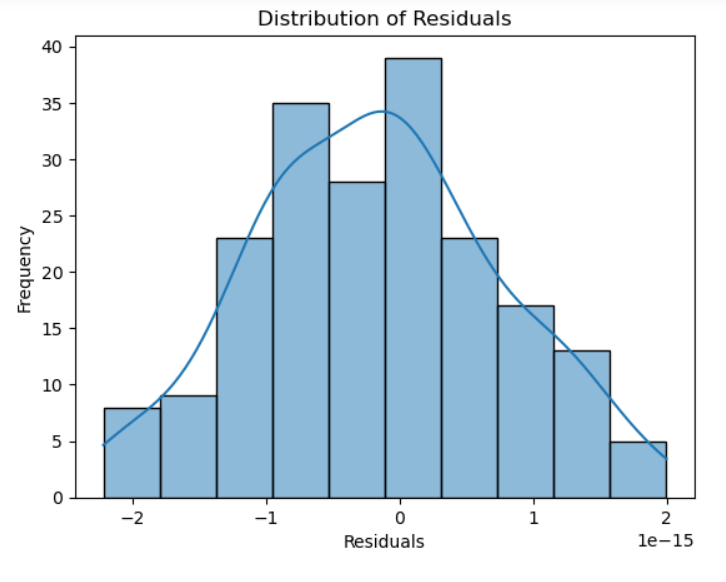
****

Fig 9 . Linear Regression Histogram Plot

**Lasso Regression-**

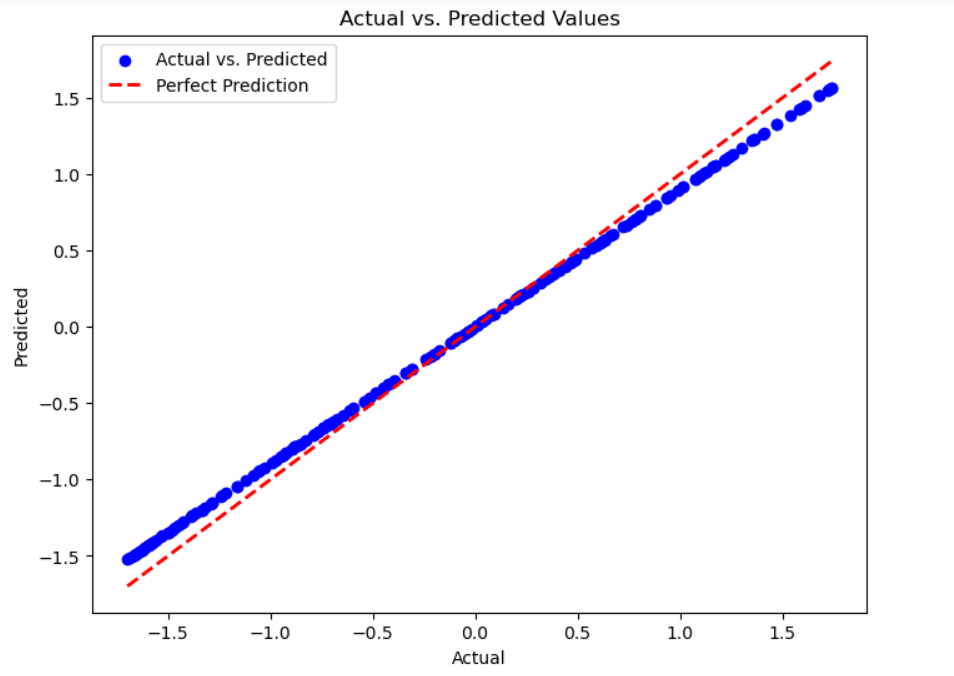
****

Fig 10. Scatter Plot of Lasso Regression

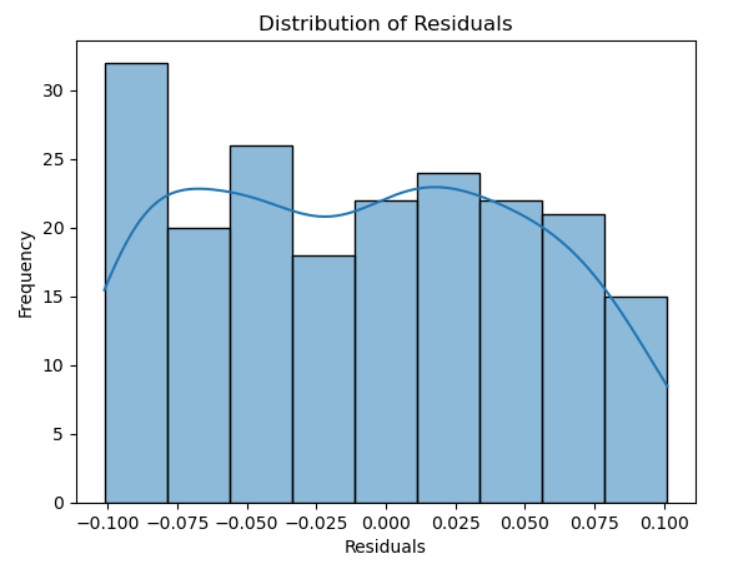
**Support Vector Machines **

Fig 11. Histogram Plot of Support Vector Machines

**IsolationForests**

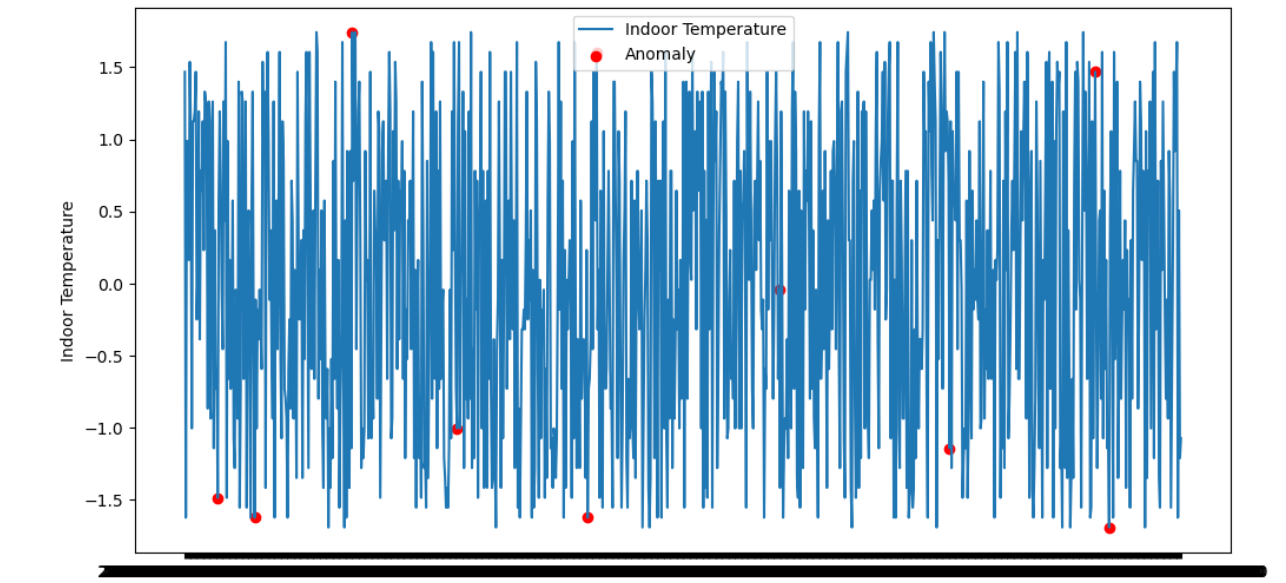
****

Fig 12. Anomaly Detection Scatter Plot

**4.3 Significance of the proposed method with its advantages**

**Significance of the Proposed Method**

The proposed method of using IoT devices to collect environmental and operational data and applying predictive models like Linear Regression and Support Vector Machines (SVMs) to forecast power consumption is significant for several reasons:

Enhanced energy management involves utilizing real-time data to ensure precise monitoring and forecasting of power consumption, enabling more effective energy management strategies. Accurate forecasting allows HVAC systems to operate at maximum efficiency, reducing unnecessary energy usage and lowering utility costs. The integration of IoT devices provides continuous, real-time data, empowering facility managers to make informed decisions quickly. Predictive analytics facilitate proactive adjustments to HVAC settings, enhancing overall efficiency and comfort. This approach also offers cost savings by reducing utility bills through optimized power utilization and lowering maintenance costs via predictive maintenance, which schedules repairs based on usage patterns to minimize downtime.

Enhanced energy management involves utilizing real-time data to provide accurate monitoring and forecasting of power consumption, resulting in more effective energy management tactics. precise forecasting enables HVAC systems to function at peak efficiency, eliminating wasted energy and lowering utility costs. The integration of IoT devices delivers continuous, real-time data, allowing facility managers to make educated decisions quicker. Predictive analytics enables proactive adjustments to HVAC settings, hence improving overall efficiency and comfort. This strategy also saves money by eliminating utility bills through improved power use and lowering maintenance costs through predictive maintenance, which schedules repairs based on usage patterns to minimize downtime.

The proposed method's scalability and adaptability make it suitable for various building sizes and types, making it versatile across industries. The system is adaptable and capable of incorporating additional data sources or new predictive models as technology advances. By gathering and analyzing multiple indicators, such as temperature, humidity, CO2 levels, and both indoor and outdoor variables, this method provides comprehensive insights into factors influencing power consumption, aiding informed decision-making and improving operational planning and resource allocation. Advanced predictive capabilities, including linear regression and support vector machines (SVMs), allow for accurate predictions of power usage. Real-time monitoring and control enable immediate responses to environmental changes and dynamic adjustments to HVAC settings, maintaining ideal conditions and reducing energy usage. A user-friendly interface with visualization tools improves the presentation of trends and predictions, making the system accessible even to users with limited technical knowledge, thus enhancing its overall usefulness.

## 

## 

## CHAPTER 5

## CONCLUSIONS AND FUTURE ENHANCEMENTS

#### Objectives

#### The primary objective of this study is to use data collected through Internet of Things devices to create statistical models that anticipate power usage. It includes establishing Internet of Things (IoT) sensors to continuously monitor a range of operational and environmental information, including humidity, temperature indoors as well as outdoors, CO2 levels, and HVAC settings. Effective data collecting and preprocessing strategies are crucial for preserving the quality and completeness of the collected data. The goal is to develop models that, using this data, effectively forecast power use and present practical advice for energy optimization and management.

The objective of the project goes beyond gathering data to include using this data to gain useful knowledge. Through the examination of intricate correlations between diverse factors and their effect on energy usage, the research seeks to offer a resilient instrument for enhancing energy efficiency. In energy resource management, these predictive models help with strategic long-term planning and enable prompt decision-making. This approach not only improves immediate operational efficiency but also gradually reduces costs and supports the goals of sustainability. Using the technique of cutting-edge analytics, the project aims to convert unprocessed data into relevant information that enables organizations to effectively manage their energy use, therefore contributing to a more sustainable and profitable future.

#### In the areas of sustainability and energy management, this project is important. Forecasting power use with precision allows for more efficient use of energy resources, reduced operating expenses, and less environmental effect. A contemporary method for managing energy use can be offered by combining IoT technology with sophisticated predictive analytics. This approach meets the growing demand for more intelligent, data-driven solutions in the face of rising energy prices and a growing focus on environmental sustainability.

#### The study also shows how technological advancements may be used to address real-world issues. Organizations can move from reaction to proactive energy management strategies through the use of real-time data. This change aligns with global environmental goals and enhances operating efficiency, emphasizing the importance of the initiative and the possibility of its impact in the energy-conscious world of today.

#### Significance

#### The suggested method enhances energy management by constantly observing and forecasting power usage. IoT devices' real-time data collecting enables prompt insights and wise adjustments, boosting efficiency. Additionally, by encouraging energy efficiency and lowering buildings' carbon footprints, this strategy helps achieve sustainability goals. due to its versatility and scalability, it could be used in a variety of industries and offers a flexible answer to energy management problems.This initiative has importance even beyond its operational benefits. The method promotes a better knowledge of the energy dynamics in buildings by enabling in-depth investigation as well as forecasting. The data gathered can influence the development of energy technology, influence policy choices, and ultimately back bigger efforts to mitigate climate change. The project's ability will result in immediate benefits in terms both the environment and economics

#### Approach Adopted

#### The method has multiple crucial phases, commencing with the implementation of Internet of Things sensors to collect real-time data on operational and environmental aspects. After that, the gathered data is cleaned and preprocessed to deal with noise, outliers, and missing values. A support vector machine (SVM) model is created to capture non-linear relationships, while a linear regression model is created for understanding linear conversations. Metrics including Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared are used to assess these models. To enhance prediction performance, model optimization, and hyperparameter modification have been carried out.

#### The best-performing model is employed for real-time predictions following the development and evaluation stages. The model shall be integrated into an intuitive dashboard or user interface as part of this deployment, giving stakeholders quick access to predictions and insights.

#### Results

The project's outcomes show how effectively IoT data might be utilized for predicting electricity usage. While the SVM model caught more intricate, non-linear interactions, the linear regression approach provided insight on the linear relationship between environmental variables and power use. The accuracy and dependability of both models were evaluated, and because the SVM model can handle non-linear data, it usually performed better. Real-time power consumption forecasting was made feasible by the predictive model's deployment, which offered helpful data for energy management.

These results highlight the need to combine advanced analytics alongside IoT. The models created in this study have the potential to significantly enhance energy efficiency and save expenses by exactly expecting power use. These models' effective implementation and applicability show that they can completely transform the energy industry.

The anomaly detection algorithm Isolated Forest, which is commonly used in machine learning, is excellent at predicting abnormalities in HVAC systems because it uses randomized partitioning of data points to determine normal cases. In contrast with traditional methods, Isolated Forest doesn't necessitate the premise of data distribution, which makes it especially useful for identifying anomalies in intricate and dynamic settings like HVAC systems. Through analysis of multiple factors such as temperature, humidity, CO2 concentrations, fan speed, and compressor status, the program identifies anomalies in which these metrics exhibit a notable departure from the predicted trends. By identifying issues before they become severe, this method—which appears in the accompanying diagram in the documentation—enables proactive maintenance techniques and optimizes overall energy efficiency.

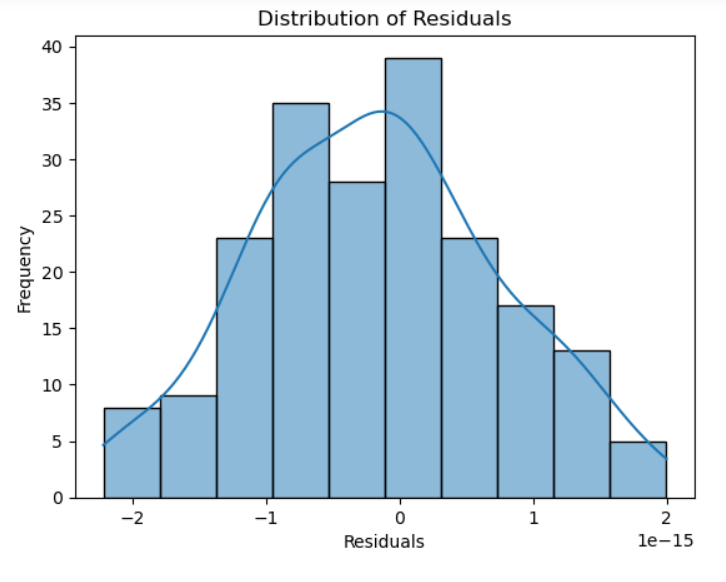
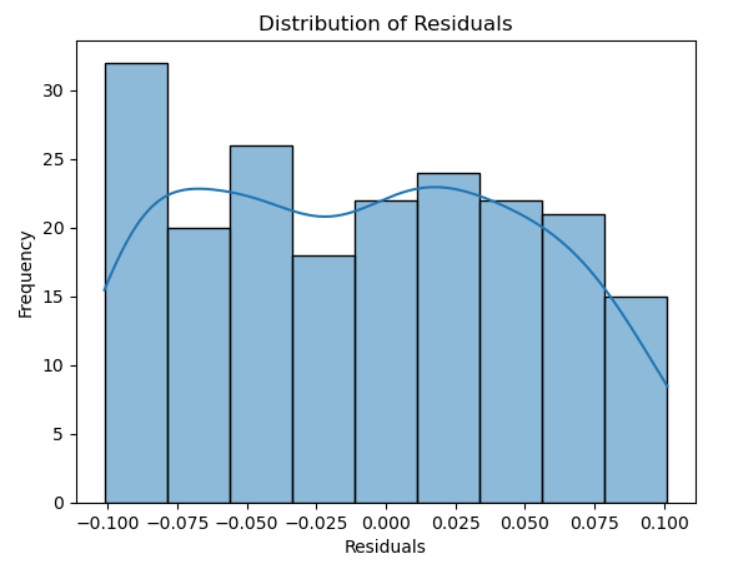


Fig 13. Histogram Plot for the Regression Model Implemented



#### Fig 14. Histogram Plot for SVM Implemented

#### 

#### Fig 15.A Depiction Of The Anomalies Detected In IsloationForest In The Implementation

#### Future Enhancements

Future advances in HVAC system anomaly detection have a great deal of potential for improving operational effectiveness and prediction accuracy. Including data from other sources, such as occupancy and real-time predictions of the weather, is a crucial strategy. By including these variables, the system improves the accuracy of anomaly detection algorithms by gaining a greater awareness of external influences on HVAC performance. This approach enables preemptive adjustments in reaction to evolving environmental circumstances, maximizing energy usage and occupant comfort in buildings.Another way to improve is via the use of deep learning and other advanced machine learning techniques. These techniques can identify complex correlations and patterns within HVAC data that typical algorithms could miss. Deep learning models make it possible to determine abnormalities with greater precision, which might result in quicker identification of potential systemic issues.

Furthermore, it is essential to improve the predictive maintenance functionalities of HVAC systems. Predictive models employ real-time metrics and before-data analysis to anticipate equipment problems before they happen. Because of this foresight, maintenance staff may plan repairs at times of low demand, which maximizes system uptime and minimizes energy use. These predictive models can be used to control other building systems, such as lighting and security, in alongside HVAC-specific applications. Building managers can accomplish shared objectives for sustainability and operational efficiency by integrating these systems under one predictive analytics framework. These models are always being taught from and adjusted to account for changing environmental dynamics and building challenges in leadership. This is rendered possible by continuing data insights.

## CHAPTER 6

## APPENDICES

**Flask Server Code-**

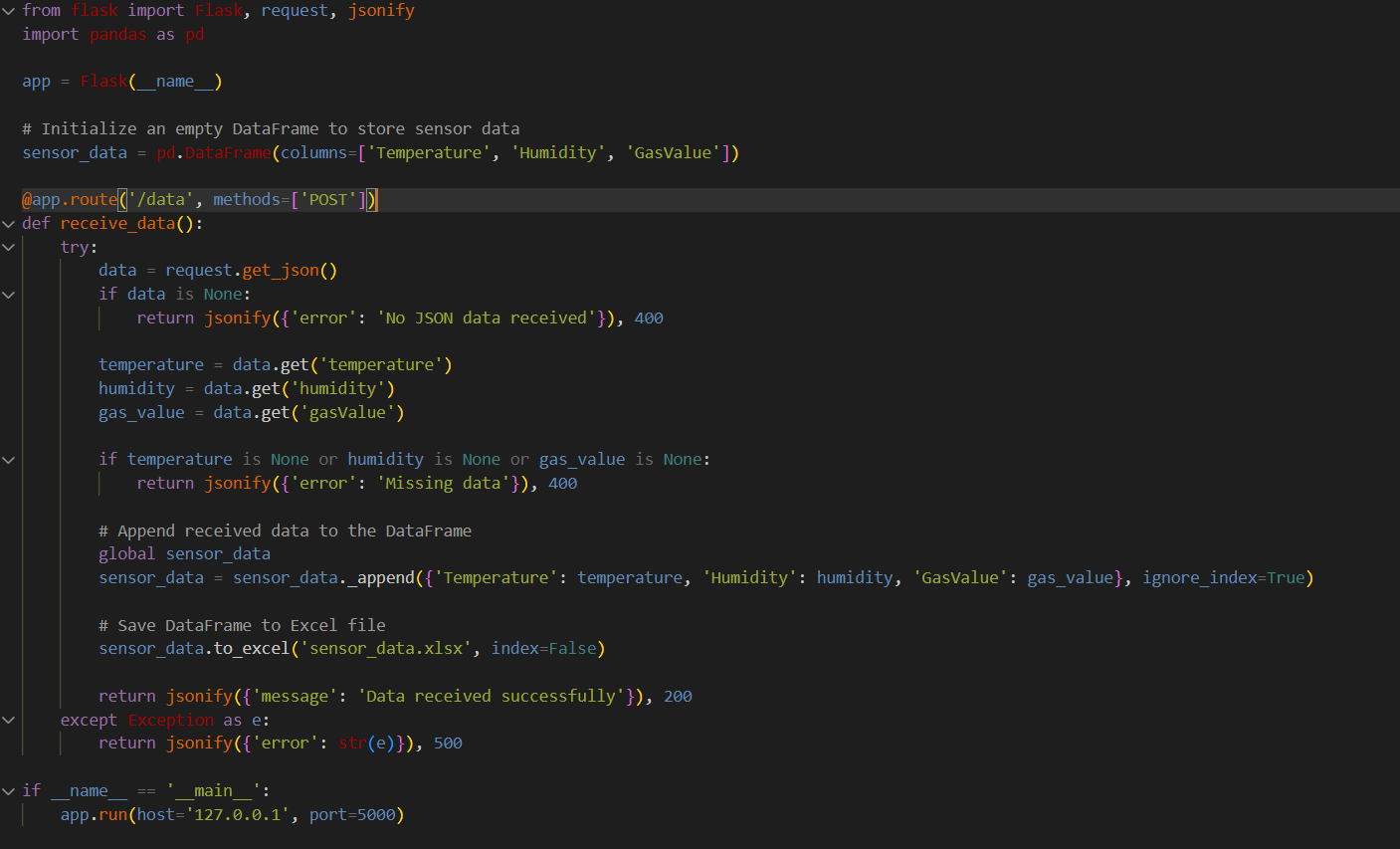


Fig 16.Flask Server Code Implementation

The Fig 16 displays the code is a Flask application that serves as an endpoint to receive sensor data via HTTP POST requests. It initializes an empty pandas DataFrame called sensor\_data to store the received data. The /data route handles POST requests containing JSON data representing sensor readings for temperature, humidity, and gas values.The application evaluates when the JSON data is present when it receives a request. It produces an error answer if it doesn't. From the JSON data, the temperature, humidity, and gas values is subsequently determined. It returns an additional error response to if any of these values are missing.The application appends the necessary data to the sensor\_data DataFrame if it is accessible. The data frame is subsequently saved to an Excel file named sensor\_data.xlsx.If there are any exceptions during the process, the application provides information about the exception along with an error answer.Lastly, the program is set up to execute at port 5000 on localhost.

**Flask Terminal-**

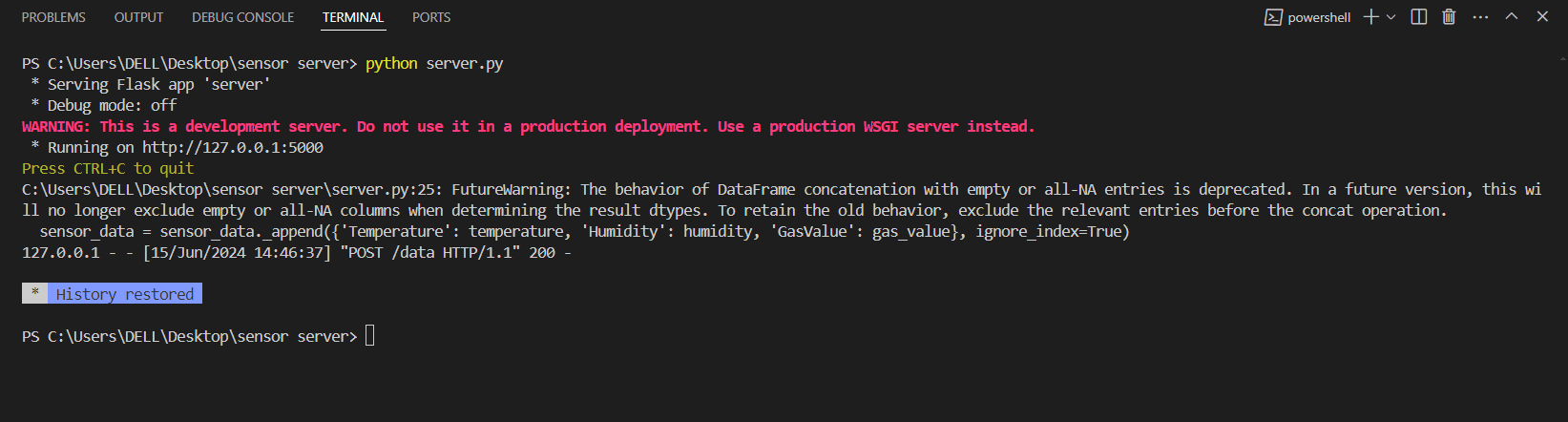
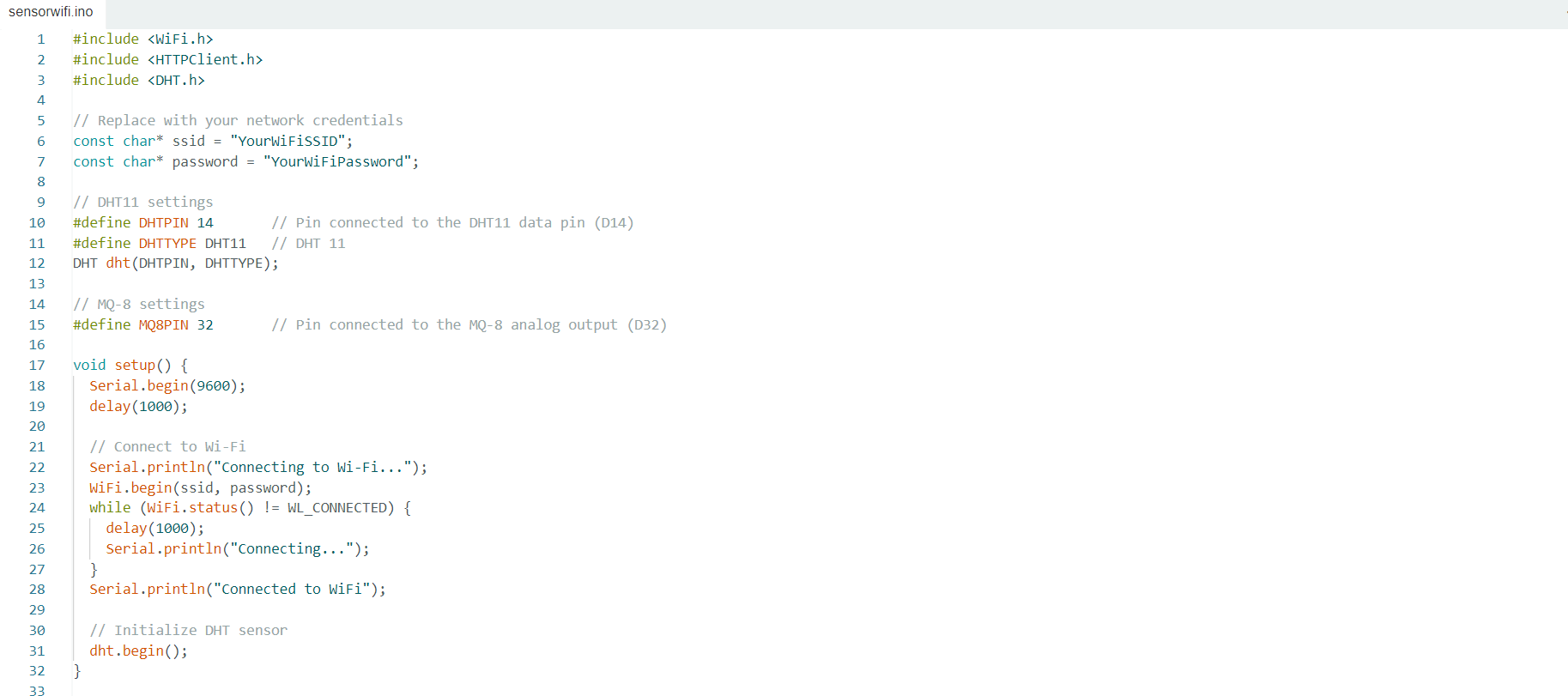
****

Fig 17. Flask Server Initialization

This Fig 17 which represents the log indicates that the Flask application named 'server' is running in debug mode, serving requests on<http://127.0.0.1:5000>. It starts with a warning advising against using the development server in a production environment and recommends using a production WSGI server instead.The /data endpoint returns a status code 200, signifying that a POST request containing sensor data was successful when it receives it.Additionally, when adding data to the DataFrame, pandas raise a FutureWarning. This warning indicates that the DataFrame concatenation behavior with empty or all-NA elements is no longer supported. In order to maintain the previous behavior, it is recommended to omit pertinent entries prior to the concatenation operation.

**Aurduino Implementation code-**

**** Fig 18.Aurduino Implementation Code

The Fig 18 represents Arduino program receives data from an MQ-8 gas sensor and a DHT11 temperature and humidity sensor in addition to connecting an ESP32 device to a Wi-Fi network. It then uses HTTP POST requests to deliver this data to a Flask server. The sketch initializes the sensors in the setup() function, defines constants for Wi-Fi credentials, and pin settings for the sensors. The loop() function reads sensor data, verifies that there are no reading errors, prints the sensor values to the serial monitor, and then uses the sendSensorData() function to send the data to the Flask server. This function constructs the JSON payload and the server endpoint. Lastly, the success or failure of the HTTP response is verified.

**Machine Learning Code-**

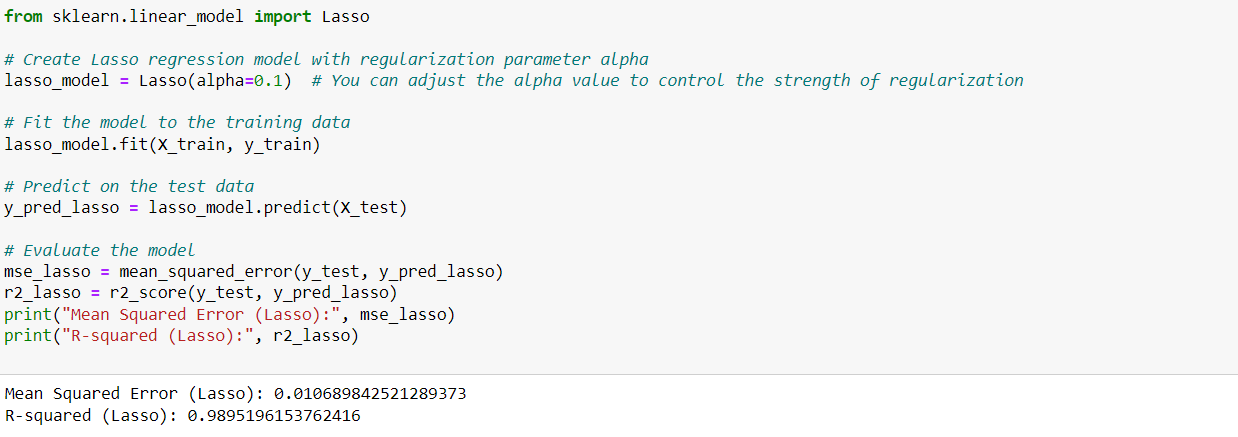
****

Fig 19. Lasso Regression Implementation Code

The Fig 19 Python code segment employs Lasso regression, a technique for linear regression with regularization, to build a predictive model. By importing the necessary libraries, initializing the Lasso model with a specified regularization parameter, fitting the model to training data, and subsequently making predictions on test data, it establishes a robust framework for predictive analysis. The evaluation stage involves computing mean squared error (MSE) and R-squared scores, offering insights into the model's accuracy and explanatory power. This approach enables practitioners to effectively assess and utilize Lasso regression for various regression tasks, balancing predictive performance with regularization strength to achieve optimal model outcomes.

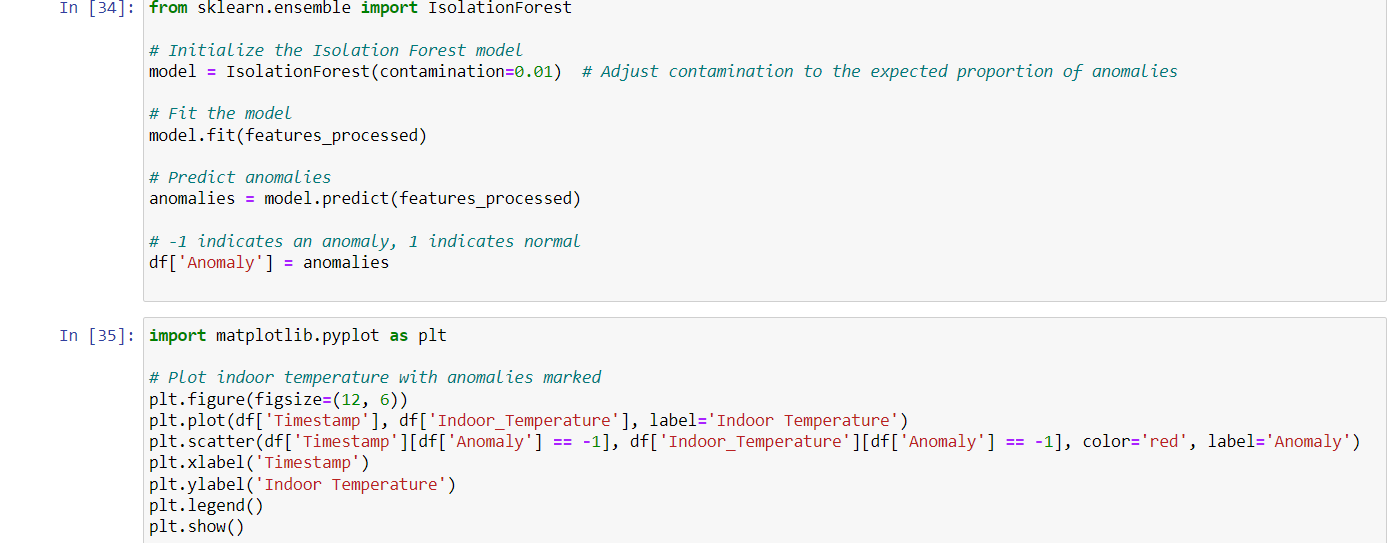
****

Fig 20. IsolationForest Implementation Code

Fig 20 shows Python code snippet employs the Isolation Forest algorithm, a popular method for anomaly detection, to identify outliers within a dataset. Initially, the Isolation Forest model is instantiated with a specified contamination parameter, denoting the expected proportion of anomalies in the data. Following this, the model is trained on processed features, enabling it to learn the underlying patterns and structure of the dataset. Subsequently, anomalies are predicted using the trained model, with each data point classified as either normal or anomalous. To visually represent these anomalies, the script utilizes Matplotlib to generate a plot of indoor temperature over time, with detected anomalies highlighted in red. This visualization aids in the quick identification and interpretation of unusual data points against the backdrop of the overall temperature trend. Overall, this code segment provides a concise yet powerful solution for detecting anomalies in time-series data, facilitating data-driven insights and decision-making in various domains.

## REFERENCES

| [1] | Borda, Davide, et al. "Development of anomaly detectors for HVAC systems using machine learning." *Processes* 11.2 (2023): 535 |
| --- | --- |
| [2] | Dadras Javan, Farzad, et al. "Machine-Learning-Based Prediction of HVAC-Driven Load Flexibility in Warehouses." *Energies* 16.14 (2023): 5407. |
| [3] | Ghahramani, Ali, Simin Ahmadi Karvigh, and Burcin Becerik-Gerber. "HVAC system energy optimization using an adaptive hybrid metaheuristic." *Energy and Buildings* 152 (2017): 149-161. |
| [4] | Najafi, Massieh, et al. "Application of machine learning in the fault diagnostics of air handling units." *Applied Energy* 96 (2012): 347-358. |
| [5] | Guo, Ying, et al. "Intelligent model based fault detection and diagnosis for HVAC system using statistical machine learning methods." *ASHRAE Trans* 119 (2013): O1. |
| [6] | Gharsellaoui, Sondes, et al. "Interval-valued features based machine learning technique for fault detection and diagnosis of uncertain HVAC systems." *IEEE Access* 8 (2020): 171892-171902. |
| [7] | Jahedi, G., and M. M. Ardehali. "Wavelet-based artificial neural network applied for energy efficiency enhancement of decoupled HVAC system." *Energy Conversion and Management* 54.1 (2012): 47-56. |
| [8] | Amasyali, Kadir, Mohammed Olama, and Aniruddha Perumalla. "A machine learning-based approach to predict the aggregate flexibility of hvac systems." *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, 2020. |
| [9] | Jang, Hyeonwoo, et al. "Design and implementation of IoT-based HVAC and lighting system for energy saving." *MATEC Web of Conferences*. Vol. 260. EDP Sciences, 2019. |
| [10] | Ebirim, Wisdom, et al. "Project management strategies for accelerating energy efficiency in hvac systems amidst climate change." *International Journal of Management & Entrepreneurship Research* 6.3 (2024): 512-525. |