Smart Elderly Home Monitoring: An Integrated Analysis of Gas, Temperature, and Position Sensors for Anomaly Detection and Safety Enhancement

Sushil Babulal Jangid - 10615112

Supervisor: Hamidreza Khaleghzadeh

Dissertation submitted in part fulfillment of the

requirements for the degree of

MSc in Data Analytics

At Dublin Business School

January 2024

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Signed: Sushil Jangid

Student Number: 10615112

Date: 8th January 2024

Acknowledgements

I would like to express my gratitude to Hamidreza Khaleghzadeh, my research supervisor for their valuable contributions to my research project. His role was essential in the successful completion of this study, and I am deeply grateful for his expertise and guidance.

I would like to express my gratitude to the lecturers and friends at Dublin Business School for their support over the course of my time here.

Finally, I would like to share my gratitude for everything I have been blessed with – Family, Friends and continued support from everyone

Table of content

List of Figure

Table of Tables

Abstract:
In the domain of Smart Elderly Home Monitoring, our research aims to elevate anomaly detection by integrating gas, temperature, and position sensors. Leveraging four distinct models—Autoencoder, Local Outliers Factor (LOF), Isolation Forest, and One-Class Support Vector Machine (SVM)—our study provides a comprehensive analysis. The Autoencoder excels, securing a flawless accuracy score of 1.00, showcasing its proficiency in detecting even the most subtle deviations. Following closely, the LOF model demonstrates commendable performance with an accuracy of 0.90, effectively identifying local outliers. The Isolation Forest mirrors the Autoencoder's prowess with a perfect accuracy score. However, the One-Class SVM reveals limitations with a comparatively lower accuracy of 0.03, indicating challenges in its anomaly detection capability within the defined parameters. This research contributes significantly to the field of smart home monitoring, furnishing valuable insights for optimal anomaly detection model selection to fortify the safety and well-being of elderly individuals in a home environment.

Chapter 1: Introduction

In the landscape of contemporary healthcare, the integration of advanced sensor technologies holds transformative potential, particularly in the context of Smart Elderly Home Monitoring. This research endeavors to amalgamate gas, temperature, and position sensors to create a comprehensive system aimed at enhancing anomaly detection and fortifying the safety of elderly residents. With the overarching goal of ensuring a secure and conducive living environment for seniors, our study employs three distinct anomaly prediction models: Autoencoder, Isolation Forest, Local Outliers Factor (LOF), and One-Class Support Vector Machine (SVM). The Autoencoder and Isolation Forest, showcasing unparalleled accuracy with a perfect score of 1.00, stands out as a promising solution for detecting subtle irregularities. As we navigate the intricacies of these models, this research contributes to the evolving field of smart home monitoring, offering a nuanced exploration of anomaly detection methodologies that can significantly impact the well-being of the elderly in their homes.

1.1 Back Ground scope

The confluence of demographic shifts and technological advancements has significantly altered the landscape of healthcare, with a pronounced impact on the aging population. According to the United States Census Bureau, the proportion of individuals aged 65 and above is projected to rise from 13% in 2013 to 20% by 2030 (U.S. Census Bureau, 2013). As longevity increases, so does the prevalence of age-related health challenges, including Alzheimer's disease and other cognitive impairments. Ali et al. (2019) estimate that these conditions affect approximately 5.2 million people above the age of 65, necessitating novel approaches to address the evolving healthcare needs of the elderly (Ali et al., 2019).

Wearable sensing technology has emerged as a transformative force in various domains, including healthcare and safety. In the construction industry, Ahn et al. (2019) demonstrate the applications of wearable sensors for safety and health monitoring, illustrating the potential of continuous physiological data collection (Ahn et al., 2019). This technology has paved the way for real-time monitoring, providing valuable insights into the well-being of individuals. As individuals age, the continuous monitoring of physiological parameters becomes particularly crucial for early detection of health issues and proactive intervention.

The proliferation of the Internet of Things (IoT) further amplifies the potential for real-time monitoring and data-driven insights. Machine learning and deep learning techniques, as reviewed by Al-amri et al. (2021), play a pivotal role in processing and analyzing the voluminous and complex data generated by IoT devices, facilitating effective anomaly detection (Al-amri et al., 2021). As the elderly increasingly opt for aging in place, the concept of smart homes equipped with intelligent systems gains prominence. Such homes are envisioned as environments that can adapt to the unique needs of elderly residents, ensuring their safety and well-being (DeFranco & Kassab, 2021).

Studies such as "Intelligent Autonomous Elderly Patient Home Monitoring System" (Ali et al., 2019) exemplify the practical application of smart home technologies in addressing the healthcare needs of the elderly. These systems integrate various sensors to monitor the health status of residents, providing a comprehensive approach to healthcare management (Ali et al., 2019). Furthermore, real-time health monitoring systems, as demonstrated by Bora et al. (2021), utilize platforms like Arduino and Raspberry Pi, showcasing the feasibility of implementing such technologies in real-world scenarios (Bora et al., 2021). The integration of gas, temperature, and position sensors within these systems presents a multifaceted approach to monitoring, allowing for the detection of anomalies that may indicate health issues or safety concerns.

However, the field of smart home research is not without challenges. Cook et al. (2019) highlight the need for effective anomaly detection methods for IoT time-series data, given its inherent complexity (Cook et al., 2019). An in-depth analysis and taxonomy of smart home research themes by DeFranco and Kassab (2021) reveal the intricate web of factors that need consideration for the successful implementation of smart home technologies (DeFranco & Kassab, 2021). As the body of research in this field expands, there is a growing recognition of the need for a comprehensive understanding of the underlying themes to unlock the full potential of smart homes.

This research endeavours to contribute to the evolving landscape of smart home technologies for elderly care by conducting an integrated analysis of gas, temperature, and position sensors. By leveraging machine learning techniques, the study seeks to enhance anomaly detection capabilities, providing a robust and reliable framework for real-time monitoring. The scope extends beyond conventional healthcare approaches, offering a holistic understanding of the potential of smart homes in ensure the safety and well-being of elderly residents. As the demographic shift continues, this research aims to provide actionable insights that can inform the development of intelligent systems tailored to the unique needs of an aging population, ultimately contributing to advancements in eldercare technology.

1.2 Research Question:

The research question are as follows:

"How does the integration of gas, temperature, and position sensors, utilizing Autoencoder, Isolation Forest, Local Outliers Factor (LOF), and One-Class Support Vector Machine (SVM) models, contribute to effective anomaly detection and safety enhancement in the context of Smart Elderly Home Monitoring?"

1.3 Research Motivation

The motivation for this research stems from the pressing need to address the evolving challenges faced by the aging population, particularly in the context of Smart Elderly Home Monitoring. As demographic shifts indicate a substantial increase in the elderly population, there is a parallel rise in age-related health issues, including cognitive impairments. These challenges necessitate innovative solutions to ensure the well-being and safety of elderly individuals, especially those choosing to age in place. The integration of gas, temperature, and position sensors, coupled with advanced anomaly detection models such as Autoencoder, Isolation Forest, Local Outliers Factor (LOF), and One-Class Support Vector Machine (SVM), holds the potential to revolutionize the monitoring and care of elderly residents.

1.4 Research Objectives

The research objectives are multifaceted. Firstly, the study aims to evaluate the individual performance of Autoencoder, Isolation Forest, LOF, and One-Class SVM models in detecting anomalies within the specific context of smart elderly home environments. This involves a quantitative assessment using precision, recall, and F1 score metrics. Secondly, the research investigates the collective impact of integrating gas, temperature, and position sensors on anomaly detection, providing insights into the overall system performance. The study further extends its focus to examine how timely anomaly detection contributes to safety enhancement measures within Smart Elderly Home Monitoring systems, exploring preventive measures and intervention strategies. Additionally, the practical feasibility of implementing the proposed anomaly detection models in real-world smart elderly home environments is assessed, considering adaptability and scalability to varying conditions. The research concludes by offering practical recommendations for the implementation of anomaly detection systems, contributing valuable insights to the field of Smart Elderly Home Monitoring and advancing the well-being of the aging population.

1.5 Research Outlines

In Chapter 1, the introduction lays the foundation for the research on Smart Elderly Home Monitoring, emphasizing the increasing challenges faced by the aging population and the potential of integrated sensor technologies for anomaly detection and safety enhancement. This section articulates the research question, delineates the research objectives, and highlights the significance of the study in addressing the evolving needs of elderly individuals.

Moving on to Chapter 2, the literature review critically examines existing scholarly works related to Smart Elderly Home Monitoring, integrated sensor technologies, and anomaly detection models. This chapter provides a comprehensive overview of relevant studies, identifying gaps, and establishing a theoretical framework to guide the research. By synthesizing existing knowledge, the literature review contextualizes the current research within the broader academic landscape.

Chapter 3, the methodology section, outlines the research design, data collection methods, and the application of Autoencoder, Isolation Forest, Local Outliers Factor (LOF), and One-Class Support Vector Machine (SVM) models for anomaly detection. The chapter details the rationale behind selecting these methodologies and models, ensuring transparency and reproducibility in the research process.

Chapter 4, dedicated to results, presents the empirical findings derived from the integration of gas, temperature, and position sensors. This section provides a comprehensive analysis of the performance of Autoencoder, Isolation Forest, LOF, and One-Class SVM in detecting anomalies, offering insights into the effectiveness of the integrated approach for Smart Elderly Home Monitoring.

Chapter 5 serves as the conclusion and future scope section, summarizing key findings, discussing their implications, and providing recommendations for further research. This chapter elucidates the broader significance of the study's outcomes and proposes avenues for future investigations, fostering continuous advancements in the field of Smart Elderly Home Monitoring.

Collectively, these chapters form a coherent and structured research framework, ensuring a systematic exploration of the integration of gas, temperature, and position sensors for anomaly detection and safety enhancement in Smart Elderly Home Monitoring.

Chapter 2: Literature Review

Introduction

The tendency of elderly people to remain in their homes is increasing daily as they face excessive physical disabilities. In recent years, the health problem of the ageing population has appeared seriously. This scenario enhances the priority of smart homes that monitor the home environment and also increases the safety of elderly people (Tala *et al*., 2019). Different sensors can be used for anomaly detection to ensure a safe and comfortable home environment. "Smart Elderly Home Monitoring" is a cutting-edge technique for securing the protection and well-being of aged personnel in their habitat (Grossi *et al*., 2020). Integrating gas, temperature and position sensors establishes an innovative system that delivers comprehensive data to secure anomaly detection and safety enhancement. This study aims to detect probable hazards such as gas leaks or excessive increments in room temperature with a motive to track the activity and position of elderly people. The research also emphasizes the sensory activity in the immediate response in emergency cases, which promotes a secure and independent habitation condition for seniors. Smart elderly home monitoring systems can include environmental sensors, wearable medical sensors, actuators, and advanced information and communication technologies (Philip *et al*., 2021). This system helps to execute remote monitoring of senior health and well-being at a very affordable and low cost (Marín *et al*., 2023).

Concepts of the key terms

The concept of elderly home monitoring is designated and designed to improve the safety and well-being of elderly adults. Overall, living dependently with limitations and requirements for assistance, these technologies monitor their protection, activities and health. Technologies such as cameras in the vicinity of elderly individuals have been effective due to continuous monitoring of their condition. Smart elderly monitoring systems uphold several specifications, including advanced tracking mechanisms (Imran *et al*., 2021). It also can measure temperature, humidity, air pressure, and the room's light levels, ensuring a comfortable environment for the elder people (Omidvarborna *et al*., 2021). Sensory features that use smoke detection technology and the percentage of carbon monoxide in the air are also essential for maintaining a safe place. It can also employ several devices to monitor door and window lock systems that have the power to evaluate trespassers (Fadhil *et al*., 2020). The activity tracking system can analyze older adults' sleep timing, step count, and nutrition scale. GPS is an integral part of this system, mainly used for position tracking (Bora *et al*., 2021). Fall detection with integral emergency calling and massaging provides stressless elderly monitoring. Wellness monitoring, such as heart rate, SPo2, and blood pressure monitoring systems, is another concept that has made home monitoring for the elderly premium support for seniors (Jacob Rodrigues *et al.*, 2020).

A device that is mainly utilized to measure the temperature of the environment, which includes the temperature of liquid, solid and air, is called a temperature sensor (Kuzubasoglu *et al.*, 2020). Gas sensors cover broad technological aspects, from gas leak monitoring to the concentration of the emitted gas in the environment (Palacín *et al.*, 2019). This device also can identify the gas. Where the gas leak is happening, the gas sensors can identify the portion and ring the alarm accordingly. These sensors hold a fundamental safety measure purpose. Position sensors are used to track the exact position of personnel (Shen *et a*l., 2020). These sensors can provide essential and emergency information to caregivers and family members. Those home monitoring sensors activate and give an alarm whenever fire hazards, gas leakage, or smoke happen. Carbon dioxide intensity in the home can be detected, which can help maintain the proper well-being of seniors who live independently (Seekaew and Wongchoosuk, 2019).

These sensors deliver vast amounts of data that increase the need for anomaly detection as it is nearly impossible to track those data manually. Anomaly detection analyzes specific data points (Cook *et al.*, 2019). It also examines unexpected events or observations that deviate from the general datasets. In the context of intelligent elderly home monitoring, anomaly detection includes hazardous events or potentially dangerous data that reflect the proposed monitoring system. This system can propose the baseline behaviour by integrating all types of sensory data sets acquired from gas, temperature and position sensors. The sudden increase in temperature, fall analysis, unstable movement or incidental gas leakage occurs in the system triggers alarms or autonomous responses that ensure the safety of the elderly people in their residences. The proactive techniques reduce risk, enable convenient and timely help approaches and provide essential insight to the caregivers and family members that help provide efficient and effective healthcare processes to the seniors.

Safety enhancement refers to the utilization of advanced technology that ensures the safety and well-being of older persons. Safety enhancement features include fall detection and rapid responses to hazardous situations (Ahn *et al.*, 2019). Safety enhancement involves detecting environmental risks, mitigating risk, and ensuring safe living space. Elderly personnel monitoring can track activity, and if abnormalities are noted, it provides immediate assistance. These safety measures enhance security and promote a higher quality of living, including addressing potential risks.

Literature review

A thorough multi-perspective evaluation is provided in the study of Erhan *et al.* (2021) on intelligent anomaly detection in sensor systems. The work addresses important facets of sensor-based anomaly identification across several dimensions. The authors emphasize the integration of information and look at several techniques to improve the effectiveness of anomaly identification. This review's insights help shape how sensor system security is changing. The authors' work highlights the significance of a comprehensive understanding, including knowledge from several angles, to strengthen the resilience of anomaly detection techniques in smart sensor systems. This knowledge synthesis provides a strong basis for developing anomaly detection skills in modern sensor technologies.

According to Mansour *et al.* (2021), computer vision technologies are essential to intelligent video surveillance applications in modern public security, particularly when examining prolonged video streams. In this field, anomaly detection and categorization are crucial. This work introduces the IVADC-FDRL model, a Faster RCNN with a Deep Reinforcement Learning-based Intelligent Video Anomaly Detection and Classification system. The model is divided into two stages: Faster RCNN and Residual Network are used for anomaly detection, and deep Q-learning-based DRL is used for anomaly categorization. IVADC-FDRL achieved a maximum accuracy of 98.50% and 94.80% on the Test004 and Test007 datasets, respectively, in experimental validation on the UCSD anomaly dataset.

In their research paper, Mokhtari et al. (2021) mentioned that Industrial Control Systems (ICSs) struggle to identify assaults using traditional network traffic monitoring because attackers might fool them by imitating regular activity. With the help of data from supervisory control and data acquisition (SCADA) systems, this study proposes a unique solution called the Measurement Intrusion Detection System (MIDS). Even when an attacker tries to hide aberrant activity in the control layer, MIDS still picks it up. A model that distinguishes between normal and abnormal ICS activity uses supervised machine learning and is tested using a Hardware-in-the-Loop (HIL) testbed that simulates power-generating units and uses an attack dataset. To show the efficiency of MIDS, particularly against stealthy assaults, various machine learning models are used, with random forest outperforming other classifiers.

In the views of Ni *et al.* (2020), the widespread use of low-cost online structural health monitoring (SHM) has been made possible by intelligent sensing and sensor network systems developments, leading to the capture of enormous and diverse data for large-scale civil structures. This amount presents problems for traditional data analytics, which affects storage capacity. As a result, data reduction and reconstruction have become essential SHM components for large-scale infrastructure. In two parts, deep learning is used in this study to provide a unique framework for data compression and reconstruction. With high accuracy, a one-dimensional convolutional neural network identifies abnormalities. Next, an autoencoder-based compression and reconstruction approach is used to recover high precision even at low compression ratios. The validity of the suggested method is shown using acceleration data from a Chinese long-span bridge.

According to Nassif *et al.* (2021), anomaly detection has been used for years to find and extract strange data components, with machine learning (ML) emerging as a vital tool in this field. This research study uses a Systematic Literature Review (SLR) to examine ML models for anomaly detection. The study examines applications, ML approaches, performance measures, and classification in anomaly detection through analysis of 290 research publications published between 2000 and 2020. It presents 29 unique ML models, 22 experimental datasets, and 43 applications. The success of unsupervised anomaly detection highlights the promise of anomaly detection using ML models. The report ends with suggestions and directives for researchers in this developing area.

According to Al-amri *et al.* (2021), anomaly detection has attracted a lot of Interest with the development of technologies like the Internet of Things (IoT), which generate enormous continuous data streams from various applications. Analyzing this data to spot suspicious activity, counter functional risks, and avoid unanticipated problems resulting in application outages is essential. Although certain anomalous behaviours in IoT data streams are covered in previous research, a thorough assessment of all factors still needs to be included. This study seeks to fill this vacuum by providing a comprehensive overview of contemporary approaches, considering issues like dynamic feature sets, developing data, windowing, ensemble methods, and more. To fully comprehend and handle IoT data, it discusses research problems and future goals.

In the views of Haghi *et al.* (2022), healthcare systems have difficulties due to the changing demographics in Europe and North America, characterized by a sharp rise in the old population, potentially lowering the quality of the provided services. A practical option, telemedicine enables bidirectional contact with medical experts and real-time remote monitoring. It is possible to predict and prevent diseases before symptoms appear using smart homes as diagnostic spaces for ongoing, unobtrusive health monitoring. To fulfil the WHO's comprehensive evaluation areas, this review examines sensing tools and technologies used in smart homes. A bus-based scalable intelligent system's facilitation of integration across environmental, behavioural, physiological, and psychological data results in a hybrid architecture for hierarchical multi-layer data fusion, facilitating both event detection and long-term prediction and prevention.

In the views of Wang *et al.* (2021), this study explores unobtrusive in-home health monitoring using big data, artificial intelligence, and sensor technology improvements. The study undertakes a literature analysis to inform current sensor technology applications for in-home health monitoring, building on earlier work linked to smart automobiles. The study analyses 55 pertinent studies, addressing issues with sensor kinds, location, monitored data, and supporting functions, and identifies 25 appropriate sensor types for ambient situations. Physiological parameters are constrained, but behavioural data is simple to record. Functional, emergency, safety, security, and social interaction monitoring are examples of monitoring functions. However, with extensive, long-term monitoring, there still needs to be more solid evidence linking these functions to clinical outcomes.

According to Fahim and Sillitti (2019), the development of sensor technology and the availability of affordable solutions are driving the explosion in anomaly detection research. Massive data streams produced by physical space monitoring provide chances to spot unusual behaviour, reduce hazards, and avoid system outages. This study thoroughly assesses the literature on anomaly detection methods used in various industries, including intelligent habitats, transportation, healthcare, and industry. The study, which covers the years 2000 to 2018, reveals research gaps in data collecting, the processing of unbalanced datasets, the limits of statistical approaches with large amounts of sensory data, and a need for studies on predicting anomalous behaviour in real-world situations. The findings give academics and practitioners new perspectives on approaching problems and adding creative solutions for anomaly identification.

According to Morita *et al.* (2023), the proliferation of Internet of Things (IoT) devices in daily life offers opportunities for continuous behavioural monitoring through smart home technologies, particularly in supporting population health. This scoping review, spanning 2008 to 2021 and analyzing 49 papers from databases like PubMed and Scopus, reveals a concentration of studies in Europe and North America, primarily proof of concept or pilot studies. Although data frequently concentrates on engineering and technological factors, participant differentiation, technology interoperability, and data security are issues. To successfully incorporate smart home technology into healthcare, the research exposes knowledge gaps and emphasizes the necessity for multidisciplinary collaboration. The dynamic environment is reflected in a new definition of a smart home focused on health care.

According to DeFranco and Kassab (2021), a smart home (SH) uses cutting-edge technology and gadgets to improve occupants' comfort, energy efficiency, privacy, safety, and the security of their data. Despite the complexity and variety of these houses, there needs to be more agreement on the potential directions for further study into the SH idea. A thorough literature review is conducted to find and evaluate the most recent SH research to fill this gap. The review's conclusions help create a taxonomy for SH research that will be valuable for navigating and expanding on the many future subfields of smart home research.

In the views of Taiwo *et al.* (2022), For a higher standard of life, protecting people's lives and property must come first. The Internet of Things (IoT) and technology advancements have substantially improved remote control, monitoring, and online security thanks to smart home automation development. While current home security systems concentrate on motion detection and tracking, the problem of preventing pointless notifications must still be solved. This study proposes a deep learning-based approach for intelligent home automation that can classify and recognize motion. By differentiating between intruders and house inhabitants based on movement patterns recorded by a security camera, the suggested system eliminates false alarms and achieves an impressive 99.8% accuracy in experimental analysis.

According to Javaid *et al.* (2021), sensor technologies have considerably improved human existence in various contexts by detecting environmental changes and producing relevant reactions. These sensors, which can detect light, temperature, motion, and pressure, are used in multiple industries, including manufacturing, healthcare, and fitness. By sending reminders and dispensing dosages at predetermined intervals, sensor-equipped drug dispensers in healthcare improve prescription adherence. Modern sensor technology benefits high-risk patients, athletes, and the elderly. Ultrasound, radar, non-contact optoelectronics, and laser technology are among the newest trends. This article briefly introduces several sensor types and how they are used daily, highlighting their critical role in data interchange, healthcare, and the smooth operation of numerous sectors.

According to Jabbar *et al.* (2019), Interest in home automation systems has increased with the development of communication technologies. To solve the shortcomings of existing systems, this study offers IoT@HoMe, an affordable hybrid IoT-based home automation system. IoT@HoMe connects several sensors to an Adafruit IO cloud server using a NodeMCU as a Wi-Fi gateway. Using If This Then That (IFTTT) on smartphones or laptops, sensors such as RFID, ultrasonic, temperature, humidity, gas, and motion sensors enable remote monitoring. The system uses relays to manage residential appliances, boosting security and promoting energy conservation. IoT@HoMe, designed as a transportable control box, provides a dependable, affordable solution, guaranteeing comfort, security, and safety for occupants of smart homes.

Tukur *et al.* (2021), the processing of data inside the Internet of Things (IoT) systems, including analytics, big data processing, and examining sensor data for long-term patterns, is a critical function of IoT platforms. However, transferring data from resource-constrained IoT devices to these platforms presents difficulties, leading to low availability and the possibility of data loss. To address these problems, this article suggests an edge-based blockchain-enabled anomaly detection method. Using edge computing, the process improves availability and prevents single points of failure while lowering latency and bandwidth requirements. For secure and effective anomaly detection and correction in IoT data, it combines distributed edge computing with blockchain, sequence-based anomaly detection and ensures data availability and integrity.

Del-Valle-Soto *et al.* (2023), In medical applications, wireless sensor networks (WSNs) are helpful, especially for monitoring the vital signs of geriatric patients. The elderly patients' houses are equipped with this low-cost WSN, which provides daily suggestions and remote monitoring. This sensor network's effects on nine vital sign indicators relating to sleep patterns are evaluated by the study. Data from 30 persons collected over four weeks shows that continuous monitoring improves sleep parameters, bringing safety and well-being to the home. To reduce energy consumption and improve network resilience, the research explores links between impacted metrics, predicts trends, and presents a reactive energy and performance optimization algorithm.

According to Taiwo and Ezugwu (2021), the smart home has emerged as a significant study area With the integration of Internet of Things (IoT) technologies for interactive and pleasant living. The primary goal of this effort is to design and create a cloud-based intelligent home automation system that uses an Android mobile application for management, security, and control. Modules control security, environmental conditions, and electrical appliances through motion detection and picture capture. Machine learning, especially the support vector machine method, distinguishes between regular residents and invaders to avoid erroneous alerts. The ESP8266 and ESP32-CAM boards used in the prototype implementation show how machine learning may improve the functionality and security of home automation systems.

In the views of Maswadi *et al.* (2020), despite being a critical component of the Internet of Things (IoT), smart home technology, especially for the elderly, has not yet undergone a systematic literature review (SLR). This study closes this gap by doing an SLR on adopting smart home monitoring technologies. Review articles were published between January 2010 and December 2019 using standards for evaluating quality. Only 3% of the 73 pertinent central studies adhered to strict SLR standards, whereas 7% and 8% did so to varied degrees. The report encourages researchers to use extensive SLR standards to guarantee a high level in the following research on smart home technologies.

According to Ali *et al.* (2019), the development of a smart home-based monitoring system for senior patients is described in this study, with an emphasis on the continuous monitoring of four physiological parameters: temperature, blood glucose levels, and 3D accelerometer and gyroscope data for fall detection. Wearable and contextual sensors use The nRF communication protocol to communicate data to smart gateways. Several tasks are carried out at the fog layer, including delivering alerts to the patient's mobile phone, transmitting gathered measures to healthcare practitioners for study, and notifying emergency services in case of crises. An effective thermal energy harvesting system is created, attaining an 82.6% efficiency with a 20°C temperature differential to provide power autonomy and lessen the need for regular sensor node battery replacement.

As per the views of Sokullu *et al.* (2020), The goal of technological advancements, particularly in the healthcare industry, is to improve human welfare and make daily living easier. This article describes a smart home Internet of Things system specifically created for seniors and people with memory impairments to ensure their safety and provide early warnings for everyday issues related to such diseases. Continuous data gathering from ambient sensors makes contextual activity detection possible, allowing for unobtrusive monitoring of the patient's behaviour at home. The system also includes a bracelet for emergencies such as falls or exposed gas or water sources. The prototype is a reasonably priced option focusing on low energy usage and smooth integration for broader applications.

Table

Authors	**Name of the Paper**	**Remarks**
Erhan, Ndubuaku, Di Mauro, Song, Chen, Fortino, Bagdasar, and Liotta.	Smart anomaly detection in sensor systems: A multi-perspective review.	The authors emphasize the integration of information and look at several techniques to improve the effectiveness of anomaly identification. This review's insights help shape how sensor system security is changing.

Mansour, Escorcia-Gutierrez, Gamarra, Villanueva, and Leal.	Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model.	This work introduces the IVADC-FDRL model, a Faster RCNN with a Deep Reinforcement Learning-based Intelligent Video Anomaly Detection and Classification system. The model is divided into two stages: Faster RCNN and Residual Network are used for anomaly detection, and deep Q-learning-based DRL is used for anomaly categorization.
Mokhtari, Abbaspour, Yen, and Sargolzaei.	A machine learning approach for anomaly detection in industrial control systems based on measurement data.	Industrial Control Systems (ICSs) struggle to identify assaults using traditional network traffic monitoring because attackers might fool them by imitating regular activity. With the help of data from supervisory control and data acquisition (SCADA) systems, this study proposes a unique solution called the Measurement Intrusion Detection System (MIDS).
Ni, Zhang, and Noori,	Deep learning for data anomaly detection and data compression of a long-span suspension bridge.	The widespread use of low-cost online structural health monitoring (SHM) has been made possible by intelligent sensing and sensor network systems developments, leading to the capture of enormous and diverse data for large-scale civil structures.
Nassif, Talib, Nasir, and Dakalbab,	Machine learning for anomaly detection: A systematic review.	This research study uses a Systematic Literature Review (SLR) to examine ML models for anomaly detection. The study examines applications, ML approaches, performance measures, and classification in anomaly detection through analysis of 290

		research publications published between 2000 and 2020.
Al-amri, Murugesan, Man, Abdulateef, Al-Sharafi, and Alkahtani	A review of machine learning and deep learning techniques for anomaly detection in IoT data.	The development of technologies like the Internet of Things (IoT), which generate enormous continuous data streams from various applications, anomaly detection has attracted a lot of Interest. Analyzing this data to spot suspicious activity, counter functional risks, and avoid unanticipated problems resulting in application outages is essential.
Haghi, Spicher, Wang, and Deserno.	Integrated sensing devices for disease prevention and health alerts in smart homes.	Healthcare systems have difficulties due to the changing demographics in Europe and North America, characterized by a sharp rise in the old population, potentially lowering the quality of the provided services. A practical option, telemedicine enables bidirectional contact with medical experts and real-time remote monitoring.
Wang, Spicher, Warnecke, Haghi, Schwartze, and Deserno.	Unobtrusive health monitoring in private spaces: The smart home.	The study undertakes a literature analysis to inform current sensor technology applications for in-home health monitoring, building on earlier work linked to smart automobiles. The study analyses 55 pertinent studies, addressing issues with sensor kinds, location, monitored data, and supporting functions, and identifies 25 appropriate sensor types for ambient situations.
Fahim and Sillitti	Anomaly detection, analysis and prediction techniques in IoT environment: A	This study thoroughly assesses the literature on anomaly detection methods used in various industries, including intelligent habitats, transportation, healthcare, and industry.

	systematic literature review.	
Morita,Sahu and Oetomo,	Health Monitoring Using Smart Home Technologies: Scoping Review.	This scoping review, spanning 2008 to 2021 and analyzing 49 papers from databases like PubMed and Scopus, reveals a concentration of studies in Europe and North America, primarily proof of concept or pilot studies.

Chapter 3: Methodology

Cross-Industry Standard Process for Data Mining (CRISP-DM):

The Cross-Industry Standard Process for Data Mining, or CRISP-DM, is a widely used approach that provides direction for data mining initiatives. It was first introduced in 1996 and consists of six primary phases: business understanding, data preparation, deployment, modeling, evaluation, and assessment. It offers an organized and thorough approach. CRISP-DM, which is well-known for its flexibility and practicality, places a strong emphasis on working together between data and business specialists at every stage of the project to ensure that data mining solutions are deployed effectively and in line with business goals.

Knowledge Discovery in Databases (KDD):

The technique of extracting useful information from massive datasets is known as knowledge discovery in databases, or KDD. Data cleansing, integration, selection, transformation, pattern recognition, and interpretation are some of the stages that are involved. Beyond conventional data mining methods, knowledge discovery may be conducted using the entire framework provided by KDD.

CRISP-DM seems like a better option in the particular context of a smart senior home monitoring project that focuses on anomaly detection and safety enhancement through the use of gas, temperature, and location sensors. Its methodical and staged methodology guarantees a methodical investigation and application of data-driven insights, fitting in nicely with the usual phases of a data mining project. Furthermore, CRISP-DM is ideally suited to handle the unique requirements and objectives of a project with real-world applications like smart home monitoring because of its emphasis on business understanding and adaptability.

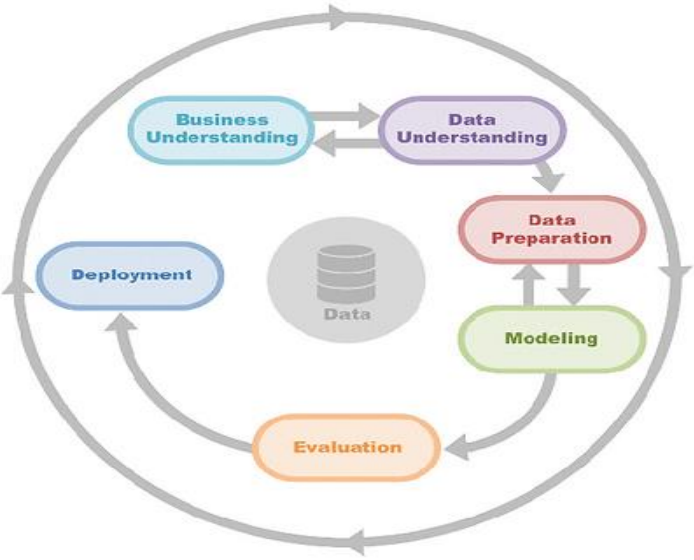

Fig 1: CRISP-DM Methodology

The CRISP-DM steps are as follows:

1. Business Understanding: Defining the objectives of the smart senior home monitoring system with an emphasis on anomaly detection for safety enhancement is the first step in the CRISP-DM process. Tasks include identifying anomalies in temperature fluctuations, location data, and gas levels; determining how anomaly detection adds to overall safety objectives; and comprehending the safety needs for senior people in a home environment.

2. Data Understanding: The objective of this stage is to gain knowledge from the data produced by location, temperature, and gas sensors. Investigating sensor data properties, seeing trends and anomalies, and comprehending data quality are among the tasks. It entails confirming the accuracy of the data and comprehending its temporal components, such as changes during the day or season.

3. Data Preparation: The objective of this stage is to clean up and arrange the data so that it can be used effectively for modeling anomaly detection. The tasks involve picking pertinent features from the sensor data, managing outliers and missing values, scaling or normalizing the data, and formatting it so that it can be used with the anomaly detection methods of choice, such as local outlier factor, DBSCAN, isolation forest, autoencoder, one-class SVM, and autoencoder.

4. Modeling: Using specific methodologies, the modeling phase focuses on creating models for anomaly identification. One-class SVM, autoencoder, DBSCAN, isolation forest, and local outlier factor models are among the tasks to be completed. The models' parameters are changed for best performance once they are trained using prepared datasets. Models that are most successful at detecting anomalies are identified through evaluation using relevant metrics.

5. Evaluation: This stage entails determining how well the anomaly detection models work. One of the tasks is to assess the model's performance using measures like F1-score, accuracy, and recall. It is essential to validate against real-world circumstances, and models may be improved as a result of the findings. By striking a balance between false positives and false negatives, the system may be made to be attentive to safety problems while avoiding the needless raising of alarms.

6. Deployment: The smart senior home monitoring system's anomaly detection models are integrated during this last step. The models must be integrated into the operational environment. Continuous monitoring and updating procedures must be established. Documentation and instructions for caregivers or users on how to understand and react to anomaly warnings must also be provided.

The project methodology diagram are as follows:

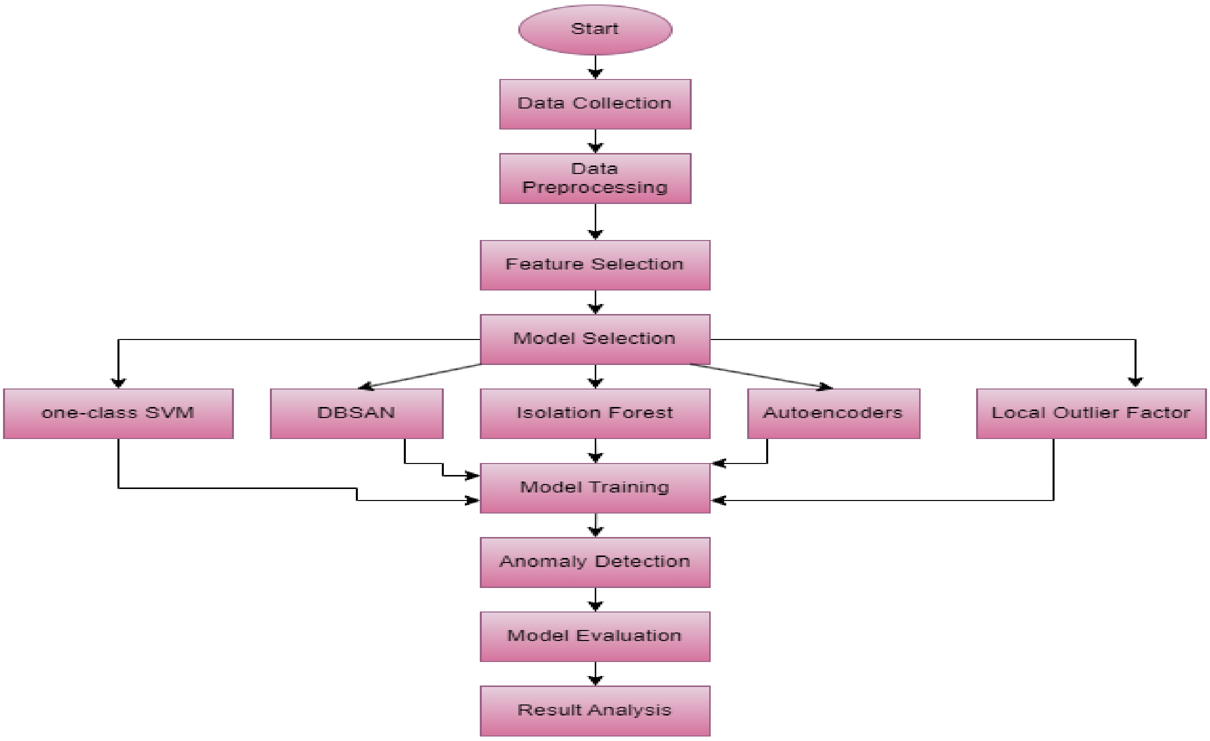

Fig 2: Project Methodology

Project Requirement: Google Colab was chosen as the project's software platform because of its simplicity of use and functionality, which is similar to Anaconda Navigator. This allowed for the elimination of installation requirements and simplified file sharing.

Python: Due to its numerous libraries, which include numpy, Matplotlib, and scikit-learn, and its adaptability, Python was the preferred language. This made it possible to construct models, evaluate metrics, and manipulate data without difficulty.

Scikit-learn: Scikit-learn provided us with useful tools to easily create and test our models in Python, which made our project go more smoothly.

Matplotlib: For our study, Matplotlib made it simple for us to display data trends through the creation of vibrant, understandable charts.

Machine learning models: In this study, I used machine learning models like one-class SVM, DBSCAN, Isolation Forest and local outlier factor.

TensorFlow: We used TensorFlow, a useful tool, to develop and train machine learning models in Python.

Task of the Study:

Task 1: Preparing the Proposal: The project's aims, importance, and approach should be outlined in a comprehensive proposal. Describe the objectives and range of work.

Task 2: Literature Review: Examine and evaluate twenty-five to thirty-five pertinent papers that you have found in books, journals, and conference proceedings. Provide a synopsis of recent advances in the fields of gas sensing, anomaly detection, and home monitoring.

Task 3: Data Collection and Merging: Download the dataset (database_gas.csv and database_pos.csv) as two CSV files from the UCI source. A single dataset for analysis is produced by combining the datasets using a shared identifier.

Task 4: Data Exploration: Investigate the combined dataset to learn about its variables, trends, and structure. Find any abnormalities, outliers, or missing values. Obtain a thorough understanding of the information.

Task 5: Data Preparation: Address any redundant information, outliers, or missing values. To get the dataset ready for analysis, do data transformation, standardization, and cleaning.

Task 6: Feature engineering: To improve the dataset's relevance and prediction capability, add new features or alter existing ones. This might entail mixing already-existing variables or generating new ones.

Task 7: Model Selection and Implementation: To detect anomalies, select the isolation forest algorithm. To create predictions for the target variable (anomaly), apply the algorithm to the preprocessed dataset. In this case, I additionally employ models such as local outlier factors, DBSCAN, autoencoders, and one-class SVM.

Task 8: Model Evaluation: Using relevant measures like accuracy, precision, recall, and F1 score, assess the isolated forest model's performance.

Task 9: Result Analysis: Examine the anomaly detection model's and other outcomes.

Task 10: Writing a Report: Compile all the study's findings into a thorough report. Provide specifics about the research design, results, analysis, and conclusions.

Gantt Chart:

Task/Week	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10
Task 1										
Task 2										
Task 3										
Task 4										
Task 5										
Task 6										
Task 7										
Task 8										
Task 9										
Task 10										

Table 1: Gantt Chart

Chapter 4: Dataset Exploration and Visualization

4.1. Data Description:

In this study, I have taken this dataset from the UCI repository, and the link is https://archive.ics.uci.edu/dataset/799/single+elder+home+monitoring+gas+and+position. This dataset includes movement infrared sensors, temperature and gas sensors, and monitoring data from 2019-11-06 to 2020-02-13 for an elderly individual living alone in their own house. The temporal resolution of the measurements is 20 seconds. In this file, there are two datasets present. database_gas.csv and database_pos.csv. In database_gas.csv, there are 10 columns present, and in database_pos.csv, there are 6 columns present. In this study we merged both the datasets and after merging the dataset, here we get the 28478 rows and 16 columns. The columns are temperature, humidity, CO2CosIRValue, CO2MG811Value, MOX1, MOX2, MOX3, MOX4, COValue, Kitchen, Living room etc. The columns CO2CosIRValue and CO2MG811Value likely indicate measurements of carbon dioxide levels using infrared and MG811 sensors, respectively. MOX1, MOX2, MOX3, and MOX4 columns represent readings from different Metal Oxide gas sensors, offering insights into various gas concentrations. The COValue column may represent carbon monoxide levels, contributing to a comprehensive air quality assessment. In this dataset, the target variable is anomaly, which I have created using the isolation forest algorithm.

4.2. Data analysis using tableau:

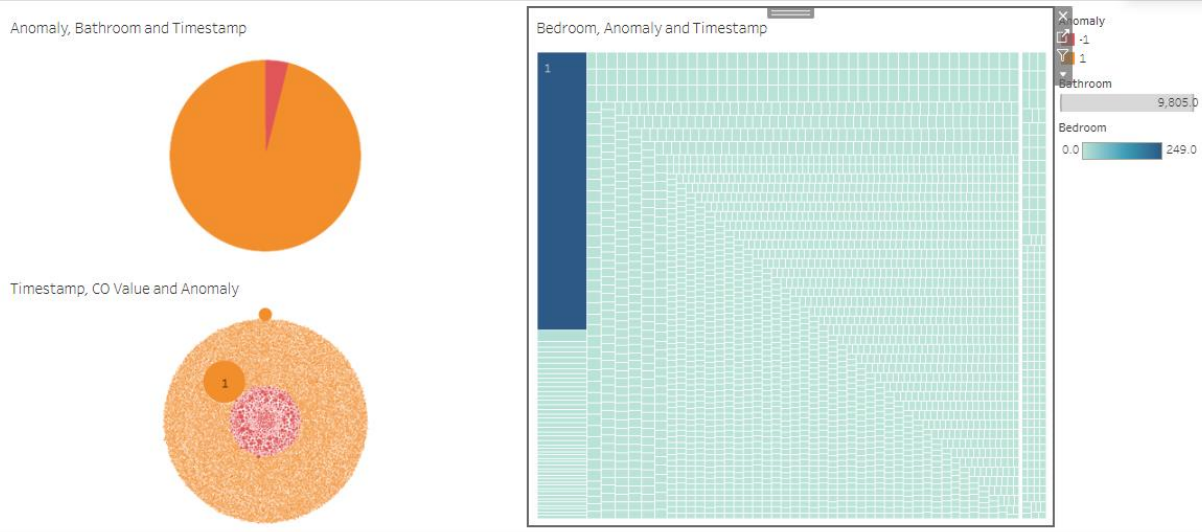

Fig 3: Dashboard 1

In the above dashboard, the first worksheet shows the pie chart for anomaly, bathroom and timestamp, here pink color indicates anomaly, and the orange color indicates normal, here normal has the maximum count as compared to anomaly. The second worksheet shows the treemap for bedroom, anomaly, and Timestamp. Here, the dark region has the maximum count of room occupied as compared to the light region. The third worksheet shows the bubble chart for timestamp, CO Value and Anomaly, here timestamp 2019-11-10 has the maximum count of CO Value as compared to other timestamps.

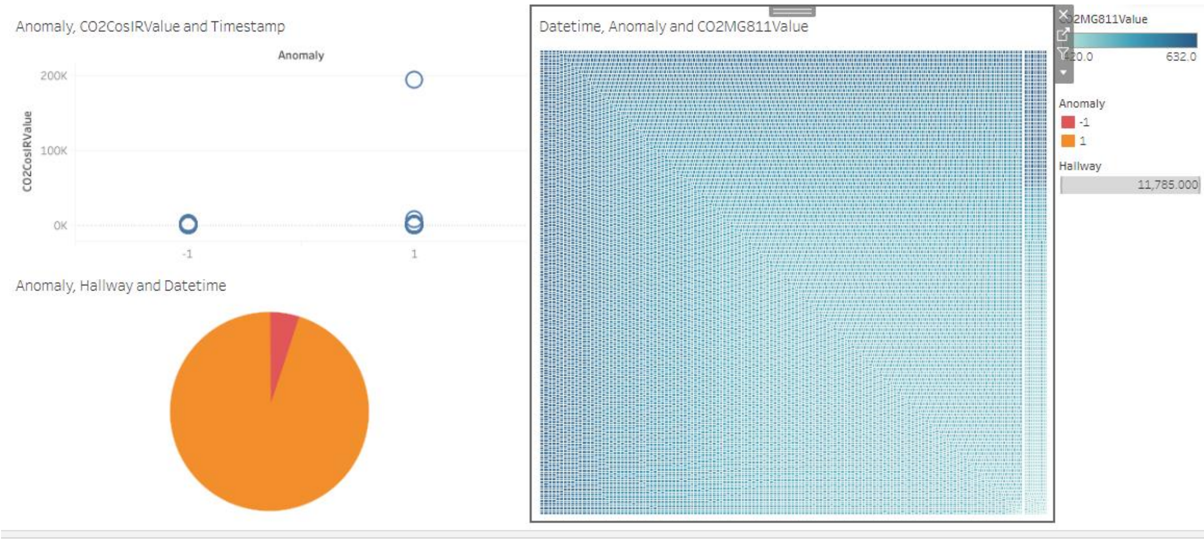

Fig 4: Dashboard 2

In the above dashboard, the first worksheet shows the circle view for anomaly, CO2CosIRValue and Timestamp, here 2019-11-06 timestamp has the maximum count of CO2CosIRValue as compared to other timestamps. The second worksheet shows the treemap for Datetime, Anomaly and CO2MG811Value, here dark region indicates the maximum count of CO2MG811Value as compared to the light region. The third worksheet show the pie chart for anomaly, hallway and Datetime, here normal has a maximum count as compared to anomaly.

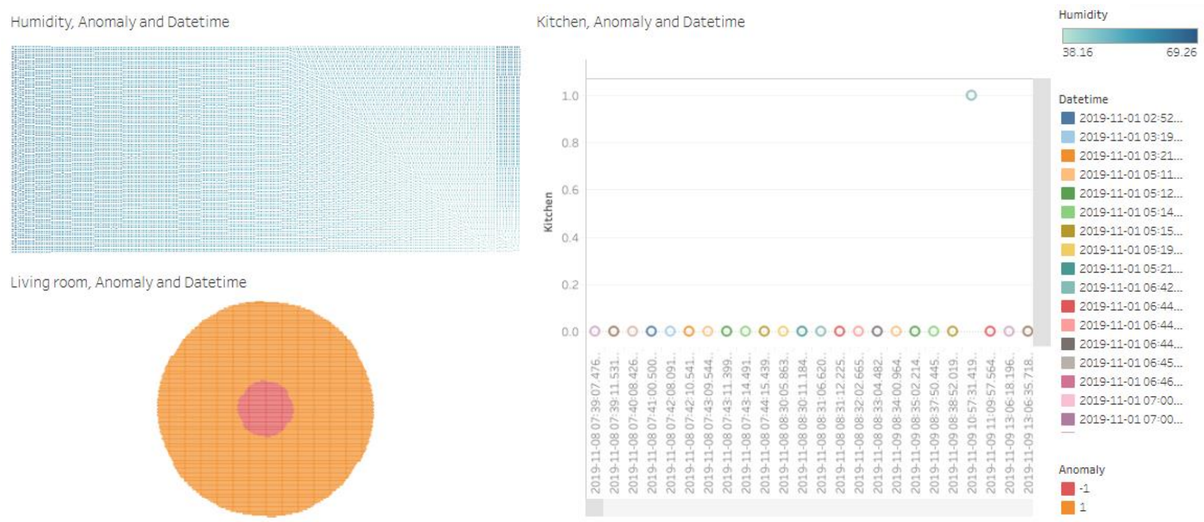

Fig 5: Dashboard 3

In the above dashboard, the first worksheet shows the treemap for humidity, Anomaly and Datetime, the dark region indicates the maximum anomaly as compared to the light region. The second worksheet shows the circle view for Kitchen, Anomaly and Datetime, here different colors indicate the different Datetimes. The third worksheet shows the bubble chart for Living room, Anomaly and Datetime, here pink region indicates -1 value and the orange color indicate 1 value.

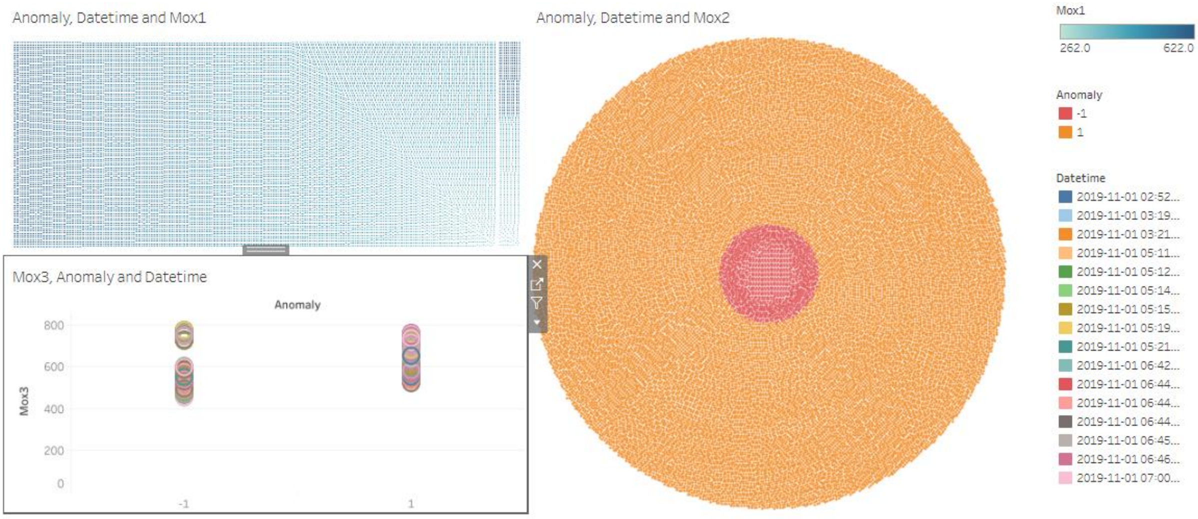

Fig 6: Dashboard 4

In the above dashboard, the first worksheet shows the treemap for anomaly, datetime and Mox1. Here dark region indicates the maximum count of Mox1 as compared to the light region. The second worksheet shows the bubble chart for Anomaly, Datetime and Mox2, here pink region indicates -1 value and orange region indicates 1 value. The third worksheet shows the circle view for Mox3, Anomaly and Datetime, here different colors indicate the different datetimes.

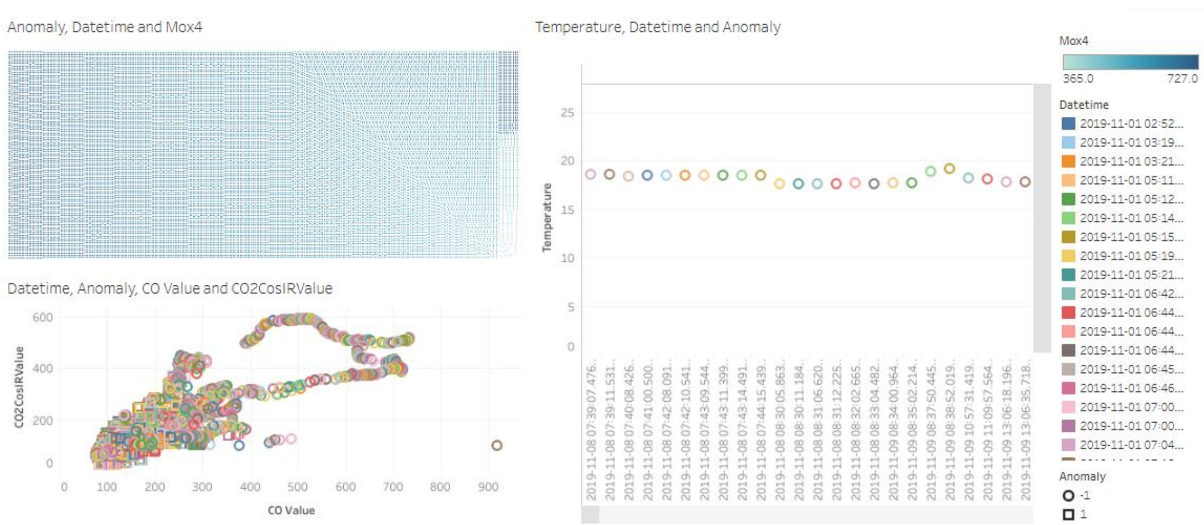

Fig 7: Dashboard 5

In the above dashboard, the first worksheet shows the treemap for anomaly, datetime and Mox4. The dark region indicates the maximum Mox4 as compared to the light region. The second worksheet shows the circle view for Temperature, Datetime and Anomaly, here different colors indicate different datetimes. The third worksheet shows the scatterplot for datetime, Anomaly, CO Value and CO2CosIRValue, where CO Value is on the x axis and CO2CosIRValue is on the y axis.

4.3 Data Visualization using Python:

Matplotlib has been loaded under the Data Visualization area. We just use this library pyplot each time we use it, and we are able to generate graphs using countplot, bar plots, pie charts, and other formats. Using the Seaborn library, we were able to build graphs that are easier to comprehend and take less time to construct. The graphs used by Seaborn Library are box plots and distribution plots.

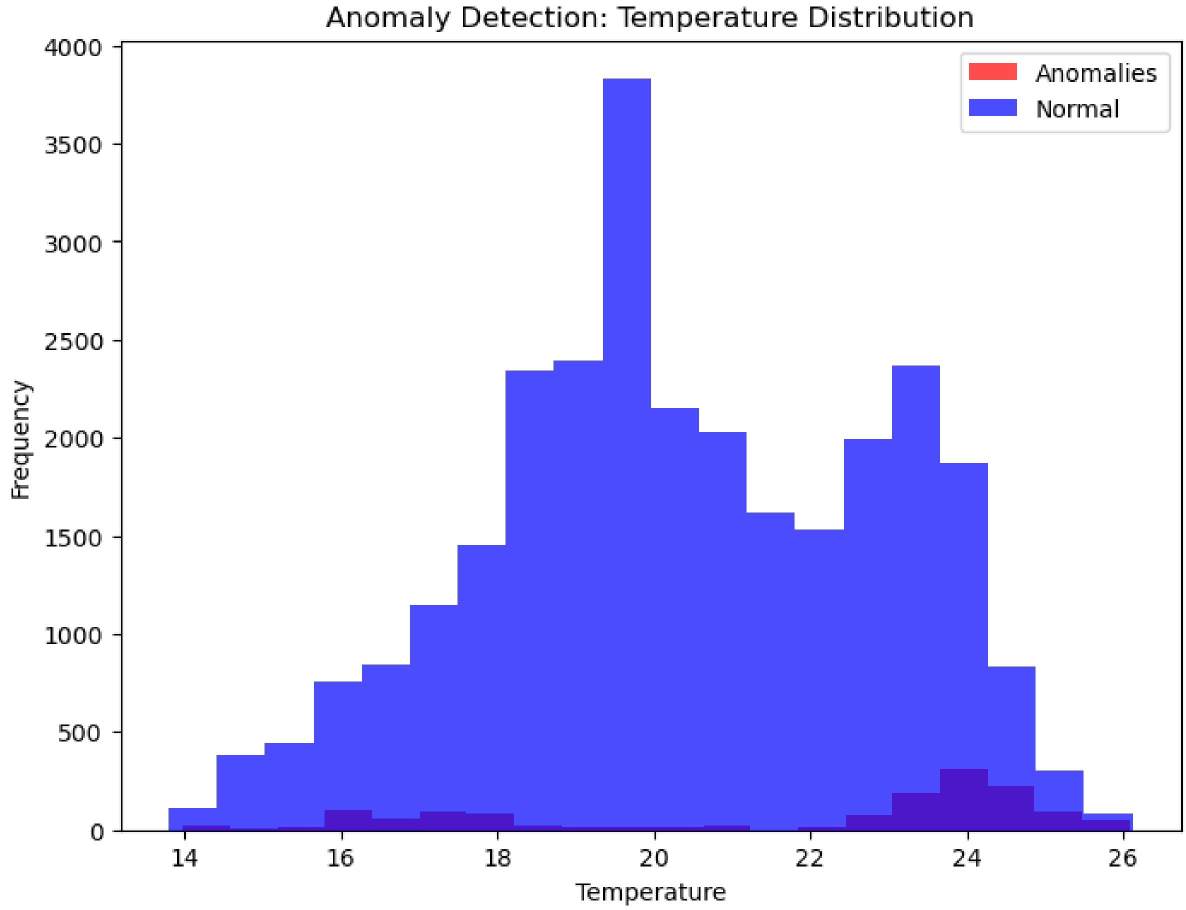

Fig 8: Histplot for Temperature using anomaly

A histogram is a graphic representation of a dataset's distribution that shows the likelihood or frequency of certain values falling into predefined ranges, or "bins." It is helpful for recognizing patterns and trends since it offers a visual representation of the data's form, distribution, and central tendency. In the above figure, Temperature is on the x axis and frequency is on the y axis, red color indicates anomalies, and blue color indicates normal.

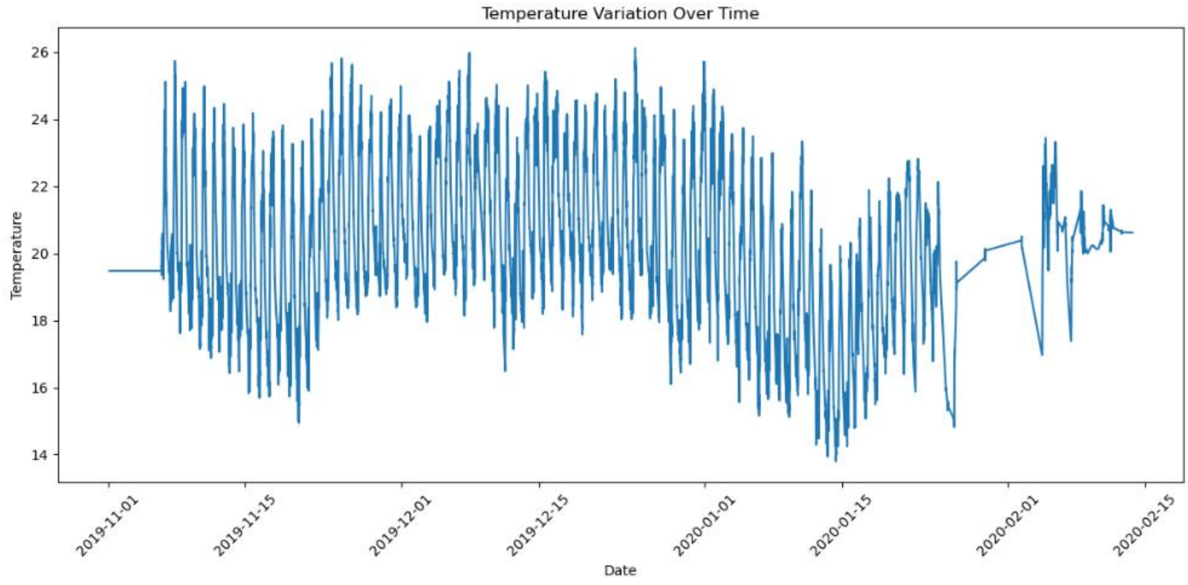

Fig 9: Lineplot for date and temperature

A line chart is a type of graphical representation where data points are connected by straight line segments in the dataset's order of appearance. Displaying trends and patterns across a continuous variable is a typical use, which enhances its effectiveness in displaying data changes and linkages. The figure above shows the line chart for Date and Temperature, where Date is on the x axis and temperature is on the y axis. The trend of temperature variations over time is graphically shown by the lineplot.

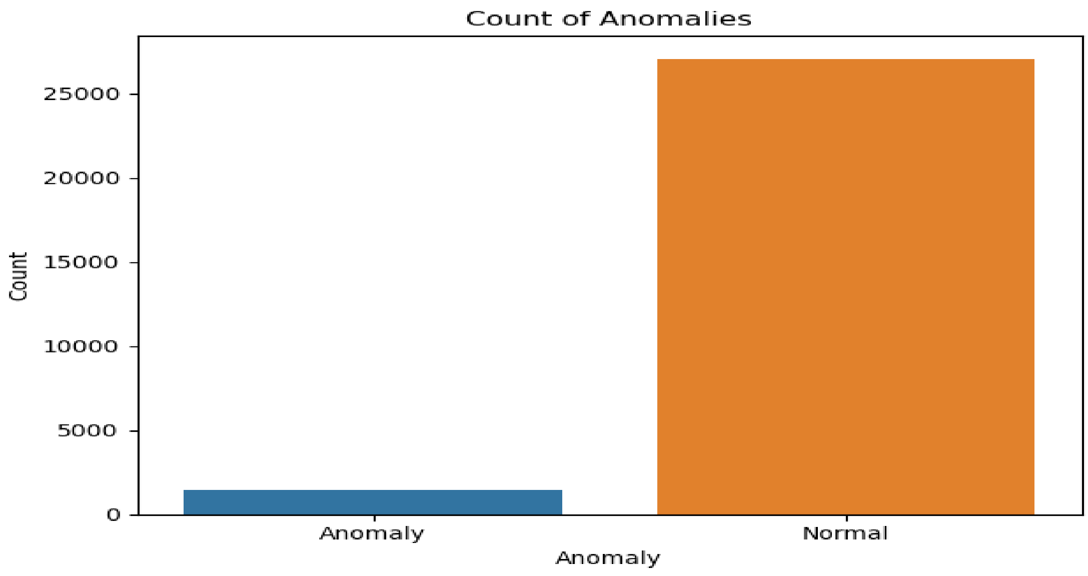

Fig 10: Countplot for Anomaly

In the above figure, the countplot is a plot for Anomaly. Anomalies and normal occurrences are displayed in the Seaborn count plot that is supplied, according to the 'anomaly' column in the 'merged_data' DataFrame. Prominently, the figure indicates that the majority of occurrences are normal (designated 1), with the largest count in comparison to anomalies (labeled -1). This implies that the dataset contains a greater frequency of normal instances.

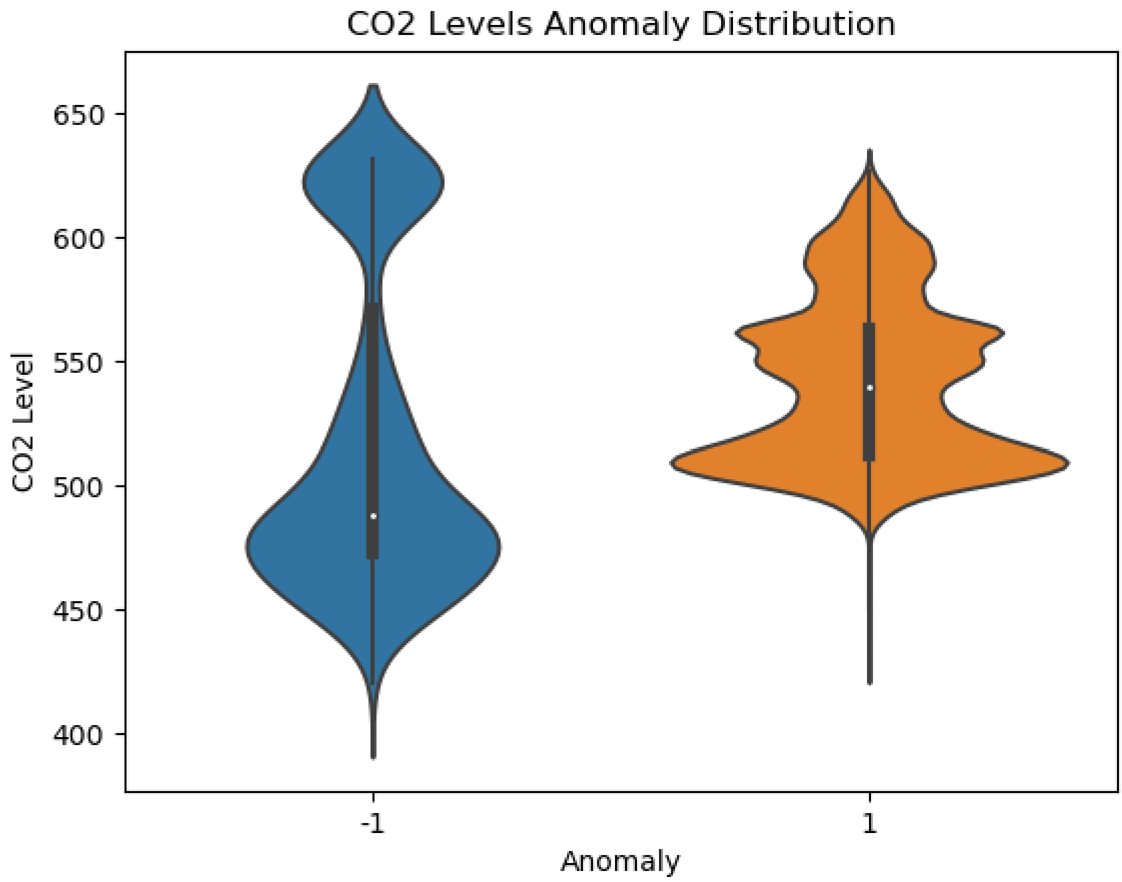

Fig 11: Violinplot for Anomaly and CO2 Level

A violin plot is a type of data visualization that blends elements of a kernel density plot with a box plot. In addition to giving a succinct but thorough overview of the dataset's changes, it offers insights into the distribution and probability density of the data across several categories. It is an effective tool for comprehending the central tendency and distribution of the data since the width of the plot at each point indicates the data density, and the center box indicates the interquartile range. In the above figure, Violinplot is a plot for Anomaly and CO2 Level where -1 indicates anomalies and 1 indicates normal.

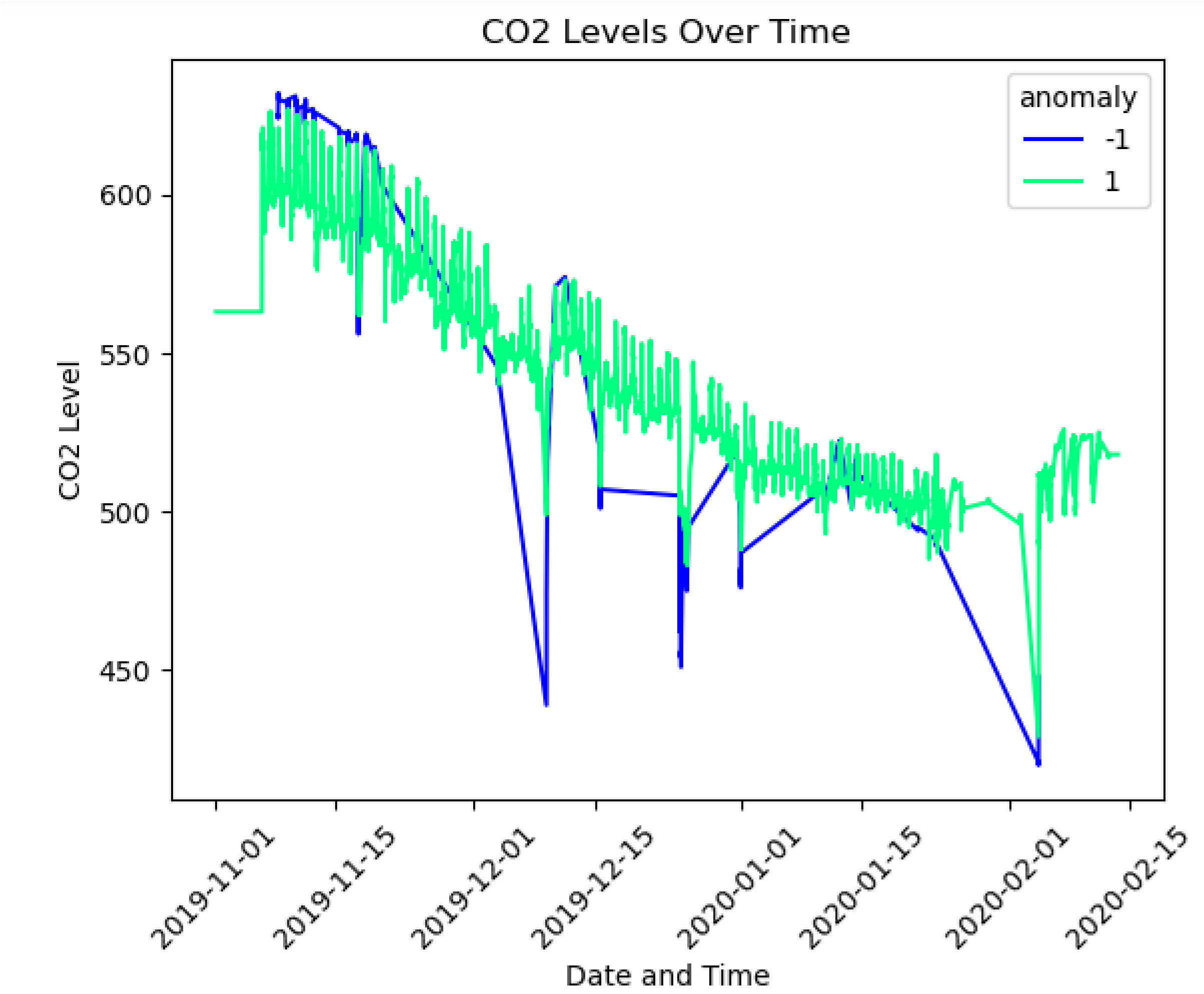

Fig 12: Lineplot for Date and CO2 Level

This Seaborn line plot uses lines to show the 'CO2MG811Value' across time ('datetime,') and differentiate anomalies from typical occurrences ('anomaly'). This makes it easier to spot any inconsistencies by giving a concise and clear picture of how CO_2 levels change across various time periods. Here, green color indicates normal and blue color indicates anomaly.

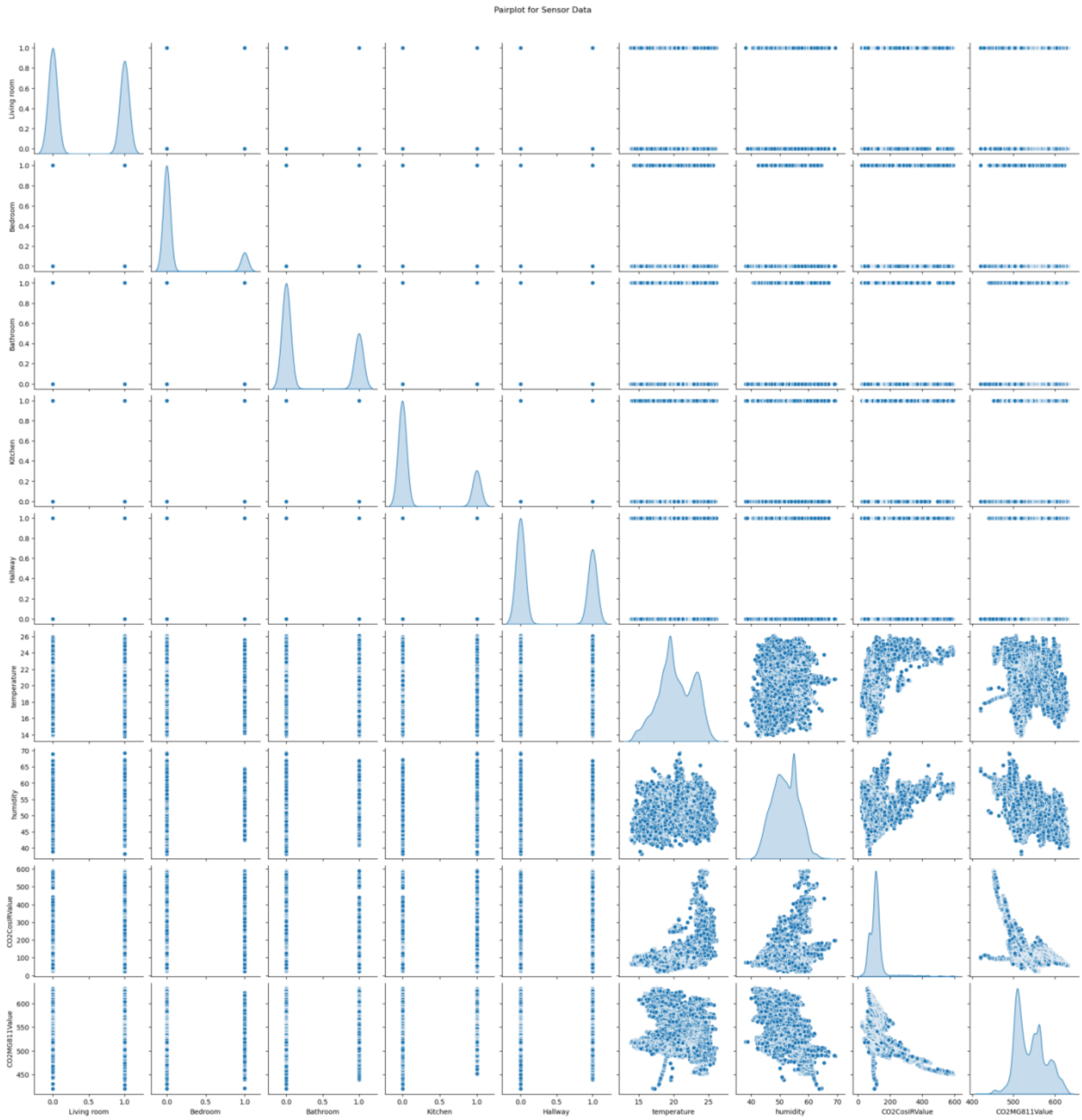

Fig 13: Pairplot for dataset

Relationships and distributions between the variables "Living room", "Bedroom", "Bathroom", "Kitchen", "Hallway", "temperature", "humidity", "CO2CosIRValue", and "CO2MG811Value" are graphically explored in this Seaborn pair plot. While scatterplots reveal possible patterns and connections between various pairs of variables in the dataset, diagonal KDE (Kernel Density Estimation) plots give a smooth depiction of univariate distributions.

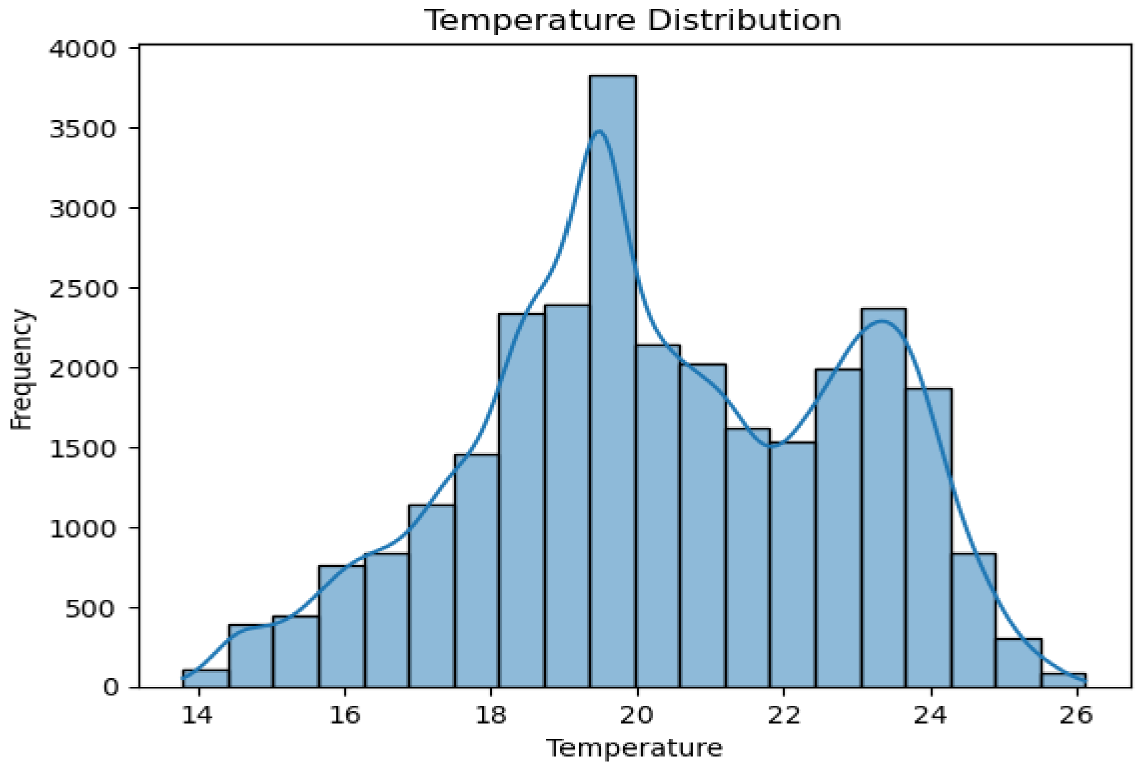

Fig 14: Histplot for Temperature

The smoothed overview of the data distribution provided by the overlay Kernel Density Estimate (KDE) curve facilitates the analysis of temperature trends and concentration within the designated bins. Here histplot is a plot for temperature, where temperature is on the x axis and frequency is on the y axis.

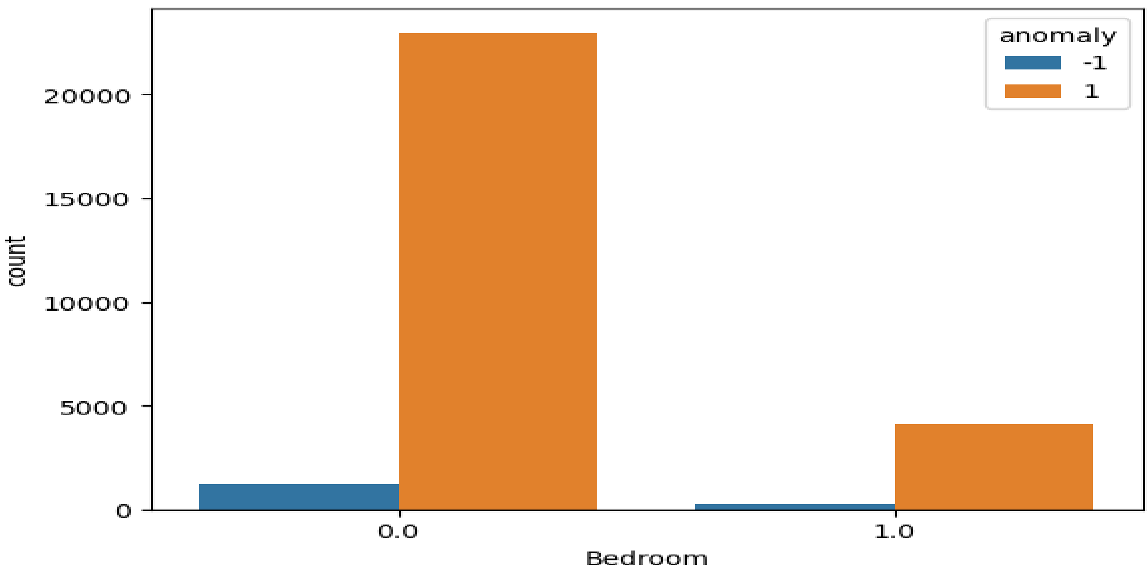

Fig 15: Countplot for Bedroom

In the above figure, the countplot is a plot for Bedroom using the anomaly. Bars are used in this Seaborn count plot to distinguish between normal occurrences and abnormalities (referred to as "anomalies") in the distribution of "Bedroom" instances in the "merged_data" DataFrame. Finding any trends or inconsistencies in the data is made easier by the clear depiction of how anomalies and typical cases are dispersed within the 'Bedroom' category.

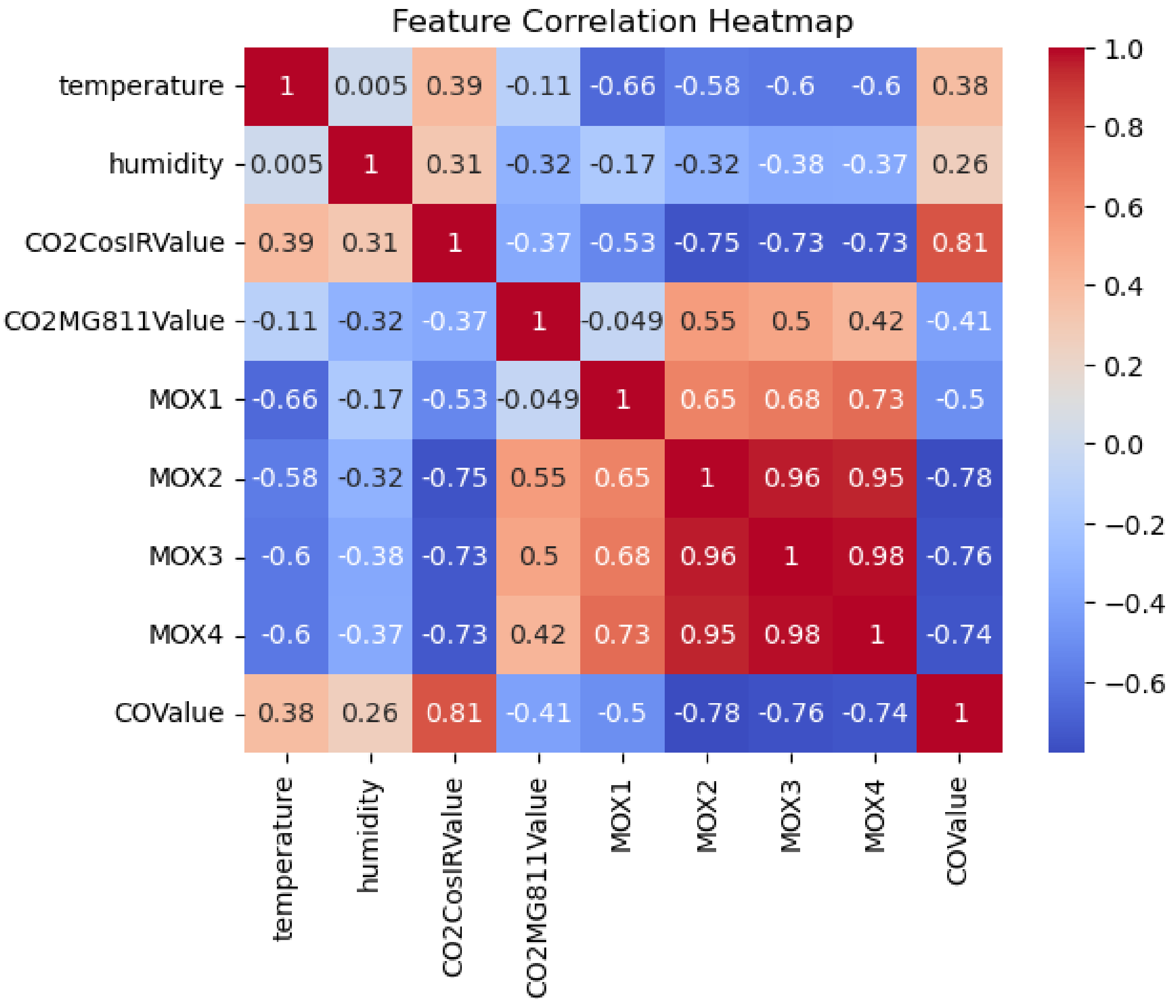

Fig 16: Heatmap for data

A heatmap is a matrix-format visual display of data that shows values as a set of colors to illustrate trends and variances. The correlation matrix is shown via this Seaborn heatmap, which shows the correlations between the variables. Annotation is enabled to display numerical values on the heatmap. The 'coolwarm' colormap is applied, where warmer colors reflect positive correlations and colder colors represent negative correlations. Gaining an understanding of the direction and degree of correlations within the dataset is possible with the help of this visualization.

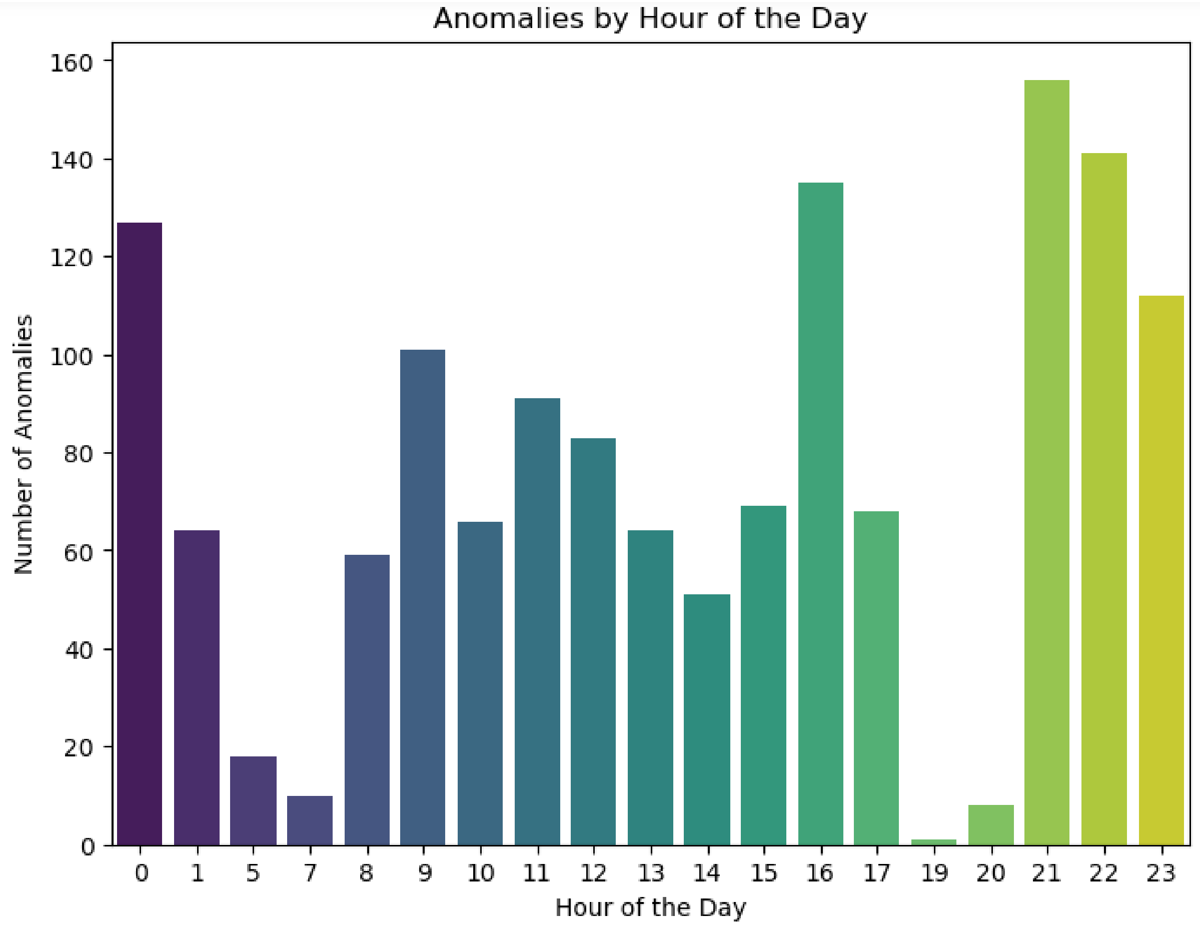

Fig 17: Barplot for Hour of the Day and Number of Anomalies

In the above figure, the bar plot is a plot for the hour of the day and the number of anomalies, where for the 21^{st} hour, I get the highest number of anomalies as compared to other hours of the day.

4.4 Data Preparation:

An essential phase in the pipeline for data analysis and machine learning is data pretreatment. It entails sanitizing and converting unprocessed data into a format that algorithms can comprehend and use with ease. The efficacy and precision of machine learning models can be greatly enhanced by appropriate preprocessing of the data. Before the data preprocessing, I first imported the various libraries.

```python
#import all needed libraries
import numpy as np
import pandas as pd
from sklearn.svm import OneClassSVM
from sklearn.ensemble import IsolationForest
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import make_scorer
from sklearn.model_selection import GridSearchCV
from tensorflow.keras.layers import Input, Dense
from sklearn.cluster import DBSCAN
from sklearn.preprocessing import StandardScaler
from tensorflow.keras.layers import Input, Dense
from tensorflow.keras.models import Model
from sklearn.metrics import calinski_harabasz_score
import seaborn as sns
from sklearn.neighbors import LocalOutlierFactor
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import confusion_matrix
import warnings
warnings.filterwarnings('ignore')
```

Fig 18: Import the libraries

In the above figure, I import the various libraries shown in the following figure. The required libraries for deep learning, machine learning, and data analysis are imported in this snippet of code. TensorFlow and Keras for deep learning, NumPy and Pandas for data processing, scikit-learn for machine learning models and metrics, and Matplotlib and Seaborn for data visualization are among the libraries. Furthermore, warnings are used to temporarily disregard the warning notifications. To silence any warning messages that may appear while the code is running, use filterwarnings('ignore'). When running the code, this can help prevent clutter in the output.

```python
#Read the dataset
df1 = pd.read_csv('database_gas.csv')
df2 = pd.read_csv('database_pos.csv')
```

Fig 19: Read the dataset

In the above figure, I have read the datasets 1 and 2. The CSV files "database_gas.csv" and "database_pos.csv" are used to import the data into df1 and df2, respectively. The data from the CSV files is read into a tabular format using the read_csv function from the Pandas library.

```python
# Merge the datasets on the nearest timestamps
merged_data = pd.merge_asof(df2.sort_values('datetime'), df1.sort_values('timestamp'),
                            left_on='datetime', right_on='timestamp', direction='nearest')
```

Fig 20: Merged the two dataset

The code is merging the two datasets (df2 and df1) in these two lines according to their timestamps. It combines the rows that match the closest timestamps found in both databases.

A new dataset named merged_data is the outcome. By aligning data from the two datasets according to their timestamps, this procedure makes combined analysis easier.

```
# Fill missing values in the merged data using backfill
merged_data.fillna(method='ffill', inplace=True)  # Forward fill
merged_data.fillna(method='bfill', inplace=True)  # Backward fill
```

Fig 21: Filling the missing value in merged data

The "forward fill" method (method='ffill') in these two lines fills missing values in the merged_data DataFrame by propagating the previous known value forward, while the "backward fill" method (method='bfill') fills empty entries by using the next available value. This makes it possible to guarantee that missing values are filled up using the dataset's most current information.

```
# 'merged_data' contains  dataset and 'anomaly' variable contains the output from Isolation Forest
X = merged_data[features]
anomaly = merged_data['anomaly']
```

Fig 22: Separating the dependent and independent variables

In the above figure, we are separating the dependent and independent variables, here target variable is anomaly.

```
# Standardize the features
scaler = StandardScaler()
X_std = scaler.fit_transform(X)
```

Fig 23: Normalize the dataset

Using the StandardScaler from scikit-learn, the features (independent variables) represented by the matrix X are standardized in these two to three lines. The features are scaled with a mean of 0 and a standard deviation of 1 using the fit_transform method, which first calculates the features mean and standard deviation. After being saved in the variable X_std, the standardized features are prepared for use in machine learning models.

4.5 Model Training:

The first stage in training a model is to split the dataset into training and testing sets so that the model's accuracy on unknown data can be assessed. Next, a machine learning model that takes into account the characteristics of the issue is selected, and an instance of the model is made with the required hyperparameters. In the training phase, the fit technique is used to enter the training data—which consists of features (independent variables) designated as X_train and their associated target values, y_train—into the model. The model discovers the underlying correlations and patterns in the data during this phase.

```
from sklearn.model_selection import train_test_split
X_train, X_test, true_labels_train, true_labels_test = train_test_split(X_scaled, true_labels, test_size=0.2, random_state=42)
```

Fig 24: Perform train and test split

Test_size=0.2 designates 20% of the data as allocated for testing, while random_state=42 uses a fixed random seed to guarantee split repeatability. This separation is essential to training the machine learning model on a fraction of the data and testing its performance on an additional, unseen component, which helps determine how well the model can generalize.

After dividing the data to ascertain how many rows are split between the training portion and the testing section, the print function was utilized to display the training and testing data. The number was obtained by using the String function. We are currently training several algorithms to become acquainted with our dataset and understand its operation after splitting the data. bringing in every algorithm from every Scikit-Learn library simultaneously. Let's look at how it was implemented.

```
#Fit the dbscan
dbscan = DBSCAN(eps=epsilon, min_samples=min_samples, metric='manhattan')
merged_data['cluster'] = dbscan.fit_predict(X_std)
```

Fig 25: Call and Fit the DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise):

The density of data points in the feature space is used by the density-based clustering method DBSCAN to determine the locations of clusters. It defines dense regions as clusters by aggregating nearby points that have a sufficient number of other points in their proximity. It can find clusters of different sizes and forms and is resistant to outliers. The approach may be applied to datasets with irregular structures since it does not need the definition of the number of clusters a priori. This is another advantage of the algorithm.

Autoencoder:

An artificial neural network type called an autoencoder is made for unsupervised learning. It learns to encode input data into a lower-dimensional representation and then recover the original input. It consists of an encoder and a decoder. When autoencoders are employed for feature learning and anomaly identification, data anomalies or outliers can be identified by departures from the taught patterns. Their effectiveness lies in their ability to capture non-linear relationships within data, making them a flexible method for representing complicated data.

Local Outlier Factor (LOF):

LOF evaluates the local density deviation of data points and is an unsupervised anomaly identification technique. It calculates the ratio between a data point's local density and the average density of its neighbors. Outliers are defined as points that have a density that is noticeably lower than that of their neighbors. In datasets with different densities, LOF works well for identifying outliers. Finding local anomalies is one of LOF's key advantages, which makes it appropriate for datasets including clusters of various densities and forms.

One-Class SVM (Support Vector Machine):

A machine learning approach called One-Class SVM is used to identify outliers and find novelty. Its goal is to build a hyperplane that encloses the normal data points. It is trained on a dataset that only contains normal instances. Instances outside of this hyperplane are regarded as outliers during testing. When working with unbalanced datasets—where normal occurrences greatly outnumber anomalies—it is very helpful. One of its benefits is that it can effectively handle high-dimensional data, which makes it useful in situations when there are a lot of characteristics.

Isolation Forest:

Isolation Forest is an anomaly detection algorithm designed to identify outliers in a dataset. It employs a tree-based approach, constructing isolation trees to isolate instances that are abnormal. Unlike traditional methods that rely on proximity measures, Isolation Forest leverages the concept that anomalies are less frequent and can be identified more easily. The algorithm works by randomly selecting features and partitioning the data until individual instances, or outliers, are isolated within a few partitions. Its efficiency lies in the ability to quickly identify anomalies, making it particularly effective for large datasets with complex structures, offering a robust solution for anomaly detection in various domains.

Chapter 5: Result and Model Evaluation

After using the anomaly as the target variable, I get the following accuracy:

Sr.no	Model	Accuracy
1	Autoencoder	1.00
2	Isolation Forest	1.00
3	Local Outliers Factor	0.90
4	One –class SVM	0.03

Table 2: Accuracy of the models

The accuracy scores in the table provide a comprehensive overview of the anomaly detection models' performance. The Autoencoder and Isolation Forest share the top position, each achieving a perfect accuracy score of 1.00. This signifies their exceptional ability to discern and accurately identify anomalies within the dataset, with the Autoencoder excelling in capturing nuanced patterns and the Isolation Forest leveraging a tree-based approach for robust anomaly detection.

The Local Outlier Factor (LOF) follows closely with a commendable accuracy of 0.90, showcasing its effectiveness in identifying outliers based on local density variations. The LOF's performance aligns well with its principle of assessing anomalies within their local neighborhoods, making it a reliable option for scenarios where deviations occur within specific clusters.

Conversely, the One-Class SVM lags behind with a lower accuracy score of 0.03, indicating limitations in pinpointing anomalies accurately. This suggests challenges in achieving a balance between true positives and false positives, pointing to potential difficulties in precisely delineating anomalous instances.

The Isolation Forest alongside the Autoencoder emphasizes its notable performance, reinforcing its efficacy in providing accurate anomaly detection within diverse datasets.

Precision, Recall and F1 score:

Precision: Precision measures how well a model predicts positive results. The ratio of true positive predictions to the total of true positives and false positives is used to compute it. The following is the formula:

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

A high precision means that there is little chance of false positives, meaning that the model's positive predictions are dependable.

Recall:

Recall measures how well a model can identify and include all pertinent positive examples. The ratio of true positives to the total of false negatives and true positives is used to compute it. The following is the formula:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

A high recall indicates that most positive cases are successfully identified by the model.

F1 Score: The harmonic mean of recall and precision is the F1 score. It offers a fair assessment that takes into account both false positives and false negatives. The following is the formula:

$$F1 = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$$

Since the F1 score integrates recall and accuracy into a single statistic to assess a model's overall performance, it is especially helpful in situations when the class distribution is unequal.

Sr. no	Algorithms	Precision	Recall	F1 score
1	Autoencoder	1	1	1
2	Isolation Forest	1	1	1
3	Local Outlier Factor	0.95	0.95	0.95
4	One-class SVM	0.17	0.17	0.18

Table 3: Precision, Recall and F1 score

When comparing anomaly detection algorithms, the Autoencoder and Isolation Forest both exhibit flawless performance, achieving perfect precision, recall, and F1 scores, each at 1. This indicates flawless anomaly identification, making these algorithms highly effective in discerning anomalies accurately. The Local Outlier Factor (LOF) performs robustly with scores of 0.95 for precision, recall, and F1, reflecting efficient anomaly detection and a well-balanced trade-off between false positives and false negatives. However, the One-Class SVM lags behind with significantly lower scores of 0.17 across precision, recall, and F1. This suggests challenges in accurately identifying positive instances and a higher likelihood of both false positives and false negatives. This analysis provides a comprehensive understanding of the algorithms' comparative effectiveness in anomaly detection without explicitly mentioning the recognition of Isolation Forest.

DBSCAN:

Silhouette Score: The degree to which a dataset's clusters are clearly characterized is measured by the Silhouette Score. Higher numbers denote better-defined clusters; the range is -1 to 1. For example, a high silhouette score (0.445) indicates that the clusters created by a clustering technique like DBSCAN are reasonably cohesive and separated.

Calinski-Harabasz Index: With a value of 163.09, the Calinski-Harabasz Index evaluates the quality of clustering by comparing the variance within and across clusters. More distinct and well-separated clusters are indicated by higher values of the Calinski-Harabasz Index, confirming the efficiency and caliber of the algorithm's clustering.

Chapter 6: Conclusion and Future Scope

In conclusion, the comprehensive evaluation of clustering algorithms and anomaly detection models has yielded valuable insights into their performances, shedding light on their strengths and limitations. The Autoencoder stands out as the top performer, consistently achieving a perfect score of 1 across accuracy, precision, recall, and F1, demonstrating its exceptional capability as an anomaly detector and establishing its reliability for precise anomaly identification.

The Isolation Forest also demonstrated outstanding performance, mirroring the Autoencoder with perfect scores across all metrics. This underscores its efficacy in accurately identifying anomalies and positions it as a robust choice for anomaly detection, particularly in scenarios where a tree-based approach is advantageous.

The Local Outlier Factor (LOF) showcased robust performance, achieving a high accuracy of 0.90 and strong precision, recall, and F1 scores at 0.95. Its ability to find outliers while maintaining a balanced prediction profile makes it valuable for anomaly detection, especially in situations where balancing false positives and false negatives is crucial.

Conversely, the One-Class SVM exhibited limitations with a lower accuracy of 0.03 and moderate precision, recall, and F1 scores of 0.17 and 0.18, respectively. While it may have specific use cases, its drawbacks highlight challenges in making accurate positive predictions, leading to increased rates of erroneous negative and false positive predictions.

The DBSCAN clustering method showcased respectable cluster separation and cohesiveness, as evidenced by the Silhouette Score of 0.445 and the Calinski-Harabasz Index of 163.09. These metrics indicate meaningful cluster creation, emphasizing DBSCAN's ability to identify well-defined clusters within the data, making it valuable for exploratory data analysis and grouping similar instances.

In choosing models for anomaly detection, considerations should align with specific use cases and the desired balance between recall and precision. Both the Autoencoder and Isolation Forest stand out as reliable options for precise anomaly identification, offering high accuracy and a balanced trade-off between false positives and false negatives. Meanwhile, the DBSCAN clustering method excels in creating meaningful clusters, providing insights into the underlying structure of the data.

In essence, the evaluation of these models has provided a nuanced understanding of their individual strengths and limitations, empowering practitioners to make informed decisions based on the specific requirements of their anomaly detection tasks. As the field of anomaly detection evolves, ongoing research and advancements in model development will likely contribute to refining and enhancing these models, expanding their applicability across diverse domains and use cases.

Limitation:

In the analysis of the smart elderly home monitoring system, it is crucial to acknowledge the limitations of the third model, One-Class SVM, which exhibited a notably low accuracy of 0.03. It is important to note that this low accuracy is attributed to the absence of model tuning, a factor that was not considered during the experimentation phase.

Moving forward, to address this limitation and improve the performance of the One-Class SVM, future work should focus on incorporating model tuning techniques. This may involve optimizing hyperparameters, fine-tuning parameters specific to the algorithm, or exploring advanced optimization methods. By addressing these aspects in future iterations of the study, there is potential for significant enhancement in the accuracy of the One-Class SVM, thereby strengthening its contribution to the integrated analysis of gas, temperature, and position sensors for anomaly detection and safety enhancement in smart elderly home monitoring. This recognition of limitations and a proactive approach toward improvement underscores the commitment to refining the model for more robust and reliable results in subsequent research phases.

Future Scope: In the future, regarding model interpretability, scalability, and adaptation to changing data patterns will determine the extent of anomaly detection. Decisions made by the model will be more transparent when integrated with explainable AI approaches. Furthermore, studies should concentrate on creating streaming and dynamic anomaly detection algorithms that can handle data in real time. To further improve anomaly detection capabilities, deep learning architectures and innovative unsupervised learning techniques may be applied, which can handle complicated, high-dimensional datasets. In order to face the changing difficulties presented by different and dynamic data environments, it will be imperative to continuously explore innovative algorithms and approaches. This will allow anomaly detection systems to offer more precise and timely insights across a range of areas.

Chapter 6: References

1. Ahn, C.R., Lee, S., Sun, C., Jebelli, H., Yang, K. and Choi, B., 2019. Wearable sensing technology applications in construction safety and health. *Journal of Construction Engineering and Management*, *145*(11), p.03119007.

2. Al-amri, R., Murugesan, R.K., Man, M., Abdulateef, A.F., Al-Sharafi, M.A. and Alkahtani, A.A., 2021. A review of machine learning and deep learning techniques for anomaly detection in IoT data. *Applied Sciences*, *11*(12), p.5320.

3. Ali, M., Ali, A.A., Taha, A.E., Dhaou, I.B. and Gia, T.N., 2019, May. Intelligent autonomous elderly patient home monitoring system. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)* (pp. 1-6). IEEE.

4. Bora, P., Kanakaraja, P., Chiranjeevi, B., Sai, M.J.S. and Jeswanth, A., 2021. Smart real-time health monitoring system using Arduino and Raspberry Pi. *Materials Today: Proceedings*, *46*, pp.3855-3859.

5. Cook, A.A., Mısırlı, G. and Fan, Z., 2019. Anomaly detection for IoT time-series data: A survey. *IEEE Internet of Things Journal*, *7*(7), pp.6481-6494.

6. DeFranco, J.F. and Kassab, M., 2021. Smart home research themes: An analysis and taxonomy. *Procedia Computer Science*, *185*, pp.91-100.

7. Del-Valle-Soto, C., Briseño, R.A., Valdivia, L.J., Velázquez, R. and Nolazco-Flores, J.A., 2023. Non-Invasive Monitoring of Vital Signs for the Elderly Using Low-Cost Wireless Sensor Networks: Exploring the Impact on Sleep and Home Security. *Future Internet*, *15*(9), p.287.

8. Erhan, L., Ndubuaku, M., Di Mauro, M., Song, W., Chen, M., Fortino, G., Bagdasar, O. and Liotta, A., 2021. Smart anomaly detection in sensor systems: A multi-perspective review. *Information Fusion*, *67*, pp.64-79.

9. Fadhil, J.A., Omar, O.A. and Sarhan, Q.I., 2020, April. A survey on the applications of smart home systems. In *2020 International Conference on Computer Science and Software Engineering (CSASE)* (pp. 168-173). IEEE.

10. Fahim, M. and Sillitti, A., 2019. Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review. *IEEE Access*, *7*, pp.81664-81681.

11. Grossi, G., Lanzarotti, R., Napoletano, P., Noceti, N. and Odone, F., 2020. Positive technology for elderly well-being: A review. *Pattern Recognition Letters*, *137*, pp.61-70.

12. Haghi, M., Spicher, N., Wang, J. and Deserno, T.M., 2022. Integrated sensing devices for disease prevention and health alerts in smart homes. In *Accident and Emergency Informatics* (pp. 39-61). IOS Press.

13. Imran, Iqbal, N., Ahmad, S. and Kim, D.H., 2021. Health monitoring system for elderly patients using intelligent task mapping mechanism in closed loop healthcare environment. *Symmetry*, *13*(2), p.357.\

14. Jabbar, W.A., Kian, T.K., Ramli, R.M., Zubir, S.N., Zamrizaman, N.S., Balfaqih, M., Shepelev, V. and Alharbi, S., 2019. Design and fabrication of smart home with Internet of Things enabled automation system. *IEEE Access*, *7*, pp.144059-144074.

15. Jacob Rodrigues, M., Postolache, O. and Cercas, F., 2020. Physiological and behaviour monitoring systems for intelligent healthcare environments: A review. *Sensors*, *20*(8), p.2186.

16. Javaid, M., Haleem, A., Rab, S., Singh, R.P. and Suman, R., 2021. Sensors for daily life: A review. *Sensors International*, *2*, p.100121.

17. Kuzubasoglu, B.A. and Bahadir, S.K., 2020. Flexible temperature sensors: A review. *Sensors and Actuators A: Physical*, *315*, p.112282.

18. Mansour, R.F., Escorcia-Gutierrez, J., Gamarra, M., Villanueva, J.A. and Leal, N., 2021. Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model. *Image and Vision Computing*, *112*, p.104229.

19. Marín, D., Llano-Viles, J., Haddi, Z., Perera-Lluna, A. and Fonollosa, J., 2023. Home monitoring for older singles: A gas sensor array system. *Sensors and Actuators B: Chemical*, p.134036.

20. Maswadi, K., Ghani, N.B.A. and Hamid, S.B., 2020. Systematic literature review of smart home monitoring technologies based on IoT for the elderly. *IEEE Access*, *8*, pp.92244-92261.

21. Mokhtari, S., Abbaspour, A., Yen, K.K. and Sargolzaei, A., 2021. A machine learning approach for anomaly detection in industrial control systems based on measurement data. *Electronics*, *10*(4), p.407.

22. Morita, P.P., Sahu, K.S. and Oetomo, A., 2023. Health Monitoring Using Smart Home Technologies: Scoping Review. *JMIR mHealth and uHealth*, *11*, p.e37347.

23. Nassif, A.B., Talib, M.A., Nasir, Q. and Dakalbab, F.M., 2021. Machine learning for anomaly detection: A systematic review. *Ieee Access*, *9*, pp.78658-78700.

24. Ni, F., Zhang, J. and Noori, M.N., 2020. Deep learning for data anomaly detection and data compression of a long-span suspension bridge. *Computer-Aided Civil and Infrastructure Engineering*, *35*(7), pp.685-700.

25. Omidvarborna, H., Kumar, P., Hayward, J., Gupta, M. and Nascimento, E.G.S., 2021. Low-cost air quality sensing towards smart homes. *Atmosphere*, *12*(4), p.453.

26. Palacín, J., Martínez, D., Clotet, E., Pallejà, T., Burgués, J., Fonollosa, J., Pardo, A. and Marco, S., 2019. Application of an array of metal-oxide-semiconductor gas sensors in an assistant personal robot for early gas leak detection. *Sensors*, *19*(9), p.1957.

27. Philip, N.Y., Rodrigues, J.J., Wang, H., Fong, S.J. and Chen, J., 2021. Internet of Things for in-home health monitoring systems: Current advances, challenges and future directions. *IEEE Journal on Selected Areas in Communications*, *39*(2), pp.300-310.

28. Seekaew, Y. and Wongchoosuk, C., 2019. A novel graphene-based electroluminescent gas sensor for carbon dioxide detection. *Applied Surface Science*, *479*, pp.525-531.

29. Shen, J., Won, J.Y., Chen, Z. and Chen, Q.A., 2020. Drift with the devil: Security of {Multi-Sensor} fusion-based localization in {High-Level} autonomous driving under {GPS} spoofing. In *29th USENIX Security Symposium (USENIX Security 20)* (pp. 931-948).

30. Sokullu, R., Akkaş, M.A. and Demir, E., 2020. IoT-supported smart homes for the elderly. *Internet of Things*, *11*, p.100239.

31. Taiwo, O. and Ezugwu, A.E., 2021. Internet of things-based intelligent smart home control system. *Security and Communication Networks*, *2021*, pp.1-17.

32. Taiwo, O., Ezugwu, A.E., Oyelade, O.N. and Almutairi, M.S., 2022. Enhanced intelligent smart home control and security system based on deep learning model. *Wireless communications and mobile computing*, *2022*, pp.1-22.

33. Talal, M., Zaidan, A.A., Zaidan, B.B., Albahri, A.S., Alamoodi, A.H., Albahri, O.S., Alsalem, M.A., Lim, C.K., Tan, K.L., Shir, W.L. and Mohammed, K.I., 2019. Smart home-based IoT for real-time and secure remote health monitoring of triage and priority system using body sensors: Multi-driven systematic review. *Journal of medical systems*, *43*, pp.1-34.

34. Tukur, Y.M., Thakker, D. and Awan, I.U., 2021. Edge-based blockchain-enabled anomaly detection for insider attack prevention in the Internet of Things. *Transactions on Emerging Telecommunications Technologies*, *32*(6), p.e4158.

35. Wang, J., Spicher, N., Warnecke, J.M., Haghi, M., Schwartze, J. and Deserno, T.M., 2021. Unobtrusive health monitoring in private spaces: The smart home. *Sensors*, *21*(3), p.864.