**USING MACHINE LEARNING TO PREDICT THE TOTAL ENERGY CONSUMPTION IN SMART HOMES BASED ON HISTORICAL DATA, WEATHER CONDITIONS AND APPLIANCE USAGE PATTERNS**

Sarah Higzi - UP2013822

School of Computing - Data Science and Analytics

Final Year Project - (PJS40 / PJE 2023/24)

Supervisor - Dr Farzad Arabikhan

Word Count - TOTAL (10,869)

May 2024

# ABSTRACT

Smart Homes. What are they? Why are they beneficial? Moreover, why have they become more prominent?

Smart homes have emerged as a critical solution to the challenges of modern-day living, from reducing household bills to limiting environmental impact. When used effectively, smart home devices empower consumers to have greater control over their daily energy usage, leading to significant energy savings and a more sustainable lifestyle.

This paper aims to answer the question: "Is it possible to use machine learning techniques to build a framework designed to forecast and optimize residential energy consumption by taking historical data, weather conditions and appliance usage patterns?" The primary objective of this study is to develop a model that can harnesses the potential of machine learning algorithms to maximize energy efficiency and management in residential homes. Utilizing tools like matplotlib and sci-kit-learn, we developed and tested several machine learning models, including regression and neural networks to target and anticipate energy use based on daily loads and hourly intervals.

The adopted method involved collecting an integrated dataset that included information on weather conditions and previous appliance consumptions provided by smart homes. This data was then split into training and testing sets. The training set was used to teach the machine learning models how to make predictions, while the testing set was used to evaluate the models' performance.

The findings of this study hold significant promise. They demonstrate that machine learning technologies and algorithms can enhance and provide more accurate predictions. However, it's important to note that the accuracy of these predictions may vary depending on the quality and quantity of the data available. Thus, enabling precise forecasting in future investigations of power consumption in household units. This could revolutionize how we manage energy resources, leading to substantial cost savings and a more sustainable future.

**TABLE OF CONTENTS**

[ABSTRACT 2](#_Toc167020825)

[LIST OF TABLES 6](#_Toc167020826)

[LIST OF FIGURES 7](#_Toc167020827)

[LIST OF SNIPPETS 8](#_Toc167020828)

[CHAPTER 1: INTRODUCTION 9](#_Toc167020829)

[1.1 Research Context and Significance 10](#_Toc167020830)

[1.2 Importance of the Study 10](#_Toc167020831)

[1.3 Aim 11](#_Toc167020832)

[1.4 Objective 11](#_Toc167020833)

[1.5 Constraints 11](#_Toc167020834)

[1.6 Log of risks. 13](#_Toc167020835)

[1.7 Legal, ethical, professional, and social issues. 13](#_Toc167020836)

[CHAPTER 2: LITERATURE REVIEW 14](#_Toc167020837)

[2.1 Introduction 14](#_Toc167020838)

[2.2 Background on Energy Consumption in Smart Homes 14](#_Toc167020839)

[2.2.1 Historical Context 14](#_Toc167020840)

[2.2.2 Evolution of Smart Homes 15](#_Toc167020841)

[2.2.3 Role of Energy Management 18](#_Toc167020842)

[2.3 Importance of Machine Learning in Energy Prediction 20](#_Toc167020843)

[2.3.1 Technological Advances in Machine Learning 20](#_Toc167020844)

[2.3.2 Impact of Machine Learning on Energy Efficiency 21](#_Toc167020845)

[2.4 Methodological Approaches in Literature 21](#_Toc167020846)

[2.4.1 Overview of Theoretical Models Used 22](#_Toc167020847)

[2.4.2 Application of Predictive Analytics Theories 22](#_Toc167020848)

[2.4.3 Relationship Between Theoretical Assumptions and Practical Outcomes 22](#_Toc167020849)

[2.5 Summary and Conclusions 23](#_Toc167020850)

[2.5.1 Conclusions Drawn from the Literature Review 23](#_Toc167020851)

[2.5.2 Predictions for Future Work 23](#_Toc167020852)

[CHAPTER 3: METHODOLOGY 24](#_Toc167020853)

[3.1 Introduction 24](#_Toc167020854)

[3.2 Research question 24](#_Toc167020855)

[3.3 Research method 25](#_Toc167020856)

[3.3.1 Sequential Steps 25](#_Toc167020857)

[3.3 Research design 27](#_Toc167020858)

[3.4 Data collection method 27](#_Toc167020859)

[3.4.1 User feedback on Dataset 28](#_Toc167020860)

[3.4.2 Descriptive feedback on Dataset 28](#_Toc167020861)

[3.5 Data analysis 29](#_Toc167020862)

[3.6 Ethical considerations 30](#_Toc167020863)

[3.7 Limitations, Validity and Reliability 30](#_Toc167020864)

[3.8 Quality Assurance 31](#_Toc167020865)

[3.9 Impact Assessment 31](#_Toc167020866)

[3.10 Summary 32](#_Toc167020867)

[CHAPTER 4: REQUIRMENTS 32](#_Toc167020868)

[4.1 Requirements Gathering 32](#_Toc167020869)

[4.2 Requirements Specification 33](#_Toc167020870)

[4.3 Predictors and Outcomes 34](#_Toc167020871)

[CHAPTER 5: DATA 34](#_Toc167020872)

[5.1 Dataset 34](#_Toc167020873)

[5.2 Data Preparation 35](#_Toc167020874)

[5.2.1 Import Libraries 35](#_Toc167020875)

[5.2.2 Load the Dataset 35](#_Toc167020876)

[5.3 Handling Missing Values 35](#_Toc167020877)

[5.4 Detecting and Handling Outliers 36](#_Toc167020878)

[5.5 Dataset Reduction 38](#_Toc167020879)

[5.6 Data Conversion 39](#_Toc167020880)

[5.6.1 Timestamp 39](#_Toc167020881)

[5.6.2 Categorical Features 39](#_Toc167020882)

[5.7 Data Preprocessing 40](#_Toc167020883)

[5.8 EDA 41](#_Toc167020884)

[5.8.1 - Columns 42](#_Toc167020885)

[5.8.2 Handling of Time Variable 42](#_Toc167020886)

[5.8.3 Distribution of Individual Variables 43](#_Toc167020887)

[5.8.4 Relationships between Variables 46](#_Toc167020888)

[5.8.5 Distribution of house and weather appliances 49](#_Toc167020889)

[5.8.6 Time Series Analysis 52](#_Toc167020890)

[5.8.7 Forecasting 56](#_Toc167020891)

[5.8 ​​Model Development 57](#_Toc167020892)

[5.8.1 Regression Model Metrics 57](#_Toc167020893)

[5.8.2 Neural Network Architecture 59](#_Toc167020894)

[5.9 Model Optimization and Validation 61](#_Toc167020895)

[5.10 Interpretation and Recommendations 62](#_Toc167020896)

[5.10.1 Train Model 62](#_Toc167020897)

[5.11 Discussion and Analysis of Findings 64](#_Toc167020898)

[CHAPTER 6: REQUIREMENT EVALUATION 64](#_Toc167020899)

[6.1 - Summary of Requirements 64](#_Toc167020900)

[6.2 Analysis of Requirement Fulfillment 65](#_Toc167020901)

[6.3 Recommendations for Future Work 65](#_Toc167020902)

[6.4 Conclusion 66](#_Toc167020903)

[CHAPTER 7: CONCLUSION 66](#_Toc167020904)

[7.1 Research Aims, Objectives, and Questions 66](#_Toc167020905)

[7.2 Key Findings 66](#_Toc167020906)

[7.3 Relation to Research Aims and Objectives 67](#_Toc167020907)

[7.4 Limitations of the Study 67](#_Toc167020908)

[7.5 Suggestions for Improvement 67](#_Toc167020909)

[7.6 Implications and Recommendations 67](#_Toc167020910)

[7.6.1 Practical Applications: 67](#_Toc167020911)

[7.6.2 Future Research Directions: 68](#_Toc167020912)

[REFERENCES 69](#_Toc167020913)

# LIST OF TABLES

[Table 1 13](#_Toc167020914)

[Table 2 14](#_Toc167020915)

[Table 3 34](#_Toc167020916)

[Table 4 34](#_Toc167020917)

[Table 5 36](#_Toc167020918)

[Table 6 42](#_Toc167020919)

[Table 7 65](#_Toc167020920)

# LIST OF FIGURES

[Figure 1 37](#_Toc167020921)

[Figure 2 38](#_Toc167020922)

[Figure 3 39](#_Toc167020923)

[Figure 4 39](#_Toc167020924)

[Figure 5 42](#_Toc167020925)

[Figure 6 43](#_Toc167020926)

[Figure 7 43](#_Toc167020927)

[Figure 8 43](#_Toc167020928)

[Figure 9 43](#_Toc167020929)

[Figure 10 44](#_Toc167020930)

[Figure 11 44](#_Toc167020931)

[Figure 12 44](#_Toc167020932)

[Figure 13 45](#_Toc167020933)

[Figure 14 45](#_Toc167020934)

[Figure 15 45](#_Toc167020935)

[Figure 16 46](#_Toc167020936)

[Figure 17 46](#_Toc167020937)

[Figure 18 47](#_Toc167020938)

[Figure 19 47](#_Toc167020939)

[Figure 20 48](#_Toc167020940)

[Figure 21 49](#_Toc167020941)

[Figure 22 49](#_Toc167020942)

[Figure 23 50](#_Toc167020943)

[Figure 24 51](#_Toc167020944)

[Figure 25 53](#_Toc167020945)

[Figure 26 53](#_Toc167020946)

[Figure 27 54](#_Toc167020947)

[Figure 28 55](#_Toc167020948)

[Figure 29 55](#_Toc167020949)

[Figure 30 56](#_Toc167020950)

[Figure 31 57](#_Toc167020951)

[Figure 32 57](#_Toc167020952)

[Figure 33 60](#_Toc167020953)

[Figure 34 60](#_Toc167020954)

[Figure 35 60](#_Toc167020955)

[Figure 36 61](#_Toc167020956)

[Figure 37 63](#_Toc167020957)

[Figure 38 63](#_Toc167020958)

# LIST OF SNIPPETS

[Equation 1 35](#_Toc167020959)

[Equation 2 36](#_Toc167020960)

[Equation 3 36](#_Toc167020961)

[Equation 4 36](#_Toc167020962)

[Equation 5 37](#_Toc167020963)

[Equation 6 38](#_Toc167020964)

[Equation 7 38](#_Toc167020965)

[Equation 8 38](#_Toc167020966)

[Equation 9 39](#_Toc167020967)

[Equation 10 39](#_Toc167020968)

[Equation 11 40](#_Toc167020969)

[Equation 12 40](#_Toc167020970)

[Equation 13 40](#_Toc167020971)

[Equation 14 40](#_Toc167020972)

[Equation 15 41](#_Toc167020973)

[Equation 16 41](#_Toc167020974)

[Equation 17 42](#_Toc167020975)

[Equation 18 42](#_Toc167020976)

[Equation 19 47](#_Toc167020977)

[Equation 20 60](#_Toc167020978)

[Equation 21 60](#_Toc167020979)

[Equation 22 61](#_Toc167020980)

[Equation 23 61](#_Toc167020981)

[Equation 24 62](#_Toc167020982)

[Equation 25 62](#_Toc167020983)

[Equation 26 62](#_Toc167020984)

[Equation 27 62](#_Toc167020985)

[Equation 28 63](#_Toc167020986)

# CHAPTER 1: INTRODUCTION

The central goal of this paper is to highlight the energy waste prevalent in modern households. Irrespective of the widespread use of smart appliances and devices, many households do not have a formalised plan of action to control energy consumption, which translates into abnormally high energy bills, a more significant environmental footprint and possibly excess wear and tear of the household equipment. This project's objective is to optimize energy usage by refining models based on data from smart home devices through machine learning and data analysis. The project also aligns with the evolution of smart homes and technology conservation by using data science to improve energy efficiency, addressing a problem that all contemporary homes attempt to solve.

## 1.1 Research Context and Significance

Smart homes are those that have several sophisticated automated systems installed, enabling technology-based central control over the entire house. These are the kinds of homes that include a network of linked appliances, often managed by a centralized location like a control system or an app on a smartphone. These devices may be linked to each another over the Internet of Things and can be managed, controlled, or seen from anywhere.

## 1.2 Importance of the Study

Smart homes can deal with problems that people face daily, for instance:

* **Energy Efficiency**: Smart homes make the heating, cooling, and electrical systems more adaptable. Cutting unnecessary energy waste, which then leads to a lower utility bill and inevitably reduce environmental impact.
* **Enhanced Comfort and Convenience:** The automation of tedious tasks and the option to control the home environment from a distance enhances the quality of life. Providing an extra level of accessibility and accommodating the people who live in the house at a better level.
* **Improved Security:** The strict nature of security systems that smart homes can provide give the homeowners a relief, as they can safety check their houses remotely and be notified if there is a security breach.
* **Adaptability to Modern Lifestyles:** The societal alterations in work and personal life dynamics have been encouraged by smart homes. The flexibility of technological integration to everyday living has met modern lifestyles, particularly for those who work from home.

## 1.3 Aim

The project aims to develop a data-driven machine learning model for optimizing energy consumption in smart homes, thereby promoting energy efficiency, cost savings, and environmental sustainability.

In addressing this primary question, the investigation could further investigate comparable sub-questions like:

* **Influence of Weather:** What is the effect of various weather conditions (heat, humidity, visibility, etc.) on the overall amount of energy consumed?
* **Usage Trends for Appliances**: Which appliances are the most energy-consuming, and how may they be used more efficiently?
* **Time-Series Examination:** Is it possible to spot daily or seasonal differences in energy usage over time, as well as trends and patterns?

## 1.4 Objective

1. Data Collection and Integration

2. Data Preprocessing and Cleaning

3. Exploratory Data Analysis (EDA)

4. Feature Engineering

5. Model Selection and Training

6. Model Evaluation and Optimization

7. Interpretation of Model Results and Feature Importance

8. Development of Energy Optimization Recommendations

9. Machine Learning Model Development

10. Optimization Strategy Formulation

11. Model Evaluation and Validation

12. Project Documentation and Reporting

## 

## 1.5 Constraints

Computing/IT Resources:

* *Personal Laptops*: My laptop is supposed to have the software and computational capacity required to undertake data analysis and simulations.
* *University PCs:* Opening hours and term time constraints may affect availability.
* *Software*: Programming languages (e.g., Python), data analysis software (e.g., Jupyter Notebook, RStudio, IntelliJ), and simulation tools (e.g., GitHub, Moodle) will be required. Licensing and access may limit software availability.
* *Data Storage*: Enough space to store and manage datasets, analytical results, and project files.

Data Resources:

* *Energy Consumption Data*: Historical energy consumption data from smart homes, including time-stamped readings from various devices and appliances.
* *Weather Data*: Information about the weather relevant to the smart home context, such as temperature, humidity, visibility, and wind speed.

Library Resources:

* *University Library:* For the literature review and preliminary study, access to academic publications, papers on research, and reference materials will be necessary. Library hours and resource availability may be limited.

Facilities/Environments:

* *Access to Data*: Accessibility to the relevant datasets may be limited due to data availability limits and privacy concerns.
* *Internet Access*: A dependable internet connection for data access, research, and discussion.

Specialized Hardware (if applicable):

* Access to and availability of such hardware may present constraints if the project calls for the use of hardware components for testing or data collection.

Human resources (if applicable):

* Project Supervisor: The project's development may be driven by the project supervisor’s availability and interest in the topic.

## 

## 1.6 Log of risks.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***No*** | ***Description*** | ***Likelihood***  ***(high, medium, low)*** | ***Impact*** | ***Mitigation/Avoidance*** |
| 1 | Data Source Unavailability | Medium | Could hinder data analysis and experimentation | Identify alternative data sources or collect backup data sets. |
| 2 | Software Compatibility Issues | Medium | May delay coding and analysis | Verify software compatibility early; seek alternatives if needed. |
| 3 | Limited Access to University PCs | Medium | May affect resource availability | Schedule PC usage during non-peak hours; plan for laptop use. |
| 4 | Participant Non-Cooperation | Low | Could disrupt collaboration (if applicable) | Establish clear communication and expectations with participants. |
| 5 | Hardware Failure (if applicable) | Low | May require repairs or replacements | Regularly maintain and monitor hardware; have backup components. |
| 6 | Data Privacy and Ethical Issues | Medium | May lead to data access restrictions | Ensure compliance with data privacy regulations; seek approvals. |
| 7 | Unexpected Technical Challenges | Medium | Could lead to delays in project tasks | Allocate extra time for troubleshooting; seek guidance when needed. |

Table

## 1.7 Legal, ethical, professional, and social issues.

|  |  |  |  |
| --- | --- | --- | --- |
| ***Legal*** | ***Ethical*** | ***Professional*** | ***Social*** |
| Data privacy – adhere to the General Data Protection Regulation (GDPR) in Europe | Informed consent for data usage | Professional Conduct: - transparency in research and reporting | Social and environmental responsibility |
| Intellectual Property and licensing agreements | Bias and Fairness - address any bias in the models |  |  |

Table

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

This chapter reviews several research and theoretical papers about the application of machine learning techniques to aid in predicting energy consumption in smart homes. The literature review covers studies that have incorporated historical data, climate variables, and fluctuations in appliance usage in models. It also explores the qualities of various machine learning models in improving energy predictability, such as decision trees, neural networks, and ensemble methods.

## 

## 2.2 Background on Energy Consumption in Smart Homes

### 2.2.1 Historical Context

Historically, energy consumption in residential buildings has become a significant concern, reflecting broader problems related to environmental impact and resource depletion. The evaluation of historical data on energy use in buildings, particularly the HVAC (heating, ventilation, and air conditioning) systems, reveals a surge in consumption rates because of the rising demand for indoor comfort levels and the increasing amount of time spent indoors. This trend has been observed since the 20th century, with energy consumption in residential buildings steadily increasing over the years (Pérez-Lombard et al., 2008).

Buildings play a significant role in exacerbating climate change, accounting for one-third of global energy consumption and one-quarter of CO2 emissions (González-Torres et al., 2022). Since the fifties, the global contribution from buildings towards energy usage has notably accelerated, accounting for between 20% and 40% in developed countries. This upward trend in consumption rates is likely to continue, pushed by population growth, comfort demands, and overall changes in building use and occupancy patterns. (Pérez-Lombard et al., 2008).

While the growing worldwide energy consumption and mounting CO2 emissions clearly require a response (Jackson et al., 2018). The advantages of globalization, rising affluence, and population growth have all contributed to this trend. Despite the pressing need to address climate change, global energy usage and the corresponding carbon dioxide emissions; the constant quest for more extensive and visually appealing buildings has resulted in a notable rise in energy usage.

Unlike the conventional residential buildings, smart homes provide a good way for energy conservation using the advanced technologies and the Internet of Things (IoT) devices. In the environment of smart homes, IoT is the term that describes the interconnected gadgets and systems that communicate locally and over the internet to manage the various household tasks automatically or remotely. These gadgets can either process, record, transmit, or do something with the data they get from their surroundings using the embedded sensors, software, or other techniques.

Smart homes employ data-based methods and automated systems to enhance energy use, which usually results in a significant decrease in energy consumption. Harris et al. (2019) have discovered that energy consumption can be cut by 25% through the adoption of smart home features such as the automation of HVAC systems, switching off the lights, and other high-energy appliances. These systems can alter the energy consumption depending on real-time data like occupancy pattern and the weather conditions outside the building, thus improving energy efficiency (Smith, 2020). Besides, the installation of the smart meters and the energy management systems in the homes, it has also been proven to be helpful for homeowners to receive the opportunity to change their energy habits, actively saving money (Brown & Green, 2021). Therefore, the smart home market has been expanding quickly, and according to the research of Johnson et al. (2022), 63% of American homes will be classified as 'smart' by 2025, which leads to combat energy efficiency and sustainability concerns.

### 2.2.2 Evolution of Smart Homes

Technology has greatly accelerated the evolution of smart homes. Today, the introduction of IoT and big data has significantly transformed the capabilities of these smart homes, enabling them to quickly adapt to users' behaviors and environmental changes.

Integration of IoT and Big Data

According to Rathore et al. (2016), the integration of IoT and big data in the planning of urban and smart homes brings forth data-driven management coupled with operational efficiency. These are the main characteristics of the intelligent configurations found in contemporary homes, instilling confidence in the effectiveness of these technologies.

Technological innovations in smart homes can consist of:

1. **Voice-Activated Assistants**: The gadgets such like Amazon Alexa and Google Home have totally redefined the way we deal with our homes by which we can control the lighting, temperature, media, and security systems by our voice.
2. **Automated HVAC Systems:** Nowadays, HVAC systems can adapt to the users' habits and the weather forecasts. Accordingly, they can automatically adjust the indoor temperatures for the best user comfort, significantly reducing unnecessary heating and cooling.
3. **Smart Security Systems:** With the introduction of the facial recognition and motion sensors systems, the owners can get the alerts and the video feeds to their smartphones, this new high-tech surveillance and remote monitoring capabilities help with security enhancement.

Technology changes, consumer demands, and the architecture and design sector have led to the rejection traditional homes. By applying these technological advances to smart homes we can allow for innovative adaptability, allowing devices to anticipate and fulfil the residents' needs.

Role of Big Data Analytics

The introduction of big data analytics is a major step forward in the global acceptance of intelligent systems. Through the analysis of the vast amounts of statistical data that are gathered from smart devices, big data analytics enables for the better, more personalized home management system. Rathore et al. (2016) stresses that this development is a big move towards enhancing energy consumption and management worldwide.

Building Management Systems (BMS) and Energy Management Systems (EMS) serve as exemplary frameworks to further understand the practical application of IoT and big data in enhancing residential living. Such systems are the perfect examples of how the latest technologies can be merged to make the best of the situation and thus you can get the maximum quantity of work done.

Building Management Systems (BMS)

BMS are key to the development of smart homes, by combining the essential systems like HVAC, lighting, and security, energy usage decreases. The Chartered Institute for Building Services Engineers (CIBSE) stresses the importance of BMS in the improvement of operations and the decrease of environmental damages.

Examples of BMS include:

1. **Honeywell Building Manager:** The BMS system that is integrated helps to control and monitor different systems hence improving the efficiency of a building.
2. **Siemens Desigo CC:** This system is responsible for conducting all outputs related to building comfort, safety, and efficiency operations from one platform.
3. **Schneider Electric’s EcoStruxure Building Operation:** This system makes it easy to manage outputs of energy, HVAC, lighting, and fire safety.

Energy Management Systems (EMS)

Energy Management Systems are the systems that are designed to manage the energy usage in a building to reduce both the consumption and the cost. Usually, they are used to check, keep track of, and regulate the electrical building loads.

Examples of EMS include:

1. **Johnson Controls Metasys® System:** Ensures the overall management of buildings are correct through detailed monitoring and control of building management systems and equipment.
2. **GridPoint Energy Manager:** This platform is compatible with the existing infrastructure. It delivers information on energy consumption in real-time and offers automated controls.
3. **Enel X JuiceBox:** An energy storage system that uses EMS capabilities to reduce energy costs.

Code of Practice for Smart Buildings

The Code of Practice for Smart Buildings is a set of norms that ensures the automation of different functions in buildings. It is designed to guide the planning, design, and integration of various building systems to achieve greater efficiency and sustainability, thus improving the quality of life of residents.

Example guidelines include:

1. **CIBSE Guide M:** Offers the most appropriate strategies for the preservation of the function and effectiveness of building services.
2. **BSRIA BG 9/2011:** A guide that gives the best methods for managing and executing building services to attain better energy efficiency.
3. **ISO 50001:** The global standard, thus, serves as a guide for the establishment, execution, maintenance, and enhancement of an energy management system.

Smart houses adapt to the behaviors of its occupants in a proactive rather than reactive manner. The specification of various BMS-controlled house systems, such the lighting and HVAC, as well as the EMS's energy-use optimization, are accompanied by the integration of external environmental conditions. The previously mentioned technological advancements could potentially be directly applied to the practical and sustainable control of smart home energy systems.

### 2.2.3 Role of Energy Management

Enhancing energy management is one of the biggest challenges in optimizing energy consumption levels in smart homes. By focusing on algorithms and real-time data techniques, we can discover trends and variables that impact data consumption.

Here are concise examples of how these technologies are applied:

* **Predictive Analytics:**

*Smart Thermostats:* The devices like ‘Nest’ for instance, employ predictive analytics to set heating and cooling systems based on the learned schedules and real-time weather information.

* **Machine Learning Models:**

*Energy Disaggregation:* ‘Sense’ is a device that dissects the energy consumption and tells homeowners how much each appliance uses, thus helping them to identify and reduce the wasteful usage.

* **Real-Time Monitoring and Control:**

*Dynamic Solar Systems*: Smart solar panels and batteries, which utilize real-time data and consumption patterns, make the decision whether to store, use or sell the energy.

* **IoT-Driven Automation:**

*Automated Lighting and Window Shades:* Systems adapt to the changing conditions through the light and occupancy sensors, thus regulating the light use and lowering the energy costs.

* **Behavioral Feedback Algorithms:**

*User Engagement Platforms*: Through applications associated with the smart meters, real-time feedback and energy-saving tips are given, thus the people are encouraged to adopt efficient behaviors.

By integrating Home Energy Management Systems (HEMS), the Internet of Things (IoT), and big data we can address upcoming problems. Through the prediction of environmental effects and maintenance of operating expenses, the flows of energy must be monitored, predicted, and controlled dynamically (María-Cano et al., 2020; Al-Ali et al., 2017).

Hannarvar and Sami (2018) and Sayah et al. (2020) revealed that to further improve energy management, such approaches can extract and analyze fundamental household energy consumption patterns by integrating advanced data analytics and machine learning. This transition of predictive approaches in energy plans clearly indicates the importance of big data in converting conventional energy plans into more sustainable and effective operations (Singh & Yassine, 2018).

The adoption of IoT technologies alongside big data analytics has begun challenging the efficiency of energy management systems, resulting in unexpected development of resource management and energy efficiency in smart homes. (Ravisagar & Srinivas, 2018; Singh & Yassine, 2018; Al-Ali et al., 2017). This effect is further enhanced by statistics revealing that the smart home market is projected to achieve $135.3 billion by 2025, growing at a compound annual growth rate (CAGR) of 11.6% from 2020. Based on recent calculations a quick development is also evident in the energy sector, where it is claimed that the rise in smart appliances and HEMS reduce household energy consumption by up to 35%.

These trends reflect a higher degree of implementation and yielding of tangible economic and environmental effects in households with advanced energy management systems. This evolution shows a change in the organization of home energy usage, moving away from conventional systems and towards more technologically advanced alternatives.

## 2.3 Importance of Machine Learning in Energy Prediction

The use and creation of computer systems that can learn and adapt without explicit instructions is known as machine learning. This is achieved by utilizing statistical models and algorithms to examine data patterns and derive conclusions from them.

### 2.3.1 Technological Advances in Machine Learning

The paradigm of machine learning is changing due to the most recent major evolution that impacts various sectors, driving more precise and sophisticated predictions. While the past decade has seen machine learning technologies, including the deep learning leveraging complex architectures like convolutional neural networks (CNNs) and recurrent neural networks (RNNs), advancements continue to be made. Machine learning methods experienced significant improvements especially in the way that they deal with unstructured large data sets (LeCun et al., 2015).

Technological advancements in the sphere of machine learning have created cutting-edge approaches to energy management, where the work of these technologies is to maximize usage effectively and minimize wastage. To put it simply, machine learning is the model for development where the speed of algorithm and the power of computers increases, hence, affecting many sectors.

According to current research, Big Data has boosted machine learning, allowing computers to recognize complicated patterns and make logical decisions much faster and more accurately. Providing us with an option to process and forecast whole datasets in real-time, thus substantially impacting different economic sectors that depend primarily on data analysis (Zhou et al., 2016).

### 2.3.2 Impact of Machine Learning on Energy Efficiency

Machine learning has largely influenced energy efficiency by creating and applying models that, with extreme precision, estimate and supervise the energy consumption of buildings. Studies have demonstrated that machine learning algorithms such as ANN and SVM, can enhance the precision of energy forecasts by up to 20% compared to the more traditional statistical methods (Ahmad et al., 2014). These models are particularly good at handling the complex and non-linear data structures predominant in energy systems leading to the development of insights and efficient energy management decisions.

Moreover, machine learning has had tremendous effects on energy systems in improving operation management. For example, it has been demonstrated that mixed models, which include several machine learning techniques, improve the algorithm's prediction accuracy by 10-15%, which reduces energy waste and optimizes its usage (Zhao, 2012). Additional work by Ahmad et al. (2014) showcased that machine learning is becoming increasingly important for the optimization of energy consumption by predicting peak load times and proposing suitable load-shedding measures in response.

## 2.4 Methodological Approaches in Literature

This study is conceptually based on predictive analytics and machine learning theory. The machine learning models used includes regression models, tree-based models, ensemble methods, and deep learning techniques. By focusing on how these technologies may be used to consistently calculate and predict energy use in household operations, we are able to create theoretical foundations that enables the development of models that can dynamically adapt to changing events and user behaviors.

#### 

### 2.4.1 Overview of Theoretical Models Used

There are several theoretical models that are key to predictive analytics and machine learning. A few among them are, regression analysis, decision tree models, support vector machines (SVM) and neural networks. Different models feature practical possibilities in terms of processing and analyzing large datasets with complicated connections. One example would be the regression models that handle continuous data perfectly, primarily when predicting energy usage depending on factors such as temperature and time. Neural networks, in general, especially deep learning models, perform pattern recognition in a significant way. They can adapt quickly to different situations but do not need to be reprogrammed on new information.

### 2.4.2 Application of Predictive Analytics Theories

Logic-based modelling and machine learning algorithms are combined to create predictive models and analytics. These models analyze the patterns to find relationships between given data. In the scope of this study, predictive analytics are exploited to anticipate the household energy consumption using the data of the weather conditions as well as the appliance usage patterns and the past metrics of the energy consumption. It makes use of complex algorithms that provide exact and useful forecasting charts which can assist in reducing costs and enhancing the overall energy use.

### 2.4.3 Relationship Between Theoretical Assumptions and Practical Outcomes

The test practicality of the predictive analytics and the machine learning models is primarily based on the assumption that the theoretical frameworks behind them are correct. These assumptions are the linearity, normality, homoscedasticity, and independence of errors. In real-world implementations, the conditions that these assumptions satisfy commonly do not hold true, which can cause discrepancies in model accuracy and dependability. A critical understanding of any such theoretical hypotheses and practical outcomes is key to working out how to improve the models and be attentive to the practical outcomes.

## 2.5 Summary and Conclusions

The sections in the report above turns attention to the fact that integrating big data analytics and (IoT) technologies have changed how smart homes consume and conserve energy, leading to more customized and effective strategies for energy management. Different machine learning techniques such as decision trees, neural networks, and ensemble approaches can been implemented to deal with the complexities associated with energy forecasting and optimization, proving necessary in this constantly evolving sector.

### 2.5.1 Conclusions Drawn from the Literature Review

The research uncovers the ability of machine learning to be effectively incorporated to improve energy efficiency in smart homes. Such approaches highlight that current technologies may output desirable outcomes, but the underlying strategies should be improved to integrate up-to-date and adapt to dynamic environmental conditions seamlessly, supporting the global drive towards environmentally friendly energy resources.

The review findings bring to light the important gaps which need further research. These gaps include data quality improvement, model overfitting mitigation and renewable sources incorporation in predictive frameworks. Another focus is on elaborating the models which are capable not only of making reliable energy consumption forecasts but also can be used to incorporate data around user behaviors and external environmental factors.

### 

### 2.5.2 Predictions for Future Work

By integrating hybrid models, we can combine the strengths of different machine learning approaches, consequently increasing accuracy in terms of prediction and adaptability capabilities in smart home energy management systems. The incorporation of real-time data with the Internet of Things can allow for more adaptable efficient energy systems. Unsupervised learning algorithms are another important area of application which can be utilized to enhance energy predictions while consuming minimal human resources.

As machine learning and data analysis gain momentum, we acknowledge the coming of impactful, user-centered energy solutions capable of diminishing energy demand. The above conclusions and projections demonstrate that smart homes have an enormous impact on the energy sector, especially when we are currently striving towards a more automated sustainable way of living.

# CHAPTER 3: METHODOLOGY

This dissertation's main claim is that by using feature selection approaches, and machine learning models we will be able to forecast a smart home's overall energy consumption based on past data, current weather, and the usage behaviors of specific appliances. This section is intended to provide and explore detailed theoretical assumptions based around the research question. Can we use and optimize machine learning models to predict total energy consumption in smart homes, by using the relevant data science methods to improve efficiency.

## 3.1 Introduction

*Hypothesis*: Machine learning techniques prove to be efficient in making energy management in smart homes more economical, thus enabling a more sustainable form of living environment.

This research project is quantitatively oriented. Its primary concern is to conduct a comprehensive statistical analysis of the discrete data on energy usage in smart homes. Precise predictive modelling becomes possible due to the datasets extensive supply of time-series data, historical energy consumption, current weather data, and appliance consumption recordings. By applying machine learning algorithms in developing models, we can forecast the amount of energy consumed in smart homes. Allowing us to determine which variables have a more significant impact on overall energy consumption,

## 3.2 Research question

The central research question of this dissertation is: "Using machine learning models to predict the total energy consumption in smart homes based on historical data, weather conditions, and appliance usage patterns?"

As a student of data science and having realized that the application of machine learning gives a deeper understanding of complex data. It is the practical knowledge gained through the course that is being applied. The project adopted feature engineering, predictive modeling, time series analysis, regression analysis, and neural networks. Using data to make accurate and effective decisions has a significant impact on the broader landscape of data science and analytics. Showcasing that our theoretical academic studies can be applied to solve real-world problems.

As the number of energy consumption issues develop internationally the need for sustainable energy products is made apparent. The issue being considered addresses a significant societal challenge, through the growing concern for mandatory efficiency in smart houses.

The demonstration of how machine learning reforms energy utilization in smart homes would mark the onset of innovation in the realm of smart technology and the IoT (Internet of Things). This area of work continually enhances performance and efficiency by meticulously analyzing data. It addresses not only issues about home improvement but also lays out a framework for implementing optimization methods in smart homes.

## 3.3 Research method

The variety of quantitative methods are chosen due to their unique capacity to produce precise, objective, and generalizable findings. The additional utilization of statistical approaches and algorithms will make it accessible to assess relationships between variables and to forecast energy usage from data history.

### 

### 3.3.1 Sequential Steps

***Problem Definition:*** The first thing that I will do is to clearly identify the problem. Using machine learning techniques for optimizing energy consumption in smart homes. This includes pinpointing certain aspects of energy consumption that the research is going to be concentrated on, in this case the role of weather conditions and of household appliances.

***Data Collection:*** Through utilization of a publicly available dataset on Kaggle entitled "Smart Home Dataset with Weather Information". The dataset incorporates electricity consumption data in minute-by-minute scale along with weather conditions and household appliances.

***Data Preprocessing:*** As a starting point, the data goes through preprocessing where things like formatting errors, irregularities or missing bits of information are corrected. This step includes dealing with missing values, normalizing data, and making transformations on variables. These steps are essential to create accurate and viable machine learning models.

***Exploratory Data Analysis (EDA):*** This step includes exploratory data analysis; this approach identifies general patterns in the data. Such as the distribution of the variables and detecting correlations within the data. This step enables us to interpret the feature importance and choose the most reliable models.

***Feature Engineering***: Depending on information from EDA, new features are developed to boost the model’s effectiveness. For this, the data might be processed, which could mean generating new variables from existing data, choosing the most important characteristics, and encoding the categorical data.

***Model Selection:*** A variety of machine learning models are contemplated with respect to their suitability of the data and the specific prediction tasks in question. Models may include regression techniques, decision trees or neural networks.

***Model Training:*** The models are selected to be trained on only a part of the dataset. This allows us to learn the relations between the features and target variable (total energy consumption).

***Model Training:*** Our research will use the models to learn the relationship between the independent variables (features) and the dependent variable (total energy consumption).

***Model Evaluation:*** The models are evaluated against the testing sub-set of the data. The primary performance indicators employed are RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and R-squared. These performance metrics are used to evaluate the accuracy and efficiency of the model.

***Optimization and Validation:*** Data is then input into these models, which are then adjusted based on the initial results to maximize the performance. This might mean altering the parameters, incorporating various feature combinations, or employing advanced ensemble techniques.

***Interpretation of Results:*** An interpretation of the results of the predictive models is formulated in the context of the research question. In addition, insights of what drives energy consumption in smart homes are identified, and the effects of factors on energy control are established.

***Documentation and Reporting:*** Lastly, the findings and the approaches used are explained in explicit detail, enabling an overall coherent summary of the research, comprising of obstacles, limitations, and recommendations for possible future works.

As a result, this systematic method ensures that the research question is addressed in its entirety, resulting in the formulation of accurate machine learning models.

## 3.3 Research design

The selected research approach for my dissertation is a quantitative one relying on machine learning techniques to predict the energy consumption in a smart home. This type of approach is justified as the source data is numeric by nature and thus requires the application of statistical methods. The proposed qualitative methodology will make the system reproducible and precise, with data science, machine learning and energy management.

## 3.4 Data collection method

***Instruments and Tools****:* This study relies on a relevant dataset from Kaggle titled "Smart Home Dataset with Weather Information." The dataset gives targeted minute-via-minute readings of household appliance energy consumption in kilowatts, as measured by smart meters, alongside concurrent climate patterns and energy consumption.

***Justification***: In this regard, the selected dataset is directly proportional to the making of an accurate forecasting tool on how much energy will be used in smart homes. A key feature of this dataset is its detailed time series data which provides an opportunity to examine exact shifts in energy consumption and the influence of weather factors. Therefore, it is in line with our research objectives of improving energy efficiency and minimizing the environmental inputs by using machine learning to predict total energy consumptions in smart homes, based om historical data, weather conditions and appliance usage patterns.

***Data Source Acknowledgment***: The dataset is free to access through Kaggle and remained the latest version released five years ago. It can be accessed through the following URL: <https://www.kaggle.com/datasets/taranvee/smart-home-dataset-with-weather-information>

### 3.4.1 User feedback on Dataset

The 'smart home data with weather data' on the Kaggle website has been used in various sectors including research, learning and application in real-life situations. This underscores the fact that it can be used for a variety of purposes in a very applicable way.

* Learning: 16 users have shared that they have utilized it for educational reasons. To provide students a chance to improve their knowledge regarding the usage of data science in smart home technology.
* Research: 5 users have adopted it for academic and research purposes. This fact underlines the capability for in-depth analysis.
* Application: 3 users have employed this dataset to address real life issues and practical problems. The implementation of this dataset in modern studies highlights its adaptability.

### 3.4.2 Descriptive feedback on Dataset

* The dataset was well-documented, which was one of the reasons two users said its quality and clarity were good for them. This indirectly leads to reliable and adequate data interpretation.
* An individual conveyed her happiness regarding the project, which remained well-maintained. Although updates were not made regularly, they were still content with the dataset.
* Three people replied that the data was well-organized and could be analyzed with minimal preparation. Suggesting that less time and effort will be made to preprocess the data, making it easier for users to conduct their research.
* A reviewer writes that several notebooks have been developed alongside the data frame. Along with other users’ work, getting access to them is an integral component of in-depth analysis, as you can use them for additional data exploration and analysis.

## 3.5 Data analysis

The specific machine learning algorithms like the regression analysis, decision trees, and neural networks are applied as they are capable at modeling complex relationships and forecast outcomes based on previous data.

***Regression Analysis:*** Used for foreseeing power consumption given continuous variables. This method made it possible to ascertain what role each of the parameters played and whether there was a direct relation between the independent variable (weather, appliance usage) and the dependent variable (total energy consumption).

***Decision Trees:*** They are used to model conditional statements, by breaking the data up into branches, the experimenter can isolate one piece at a time, for example, an appliance or weather condition that is consuming energy. This method has proved itself to be extremely effective in such cases where interrelationships between different variables are non-linear.

***Neural Networks:*** Most importantly, neural networks are exceptionally useful for solving complex tasks and discovering hidden patterns in data which would not be immediately obvious. Like the human brain, they are designed to improve the accuracy of predictions through self-studying, by employing different mechanisms.

A range of these methods are selected based on their applicability to the research question hence providing a deeper understanding of the elements and diverse parameters affecting the energy consumption in the smart homes, hence forecasting future usage with extreme accuracy. These methods validate the proposed hypothesis of identifying the power of machine learning in this context as well as proving that it can be used to improve energy efficiency and sustainability in residential places.

## 3.6 Ethical considerations

The research below centers on maintaining top ethical standards, and since it involves the households' energy data the ethical standards is more crucial and sensitive and therefore must be maintained.

Below are the essential ethical considerations for this study:

* ***Data Privacy and Anonymity:*** Privacy will be ensured and validated by anonymizing the dataset to remove any personally identifiable information. Hence preventing the misuse and respecting officials’ privacy
* ***Consent and Transparency:*** Verifying that participants' data collection has been consented to by participants and that they were informed about what would be done with their data is essential.
* ***Avoidance of Bias***: An analysis will continuously look for and neutralize any data biases. Moreover, it shall make sure that the machine learning models do not amplify these biases.
* ***Impact Assessment:*** The impact of the new discoveries in terms of society, environment, and economy will be thoroughly assessed.

Taking these steps helps to enunciate that the study is of high ethical standards, as participants are guaranteed privacy, and the results are of high quality.

## 

## 3.7 Limitations, Validity and Reliability

The key limitations are the availability of real-time information and possible biases in the dataset used, which may lead to the lack of relevant application. Other constraints such as time and resource availability could influence the depth of model building, as well as their training and validation. To comfort these issues and protect the research in its integrity, certain methods have been introduced.

Validation of the instrument is made possible through robust testing of the algorithms against standard benchmarks to ensure that they capture the energy expenditure almost precisely as designed. The reliability of the data is assured by the usage of a well-prepared dataset from a trustworthy source. Error detection methods were used to remove outliers, additional appropriate responses where additionally considered to handle missing data. Cross-validation was used during the test and training section to prevent the models from overfitting. These measures accumulatively will help to strengthen the reliability of the research outcomes and tackle the consequences of the given limitations.

## 3.8 Quality Assurance

To guarantee the quality of the research findings, a variety of rigorous cross-validation methods are applied. In this regard, the model is split into several datasets to help adjust and validate the machine learning models. In addition, by using this approach, we not only decrease the chances of overfitting but also ensure that models perform equally well on different parts of data, providing a more generalized estimate of the model's performance.

Moreover, the review equally checks and scrutinize existing literature in similar studies for consistency and reliability. To validate the research conclusions, this comparison underscores how the research findings are either in conformity with or divergent from the preceding data or theories that further improves the credibility of the research results.

## 3.9 Impact Assessment

The consequences and repercussions of this study would be, hence, far-reaching, especially as it relates to advanced energy technologies that are always evolving to incorporate real-time environmental and behavioral parameters. Data gathered from the research will benefit the overall ambition but minimizing energy waste and lowering ecological impact through giving significant contribution of the insights.

## 3.10 Summary

The methodology of the dissertation is based on machine learning techniques to predict total energy consumption in smart homes using historical data, weather conditions and appliances usage patterns. A quantitative method was developed, which involved studying a diverse dataset from Kaggle with statistical methods. The central hypothesis assumes that machine learning models make use of feature selection techniques to accurately estimate the energy consumption.

Thus, the procedure consisted of data collection, data preprocessing, exploratory analysis and prediction and model testing. The approach, adopted for this purpose, ensures that it is in line with the objective of the dissertation which lies in attaining energy efficiency in smart homes by implementing step by step systemic approach to develop reliable predictive models. This strategy is not only responsible for the right path of consumption management but is also a stage to develop smart home technology.

# CHAPTER 4: REQUIRMENTS

## 4.1 Requirements Gathering

During this interval, requirements were gathered from different sources like academic literature, example reports and analysis of current technologies and datasets. The main idea behind creating the prioritized list was the need for an all-inclusive set of problems that the project needed to fix, to successfully predict energy usage in smart homes through Machine Learning. The need for the project to be reliable, interpretable, and accurate are vital for a successful outcome and understanding of the question at hand.

Requirements prioritization is the process of defining an order for completing tasks based on

their relative importance to the project at hand (Hatton, 2008). MoSCow is a prioritization technique, that implies the identification of the top rank requirements. These requirements are split into four groups:

* ***Must Have:*** these requirements must be fulfilled, otherwise the project has failed.
* ***Should Have:*** these requirements are a high priority but are not critical to the project's success.
* ***Could Have:*** these requirements are desirable, but less important than requirements in the should have and must have category.
* ***Won't Have:*** these requirements will not be implemented in the current project but could be in the future.

## 4.2 Requirements Specification

This section explains detailed requirements that were gathered and categorized as functional, non-functional and user requirements, both are key for model prediction to be developed.

|  |  |  |
| --- | --- | --- |
| ***Requirement*** | ***Description*** | ***Priority*** |
| Machine Learning Environment Setup | Set up a Python 3 environment with libraries like NumPy, Pandas, Scikit-Learn, TensorFlow, and Keras for developing machine learning models. | Must Have |
| High-Quality Dataset | Ensure access to a clean, comprehensive dataset for model training. Relevant data is collected from the "Smart Home Dataset with Weather Information" on Kaggle. | Must Have |
| Data Pre-processing | Implement data cleaning, normalization, and preparation pipelines. | Must Have |
| Model Development | Develop multiple neural network architectures to compare performance. | Must Have |
| Model Evaluation | Test models on validation sets to ensure accuracy and effectiveness. | Must Have |
| Model Optimization | Techniques like hyperparameter tuning for improved model performance. | Should Have |
| Model Deployment | Deploy the model in a cloud environment for accessibility, such as Kaggle or Goggle Collab. | Should have |
| Extended Data Features | Include additional data features to potentially enhance predictions. | Should Have |
| Extended Device Integration | Integrate with additional smart home devices for enhanced data collection. | Could Have |
| Advanced Visualization Tools | Develop advanced tools for visualizing energy consumption and predictions. | Could Have |
| User Customization Features | Allow users to customize alerts and energy usage reports. | Could Have |
| Full Automation of Appliance Control | Automated appliance control based on predictions. | Won’t Have |
| Internationalization | Support for multiple languages and international data formats. | Won’t Have |
| Comprehensive User Training Tools | Develop extensive training modules for system use. | Won’t Have |

Table

By evaluating user requirements, the project can prioritize development efforts effectively to ensure the successful implementation of a machine learning system for predicting total energy consumption in smart homes.

|  |  |  |
| --- | --- | --- |
| ***User Requirement Priority*** | ***User Requirement Description*** | ***Evaluation*** |
| High | Accurate Prediction: The system must accurately predict total energy consumption based on historical data, weather conditions, and appliance usage patterns. | This requirement is crucial for the effectiveness of the system in optimizing energy consumption in smart homes. |
| High | Real-Time Data Integration: The system should integrate real-time weather data and appliance usage patterns to enhance prediction accuracy. | Real-time data integration ensures up-to-date predictions, improving energy management in smart homes. |
| Medium | Scalability: The system should be scalable to accommodate varying sizes of smart homes and different data volumes. | Scalability is important for the system to adapt to different home sizes and data loads without performance issues. |
| Medium | User-Friendly Interface: The system interface should be user-friendly, allowing homeowners to easily access and interpret energy consumption predictions. | A user-friendly interface enhances user experience and encourages homeowners to actively engage in energy management. |
| Low | Data Privacy and Security: The system must prioritize data privacy and security to protect sensitive information related to energy consumption. | While important, this requirement is lower as it is a common consideration in smart home systems. |

Table

## 4.3 Predictors and Outcomes

***Predictors:*** Appliance Usage and Weather Conditions.

***Outcome:*** Total Energy Consumption.

* Use [kW]: Total energy used.
* House Overall [kW]: Aggregate of all appliance usage.
* Gen [kW]: Total generated energy.

# CHAPTER 5: DATA

## 5.1 Dataset

The dataset includes:

* Energy Consumption Data: Measurements from smart meters for individual appliances.
* Weather Data: External conditions such as temperature and humidity impact energy usage.

## 5.2 Data Preparation

### 5.2.1 Import Libraries

A library is a collection of existing functions that can be used in your code. Importing pre-defined modulus allows you to access and utilize functions, classes, and variables. The statement serves as the gateway to pre-written code that allow users to perform specific tasks.

### 5.2.2 Load the Dataset

Equation 1

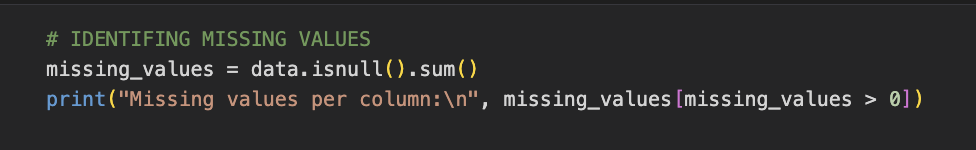
Dataset shape: (503911, 32)

As fetched though the .shape() property a value is returned. This value states the dimensions of the dataset in terms of rows and columns. In this case the dataset has 32 rows (variables) and 503911 columns.

## 

## 5.3 Handling Missing Values

Due to the extensive size of the dataset, the proper management of missing values will be of utmost importance. The results of data analysis are considerably dependent on the ways in which the missing values and outliers are processed. (Kwak SK et al, 2017). To determine the suitable approach for addressing these values, it is crucial to ascertain the quantity and significance of these rows.



Equation

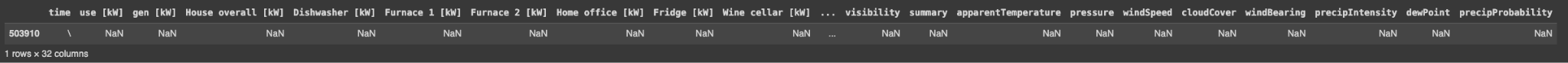
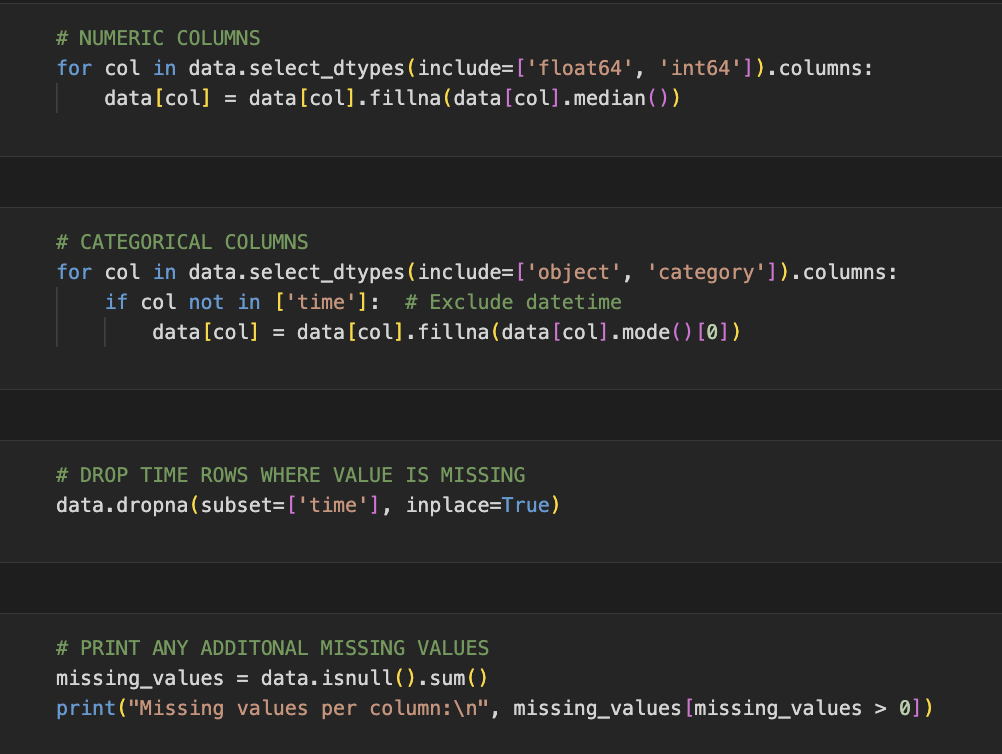


Table 

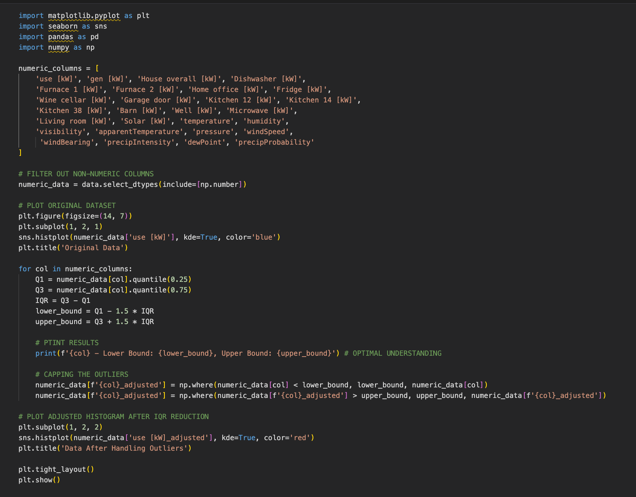
Equation 3

In this scenario, the missing values are present in the entire last row, thus it is necessary to remove it before proceeding. For numerical values either the mean or median can be used, while categorical missing values can be represented by the mode or a placeholder such as "unknown" or "NaN". In the case of missing datetime values, it is essential to remove them. *Equation 3* showcases the different methods in which we have processed the missing values.

## 5.4 Detecting and Handling Outliers

Equation 4

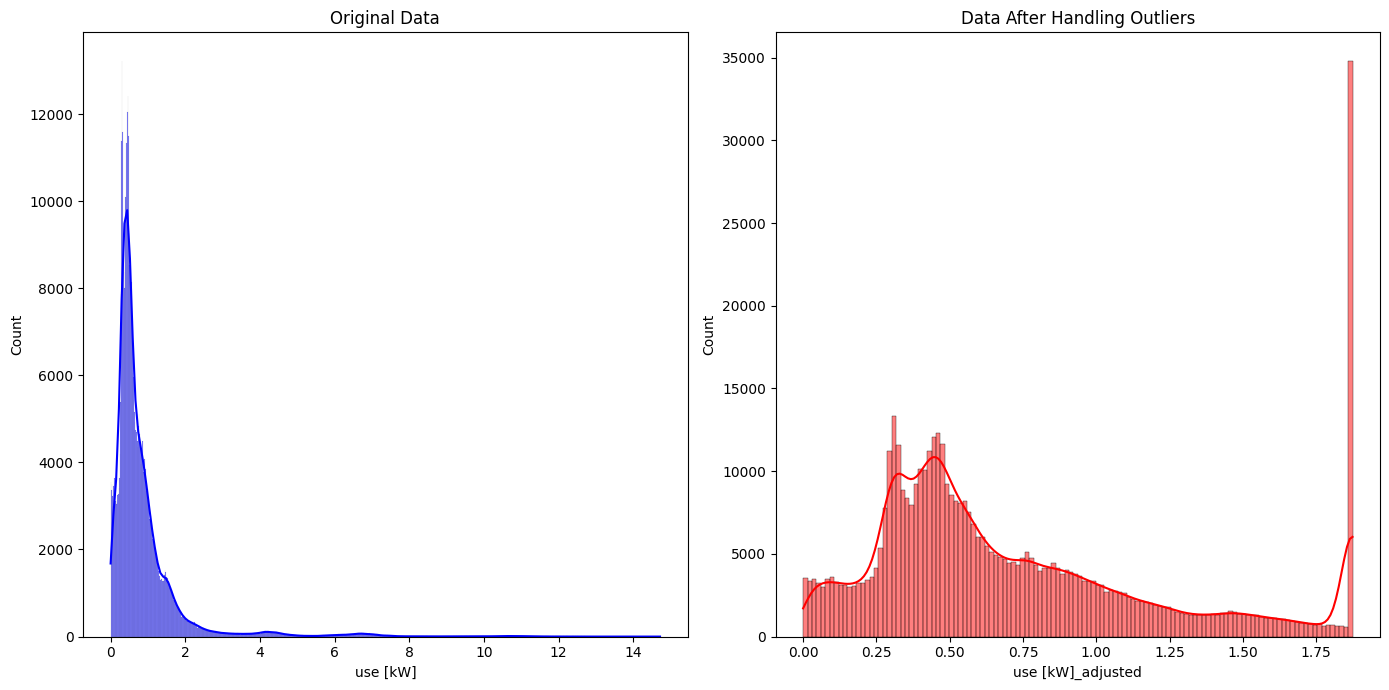
Initially the first method of removing outliers was very drastic and as a result, the number of records in the dataset was reduced greatly, so later the approach was revised and altered. The primary method showcased in *Equation 4* started with the decrease from 503,911 entries into only 5,858. This drop is a significant one, removing a great amount of important information data, hence future calculations may not reflect the real qualities of the overall dataset. This significant drop of the data can affect the accuracy and generalization quality of statistical models created from this dataset possibly resulting in their wrongful function on new, unseen data. Moreover, ignoring outliers without full understanding of their origins may cause significant data loss of necessity for comprehensive analysis.



The code in *Equation 5* starts by precisely identifying all the numeric features that the study covers. This kind of a direct exclusion guarantees that the outlier detection process only involves features that are relevant. The data is further filtered to keep only the numeric ones. The first visualization plot facilitates the understanding of the original data distribution before any adjustments and serves as a benchmark for comparison.

Equation 5

For each column, the Interquartile Range (IQR) is calculated to find the interlevel of its distribution. The code prints out the lower and upper capping limits which are used to deliver the code transparency that can be reviewed manually and adjusted accordingly. Lenient capping is applied using 1.5 times the IQR from the Q1 and Q3. This method adjusts the extreme values without removing them, that way the integrity of dataset is preserved, and zero data loss occurs.



The importance of handling outliers was demonstrated through a histogram. The side-by-side visualization of the data before and after the outlier handling clearly brings about the comparison of the impact of the outlier handling with the original data (*Figure 1*).

Figure 1

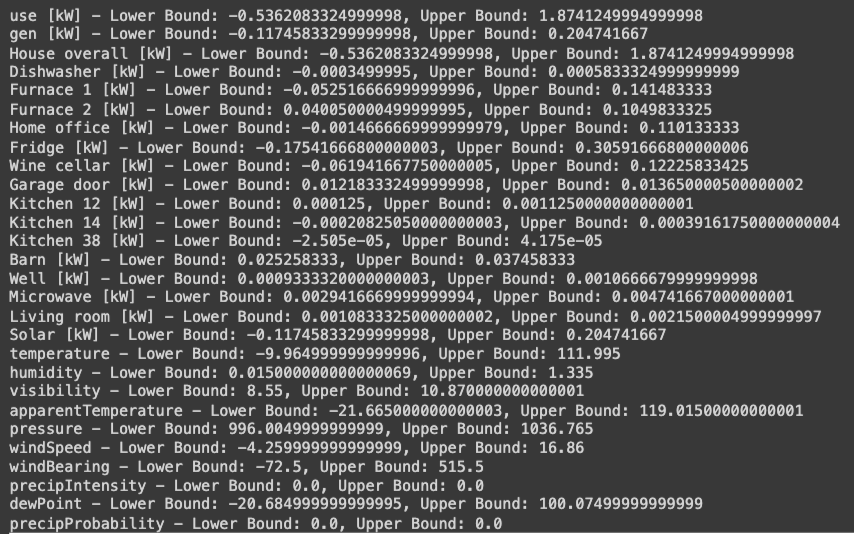
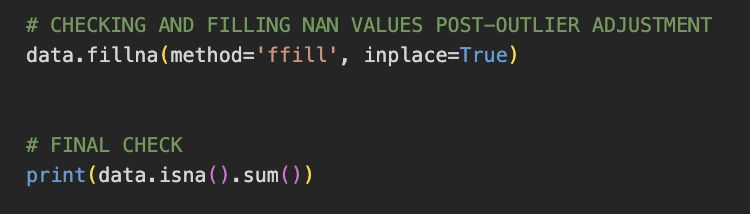
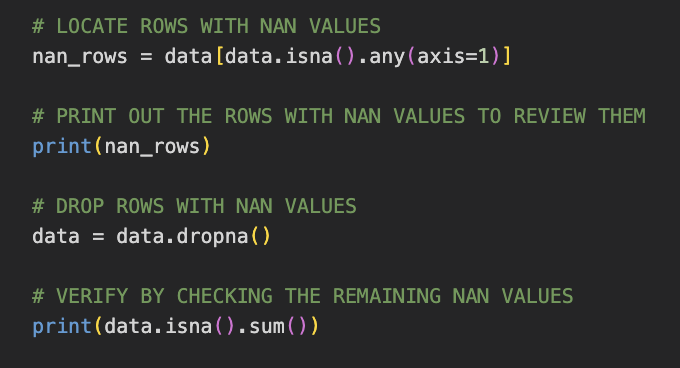


Figure 2

The thresholds in *Figure 2* were established through the "IQR method," a statistical technique applied to identify possible outliers in the data stream. These bounds vary largely across different variables, as these variables have various degrees and kinds of scales and distributions.

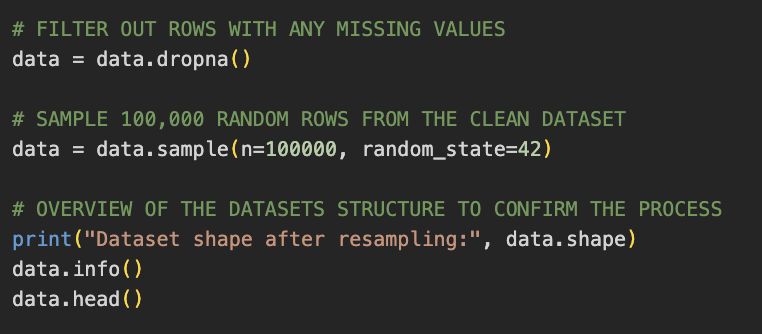


Equation 6

Anomalies turned to NaN Values will remain NaN because:

Equation 7

* They show the distribution of variables and their features without imputed values.
* Allows suitable imputation methods based on the patterns and connections found.
* Highlight other data quality issues or points that could hinder handling missing values.
* Allows for non-biased assumptions.

Due to the extensive size of the dataset, the nan values needed to be removed to complete further analysis. 

## 5.5 Dataset Reduction

I have removed all anomalies before reducing the dataset to 100,000. These data anomalies can skew results, leading to inaccurate insights and decisions.

Equation 8

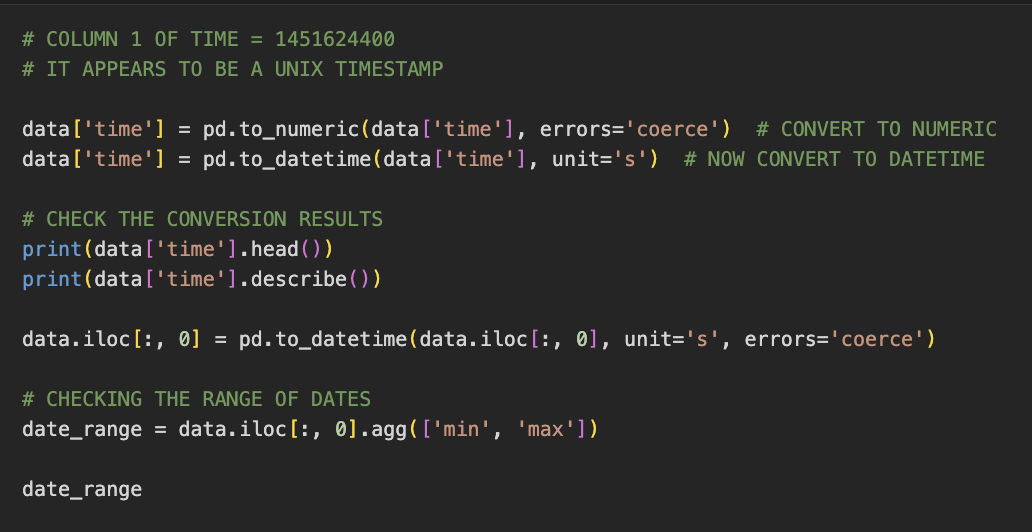
Handling Anomalies before data reduction:

* It facilitates in creating data-driven decision-making process, thus allowing for more focused and resource-efficient data cleaning.
* More tailored and insightful
* This will enable you to understand which features are informative in providing the most reliable imputation method—imputation, i.e., fill or remove with NaN.

## 5.6 Data Conversion

### 5.6.1 Timestamp

The conversion of a timestamp data to a datetime variable is essential for multiple reasons. Firstly, extractions become easier and more manageable when converted, allowing for new time-date features to be derived. Secondly, it allows improves interpretability, allowing the creation of time series plots, aiding in the visualization of trends, patterns, and anomalies over time.



Equation 9

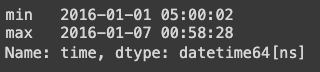


Figure 3

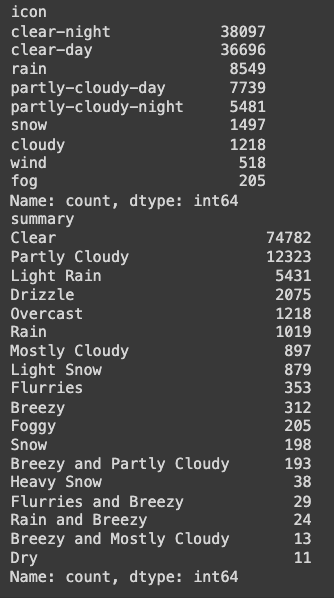
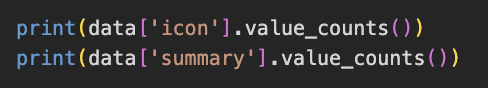
The metadata from Kaggle indicates that the dataset spans 365 days; however, upon detailed inspection, the data only covers the first week of January. The output provided in Figure 3 shows the dataset's earliest record being at 5 AM on January 1, 2016, to just before 1 AM on January 7, 2016.

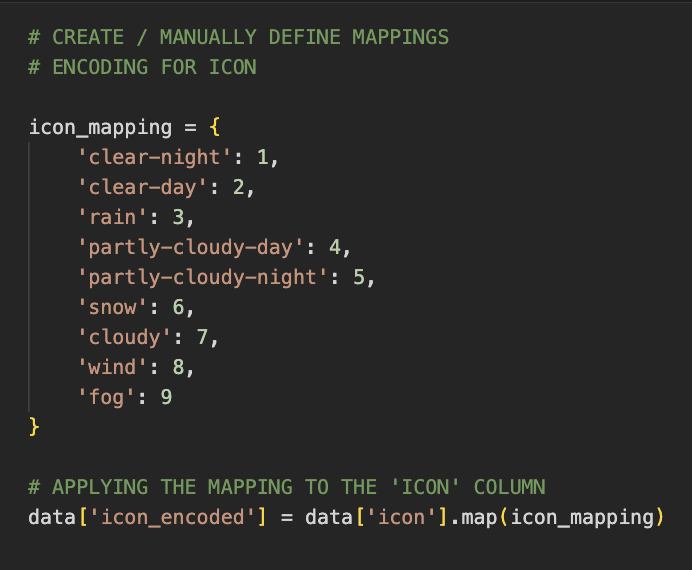
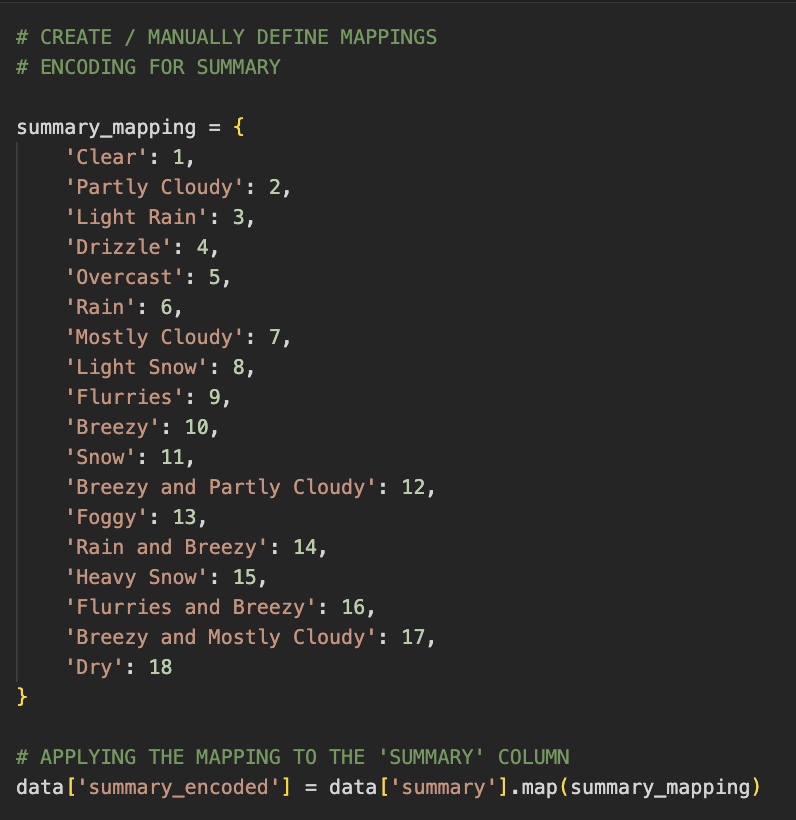
Figure 4

### 

### 5.6.2 Categorical Features

Initially, the dataset variables 'icon' and 'summary' were categorical strings representing weather conditions.

Equation 10

To transform these categories into a format suitable for modelling. *Equation 11/12* showcased a manual encoding scheme for each variable; this was defined where each weather condition/ weather summary was assigned a unique integer. These new conversions were then converted into new columns which contained the new corresponding integer codes.

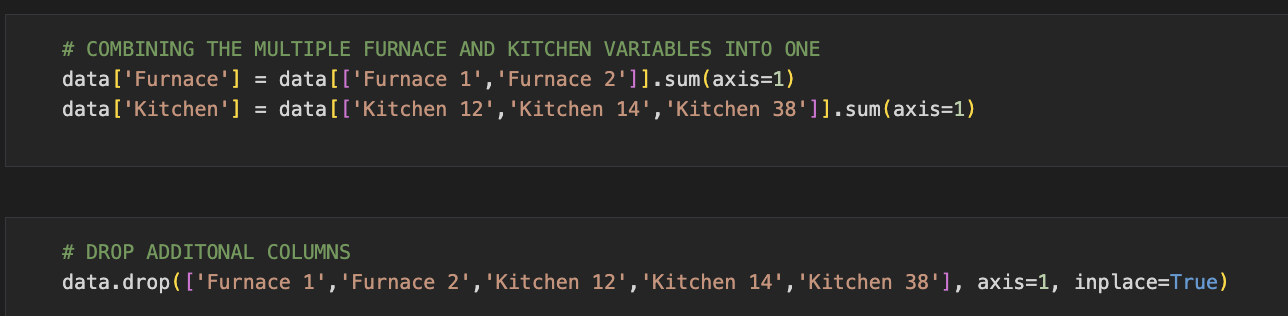
Equation 11

Equation 12

## 

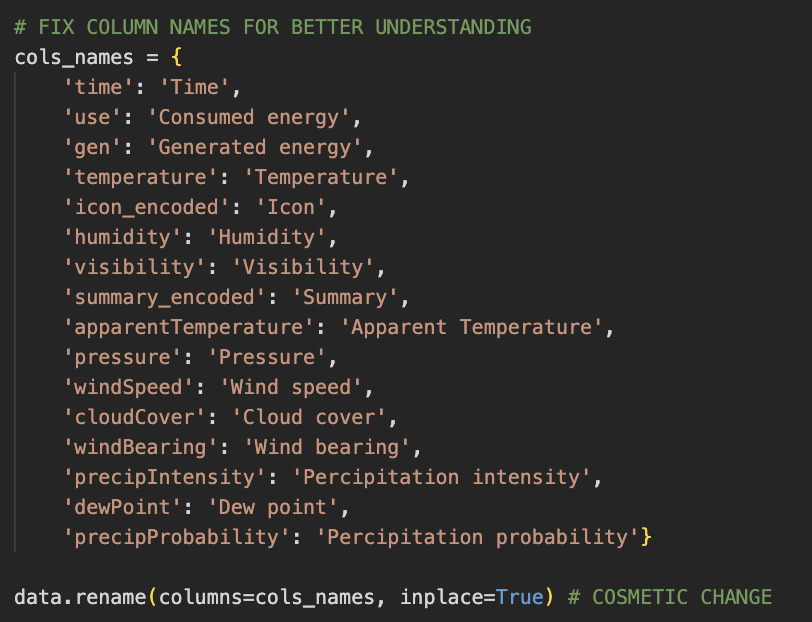
## 5.7 Data Preprocessing

Equation 13

Combining certain variables in the dataset into single columns is motivated by simplifying the dataset and enhancing the analytical robustness. For the columns related to 'Furnace' and 'Kitchen', the original data separates the energy consumption into multiple specific appliances or locations within those categories. Summing these into a single column reduces the number of features. This approach avoids the granularity of separate metrics, which is unnecessary for broad energy usage analysis or predictive modelling.

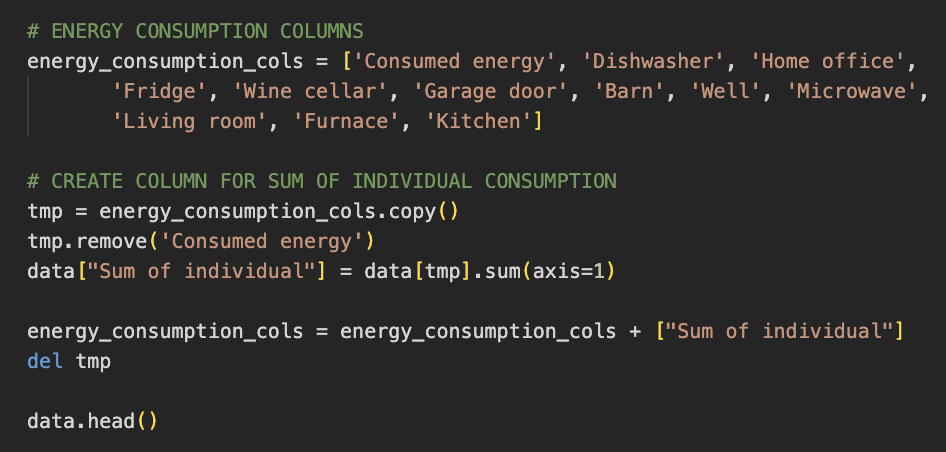
Equation 14

After combining the two furnace variables and three kitchen values, the original variables were removed, along with the icon and summary columns.



Equation 15

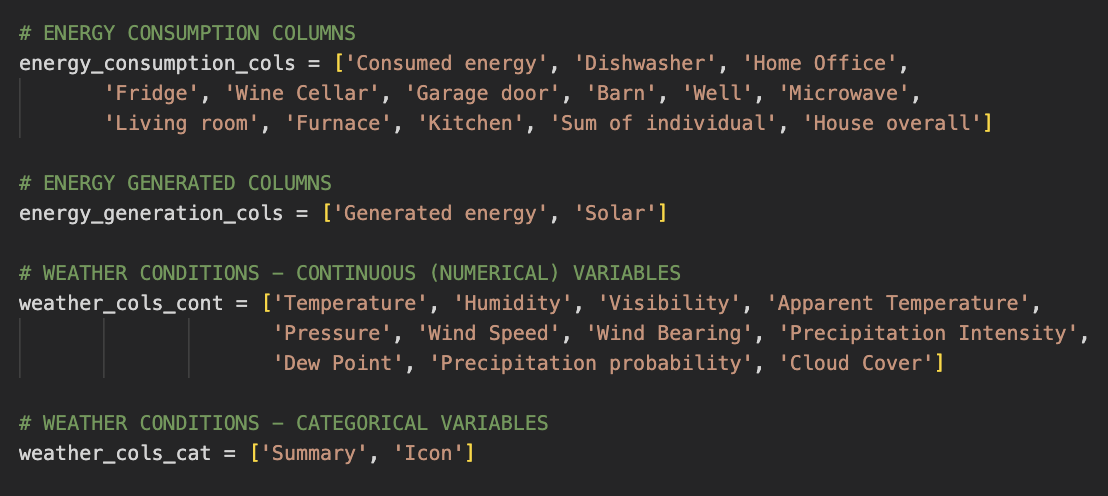
The following change was renaming the columns to more descriptive/understandable names. *Equation 15* showcases how the cosmetic change, columns with intuitive names, allows for more straightforward graphs and tables. Therefore, future stakeholders and researchers can understand what each column represents without needing to refer back to the original data dictionary.

Before moving on to exploratory data analysis, a new column was created. The sum of individual consumption aggregates the total energy consumption from separate devices. By summing the energy consumed by individual categories or devices, an analysis of the total energy usage distributed among the different parts of the household can be made. This allows for future study of where energy is used most and where potential savings could be made.

Equation 16

## 5.8 EDA

This chapter covers a vital step called Exploratory Data Analysis (EDA), which will bring the data's intricate patterns and overarching connections to light. The core responsibility of EDA is to carry out data exploration that will inform the process of building a machine learning system whose primary goal is to predict total energy consumption in smart homes. By analyzing historical data, weather conditions, and power appliance usage trends, we aim to pinpoint key factors that affect energy consumption. This will not only discover essential patterns and outliers, but it’ll also contribute to feature selection, resulting in a precise and easily interpretable model. Assisted with different visualisations and statistical summaries, EDA gives an overall view of the data, which shows possible implications such as correlations and dependencies that can help the effective use of energy and increase the efficiency of smart home systems.



Equation 17

### 5.8.1 - Columns

Firstly, we grouped the columns by type, *Equation 17*.

1. Energy Consumption = energy\_consumption\_cols
2. Energy Generated Columns = energy\_generation\_cols
3. Weather Conditions = weather\_cols\_cat & weather\_cols\_cont

### 5.8.2 Handling of Time Variable

The statistical summary doesn’t work for the datetime data type as it is not numeric. To analyze trends, we must extract features from them, for example, part of the day, hour of the day, day of the week, day of the month, month, and year.

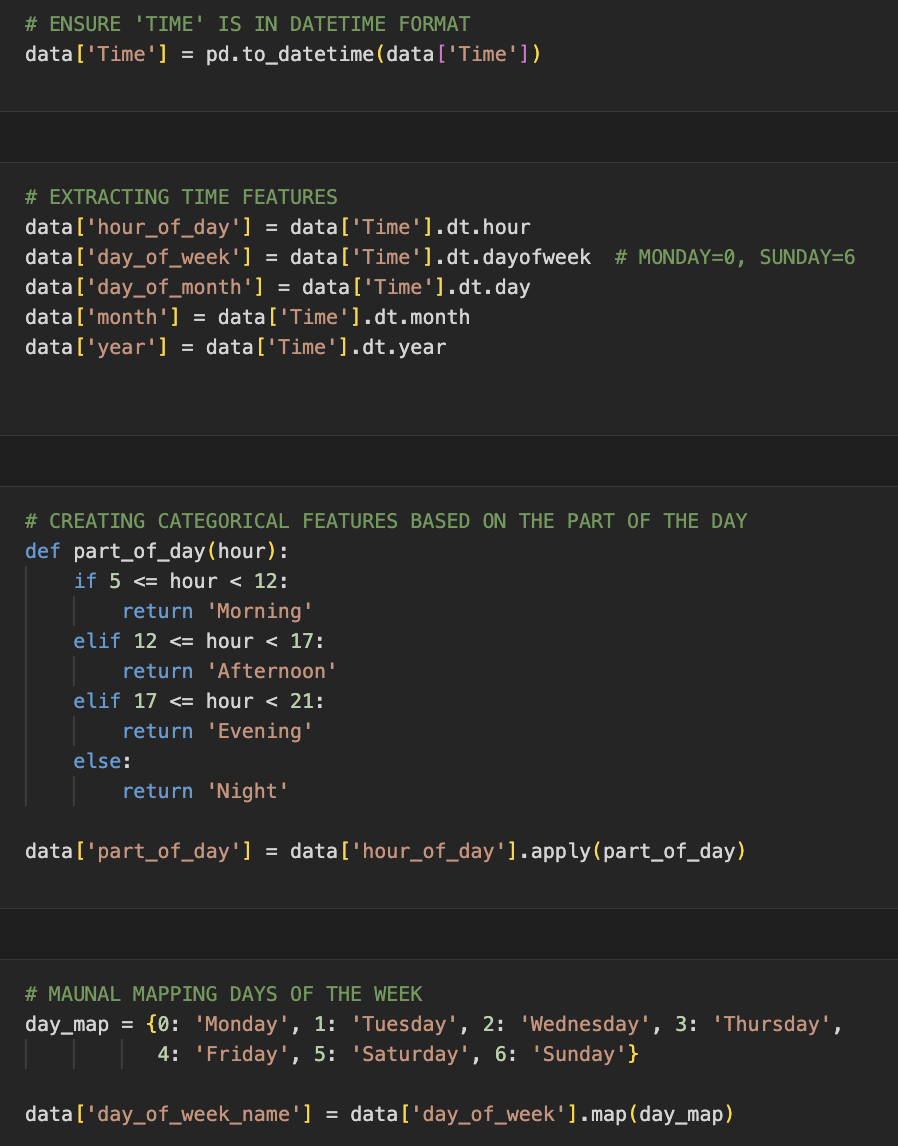
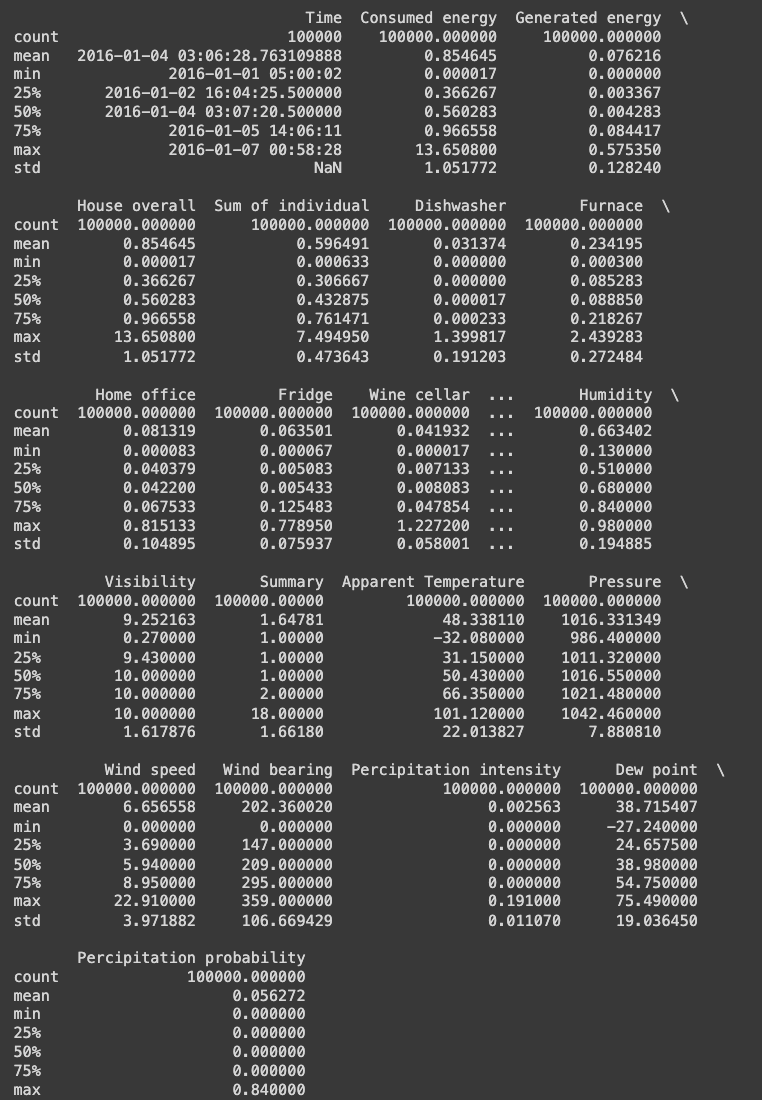


Figure 5

Equation 18

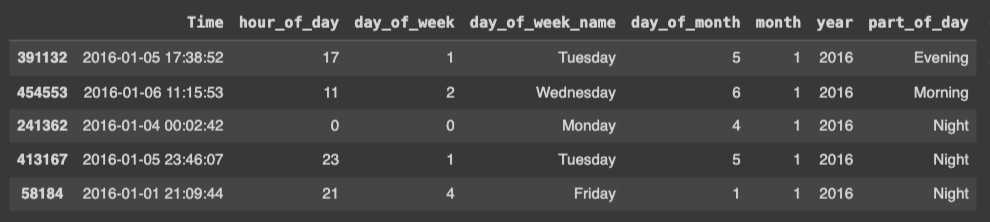
As discussed previously in 5.6 the min and max dates span over the first week of January 2016. Due to this, I can remove the month and year columns as they hold no value. Retaining the ‘month’ and ‘year’ columns would unnecessarily increase the dimensionality of the dataset without offering any benefit towards analysis or modeling.

Table 6

### 5.8.3 Distribution of Individual Variables

Our intention is to understand the correlation of individual variables with the goal of discovering the patterns and outliers that influence energy consumption. This kind of data analysis plays an essential role in the stage of feature selection and the building of the smart home energy usage model so that it is as accurate as real-world behavior.

|  |  |
| --- | --- |
| Figure | Figure |
| *Figure 6/7* shows that the data is evenly distributed across most days, with Thursday having significantly fewer entries. | |
| Figure | Figure |
| Night has the highest number of entries, followed by morning, afternoon, and evening. *Figure 8/9* | |
| Figure 10 | |
| Both histogram and density plot visualisations of consumed energy exhibit a right-skewed distribution. Values appear to be concentrated towards the lower end, with a peak around 1. Notably, there are several.  significant outliers visible in the box plot.  Figure 11 | |
| Figure 12 | |
| The histogram and density plot for the generated energy display a strong concentration near zero. This indicates that most households generate little to no energy, as evidenced by the sharp peak near zero. Moreover, the steep decline in energy generation further supports this conclusion. The box plot depicts a long tail of outliers, which could suggest that most households do not possess environmentally friendly means of energy generation. However, it is worth noting that newer buildings may include solar panels, thereby generating energy.    Figure 13 | |
| Figure 14 | |
| The histogram and density plot depict a relatively uniform distribution with modest peaks in the mid-range, indicating a balanced distribution of temperature values. Additionally, the box plot displays a broad range and a few outliers at the lower end, which highlights that the temperature remains within a consistent range. The diverse range of readings is beneficial for conducting an extensive investigation of energy consumption patterns.  Figure 15 | |
| Figure 16 | |
| All three graphs indicate a uniform distribution over a broad range of values. The density plot reveals a generally consistent rise with peaks suggesting various humidity levels. The box plot shows a wide interquartile range and a stable spread of data. There are no apparent outliers, suggesting a relatively constant distribution.  Figure 17 | |

### 5.8.4 Relationships between Variables

This section emphasizes how estimating the overall energy usage of smart homes requires an understanding of variable distributions. By looking at how energy use is distributed across several factors.

|  |  |
| --- | --- |
| Figure 18 | The first correlation matrix shows the correlation coefficients between two variables. The closer the number is to 1 and -1 the stronger the correlation is. The redder the boxes are the more evident the positive correlation, and the bluer the unit, the more evident the negative correlation.  *Figure 19* was visualized with a threshold so only the variables with high correlation are highlighted. This allows for more in-depth analysis and easier selection for feature importance. The final graph is a Half−matrix plot. |
| Figure 19 | Equation 19  ^ Code to create threshold for correlation matrix |
| Figure 20 | ***Highly correlated variables:***   * Consumed energy & House overall = 1 * Solar & Generated energy = 1 * Apparent Temperature & Temperature = 0.99 * Apparent Temperature & Dew point = 0.90 * Temperature & Dew point = 0.88 * Furnace & Sum of individual = 0.60 * Summary & Icon = 0.58 * Cloud Cover & Summary = 0.57 * Sum of individual & Fridge = 0.54 * Consumed energy & Sum of individual = 0.52 * Sum of individual & House overall = 0.52 * Visibility & Cloud cover = -0.55 * Visibility & Humidity = -0.59 * Kitchen & Fridge = -0.60 * Day of the month & Day of the week = -0.63 * Day of the month & Microwave -0.74 |

***High Correlation (>=0.9):***

* Consumed Energy & House Overall: This suggests that the household's overall energy consumption strongly predicts its total consumed energy.
* Solar & Generated Energy: Solar energy generation is directly linked to the total generated energy.
* Apparent Temperature & Temperature: These two are almost interchangeable due to their high correlation.
* Apparent Temperature & Dew Point, Temperature & Dew Point: High correlation suggests that temperature variables are closely linked to humidity levels represented by the dew point.

***Moderate to High Correlation (0.5 - 0.9):***

* Furnace & Sum of Individual: This indicates that furnace usage is a significant component of total individual appliance energy usage.
* Summary & Icon, Cloud Cover & Summary: These correlations suggest that weather summaries are consistent with specific weather icons and cloud cover, which might influence energy usage patterns.
* Sum of Individual & Fridge, Consumed Energy & Sum of Individual, Sum of Individual & House Overall: These correlations indicate that individual appliance usage patterns (like the fridge) significantly predict overall energy consumption.

***Negative Correlations:***

* Visibility & Cloud Cover, Visibility & Humidity: Lower visibility relates to increased cloud cover and humidity, therefore, could influence temperature control and, as a result, energy usage.
* Kitchen & Fridge: This negative correlation might indicate that when one is used more, the other is used less, possibly due to user behaviors patterns.
* Day of the Month & Day of the Week, Day of the Month & Microwave: These illustrate some frequent patterns in energy use, which may be tied to specific days or events.

### 5.8.5 Distribution of house and weather appliances

|  |  |
| --- | --- |
| Figure 21  Figure 22 | The chart of Total Energy Consumption and Total Energy Generation time series illustrates how energy use and production fluctuate on a weekly basis.  The Energy Consumption schedule sharply drops on day three and peaks very soon after that, stabilizing on day six. Conversely, Energy Generation progressively rises to a mid-period peak and then decreases.  We can create a couple of assumptions from these depictions from the time-series analysis.   * Generation of energy, from renewable sources, increases with good weather conditions and decreases under bad weather. * Energy consumption is driven by household demand, which fluctuates based on daily activities.   The graphs show the disconnect between the highest energy production and the use of these sources. Underpinning the necessity of incorporating better storage solutions, by optimizing consumption patterns and creating a balance between supply and demand. |
| Figure 23 | This graph shows the density distribution of energy consumption across various household appliances.  Insights =   * Most appliances have a peak usage below 0.2 kW. * Few appliances like the dishwasher show energy consumption up to 1.4 kW. * There is a variance in energy consumption across different appliances.   Relevance =  These distributions can help identify which appliances have significant energy consumption, influencing model features. Understanding typical and extreme values helps in establishing baseline energy usage profiles. |
| Figure | Existing weather distribution graphs give the overall picture of different weather conditions parameters and their frequency distributions.  The Temperature (Distribution) displays a “bimodal pattern”, meaning two temperature ranges, presumably related to the day and night cycles.  The Apparent Temperature Distribution is almost the same as the Temperature Distribution and presents two exceptional peaks, which is an example of how temperature is perceived including high humidity and wind.  The Pressure Distribution is seen to have a bell shape curve centered on 1015 hPa with most readings falling around this value.  Distribution of Wind Speed has the right slope, and most of the occurrences are of low speed with less and less tapers as their speeds increase.  Intensity of the Cloud Cover Distribution is a combination of peaks revealing the constant changes in the environment and therefore, clear, and overcast skies stretch wider as the single most common feature.  The Distribution of Precipitation Intensity is highly right-skewed, which implies that during the recorded time period, the intense precipitation was not as frequent.  The graph of Dew Point Distribution exhibits double-peaking like Temperature Distribution and Apparent Temperature Distribution graphics, indicating high levels of humidity level in the air.  Combining such patterns into model trajectories will enable us to validate forecasts regardless of whether they are typical or not. |

### 5.8.6 Time Series Analysis

|  |  |
| --- | --- |
| Figure 25 | Shown by the time series chart of appliance energy use the appliances show different usage patterns and power trends for several home devices.  *Figure 25* –  Often, the dishwasher exhibits cyclical energy consumption during peak times, which means that its operation is likely to occur during meals. This discovery can be useful for identifying peak consumption periods.  Home office reflects the same pattern under minor fluctuations that indicate constant energy usage in accordance with regular work routine which implies it to be stable over the time in energy forecasting.    The wine cellar had a notable mid-point peak, probably associated with environmental temperature variation, pointing to the necessity of temperature data inclusion in predictions. |
| Figure | *Figure 26* -  The garage door represents low usage but with clear peaks indicating an individual's particular times of entry or exit, which are helpful for predicting the energy usage when homeowners are in and out of the house.  The microwave shows a continuous uprise of which peak at mealtimes, this notable change can be used to consider what happens on the daily bases.  A furnace's peaks (corresponding to heating requirements), and troughs (corresponding to temperature change) appear at the start and end, with a slope in between. This suggests the necessity of seasonal readjustments in predictions.  The kitchen has high usage rates, and the fluctuations are in perfect congruence with the mealtimes, hence the need to identify and highlight peak usage periods. |
| Figure 27 | The fridge illustrates a mid-period peak with descending pattern characteristic of cooling cycles and constant consumption.  The barn exhibits multi-peak and trough form, implying frequent variations of activities, which directly generate variability of energy forecasts.  Energy consumption in the living room varies significantly and shows two distinct elevated peaks. Inferring that there must be a direct relationship between energy consumption and the varying occupancy and activity patterns, as most variables seem to fluctuate at 1/05-1/06.  Having these insights put along with weather-forecasts and machine learning tools can markedly improve the energy consumption prediction accuracy in smart home systems by resorting to both habitual and external affecting trends. |
| Figure | The Temperature Time-Series and Dew Point Time-Series display parallel trends, with their upward and downward movement in synchrony. This suggests that as the temperature rises, the humidity increases proportionately, thus showcasing a close relationship between dewpoint and temperature.  The Visibility Time-Series and Cloud Cover Time-Series disclose the inverse proportion between the higher cloud cover and visibility decrease; with this fact we can state that thick cloudiness is often associated with the reduced atmospheric clarity.  The series plots of Pressure and Wind Speed are also related as variations in atmospheric pressure events have found to contribute to fluctuations in wind speed, which can be used to predict wind patterns. |
| Figure | The respective spikes in the intensity time-series of precipitation line up with increased cloud cover because usually clouds are directly related to heavier chances of precipitation.  These interdependencies are one of the major factors for the improvement in the precision of energy consumption prediction models, as they show how weather parameter affecting others can influence energy consumption in smart homes.  Through recognition of the relationships, predictive models become more capable of factoring in and coping with the intricacy in the dynamics of weather conditions variations between various weather conditions. |

### 5.8.7 Forecasting

|  |  |
| --- | --- |
| Figure | The Energy Consumption Forecast over a given time-series, which illustrates the past daily energy consumption (in blue) and the forecasted consumption (in red) based on an ARIMA model.  The daily electrical energy usage varies, with a noticeable peak and troths followed by a stable phase, according to historical statistics. Thus, predicting the following thirty days is the responsibility of the ARIMA model, with an order of (1, 1, 1).  According to the horizontal charted line, the model expects smooth consumption with constant levels across the whole forecasting period. This shows that the model mainly captures the long-term trend. The biggest issue with the flat predicted line is that it fails to emphasize the short-term changes that were previously observed in the historical data.  Although the ARIMA model offers a strong basis, it may be significantly enhanced by including accurate meteorological data and optimizing the current model's parameters. The following comprehensive method enables a more precise recording of historical consumption variations and contributing variables like the climate and routine household tasks. This leads to more accurate energy consumption projections for smart home control. |

## 5.8 ​​Model Development

### 5.8.1 Regression Model Metrics

|  |  |
| --- | --- |
| Figure 31 | With the purpose of determining starting analytics, the evaluation was conducted of primary machine learning models with Gradient Boosting, Random Forest, and Linear Regression as the focus of the analysis. *Figure 31* displayed the performance indicators.  Through these models, GB demonstrated incredibly high errors, and the RF and LR turned out quite good.    Figure 32 |
|  | After the initial study we widened the investigation to possible further variables which we found to improve the prediction accuracy. This brought to include SVR (Support Vector Regression), Lasso, Ridge, ElasticNet, Decision Tree, Bayesian Ridge, and Neural Network.  Perfect R^2 scores of 1 demonstrated the exceptional prediction accuracy of machine learning models, including Bayesian Ridge, Decision Tree, Random Forest, Ridge, and Neural Network models.  Gradient Boosting and Neural Network maintained strong performance with high R^2 values close to 1, despite the initial high errors observed.  SVR also performed well with an R^2 of 0.9880. On the other hand, Lasso displayed the weakest performance, with a low R^2 and high MSE, indicating poor predictive capability, while ElasticNet showed moderate performance. |
|  | This step-by-step method, beginning from a limited number of models and then progressing to a more varied collection, helped us to pinpoint the models that were superior for predicting the overall energy consumed in smart homes.  From the results of this comprehensive study, Random Forest, Gradient Boosting, in addition to Neural Networks, indisputably appear to be the only methods that provide the highest level of predictive accuracy, thus, making them very appropriate for the complicated tasks such as forecasting. |

### 5.8.2 Neural Network Architecture

|  |  |
| --- | --- |
| Figure | The Actual and Estimated Energy Consumption graph illustrates the model's ability to pinpoint actual energy utilization. The scheme contrasts the predicted consumption (Y-axis) with the actual consumption (X-axis), the line of a perfect prediction is illustrated in red. Almost all points fall along the line, implying high accuracy of the predictions. The Mean Absolute Error (MAE) is low at 0.09, reflecting minimal deviation between actual and predicted values, which further confirms the model's quality and effectiveness.    Equation |
| Figure | The training and validation process initially involved 10 epochs, where significant improvements were observed in both training and validation metrics.  *Equation 20* highlights that in the first epoch, the training loss was 0.0906 with an MAE of 0.1659, while the validation loss and MAE were 0.0075 and 0.0598. By the tenth epoch, the training loss decreased to 0.0205, and the MAE to 0.0692. The validation metrics showed slight rise, indicating possible overfitting. |
| Figure | Equation |
| Figure | The training was prolonged to 50 epochs. The further refined model produced an MSE score of 0.0093 and MAE to 0.0661. The data maintained consistent loss reduction over time.  This extended training showcased the model's robust learning and generalization capabilities, effectively balancing accuracy, and overfitting concerns, reinforcing its suitability for predicting energy consumption in smart homes.  The combination of historical energy data, weather conditions, and appliance usage patterns, modelled through a neural network, demonstrated effective prediction capabilities, as evidenced by low error metrics and close alignment between actual and predicted values |

## 5.9 Model Optimization and Validation

The model optimization and validation process involved advanced techniques such as hyperparameter tuning and cross-validation to enhance the accuracy of predicting total energy consumption in smart homes. The neural network model was refined using a Random Search tuner to identify the optimal combination of parameters. This process is crucial for aligning the model’s predictive capabilities to leverage historical data, weather conditions, and appliance usage patterns to forecast energy consumption accurately.

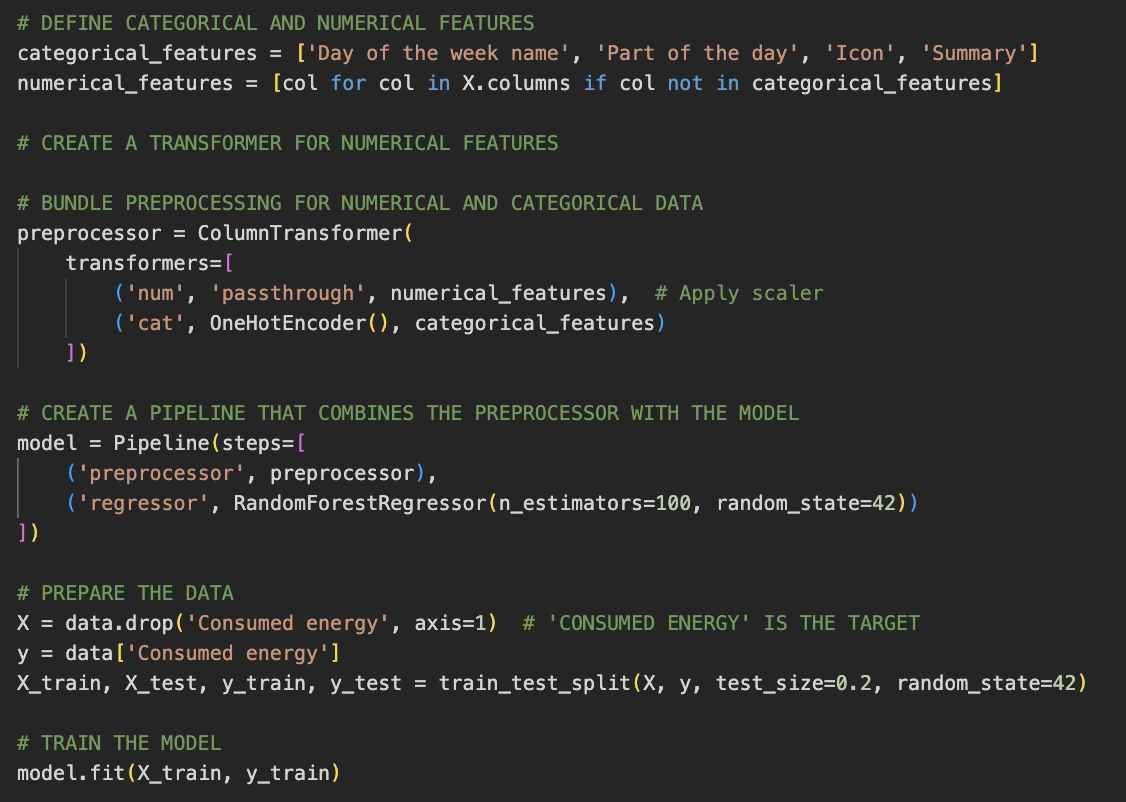
|  |  |
| --- | --- |
| .  Equation    Equation | A function called build\_model was designed to build a neural network with customizable parameters.  Varying the number of units in dense layer from 32 to 512 was incorporated as step 32. In addition, the dropout rate is also adjusted between 0. 0 and 0. 5.  Such adjustments helped the model strike a balance between the complexity and preventing cases of overfitting.  The model uses the Adam optimizer and MSE loss function with MAE metric as the supplementary indicator for training accuracy evaluation.  The RandomSearch tuner was configured for minimizing MAE as validation objective, 10 trials maximum and a minimum of three trainings to every trial to provide accurate assessment.  The best model for tuning 10 epochs achieved a validation MAE of 0.000918. This highlights a notable increase in accuracy of the predictive power. |
| Equation    Equation | The tuning process ran for 50 epochs, incorporating an early stopping callback to prevent unnecessary training once validation loss ceased to improve.  Our perfect model structure reached 192 units in the dense layer and a dropout rate of 0.2 comprising a total of 13,057 trainable parameters.  The best model demonstrated outstanding performance on the test data, with a test MSE of 4.97 × 10^-7 and a test MAE of 0.000373. |
| Equation    Equation | The results relate directly to the main goal as these can bring evidence of high accuracy regarding the model’s capability to effectively predict energy consumption. |

## 5.10 Interpretation and Recommendations

### 5.10.1 Train Model

The model training and feature importance analysis process incorporates a pipeline involving a RandomForestRegressor, with preprocessed steps for categorical and numerical features. One-hot encoding was used to transform categorical features, while numerical features were directly passed through. This series of steps in a ColumnTransformer class will make preprocessing and model fitting seamless.

|  |  |  |
| --- | --- | --- |
| Figure | |  | | --- | |  |   Figure 38 |



The data was split into training and testing sets, and the model was trained to predict ‘Consumed energy.’ After training, feature importance was extracted and normalized using MinMax scaling. The resulting importance values were visualized on a horizontal bar chart with a logarithmic scale, highlighting the most significant predictors of energy consumption.

Equation 28

|  |  |
| --- | --- |
|  | 'House overall,' 'Wind speed,' 'Part of the day: morning,' and 'Sum of individual' are the main factors in forecasting energy usage, according to the feature significance analysis.  The existence of these variables proves their critical nature for the provision of energy consumption forecasts; they stress the fact that the home factors, weather, and time of day are undoubtedly of the greatest significance.  This customized procedure featured strong model training and efficient preprocessing. The assessment of these parameters advances the objective of energy forecasting in smart homes using machine learning. |

## 5.11 Discussion and Analysis of Findings

Chapter 5 provides evidence-based analysis and deciphering to predict smart energy consumption in the home. The chapter underlines the necessity of data preparation to address missing values, outliers, and getting categorical variables right. All these play a signal role in training the models.

Exploratory data analysis proved that temperatures (and dew points) and cloud cover (and visibility) have such strong correlations, which is essential for understanding consumption dynamics. The evaluation of machine learning algorithms, especially the Random Forest, Gradient Boosting, and Neural Networks, has proved their great predictive capability, with them having high R² and low MSE values, respectively.

Variable importance analysis showed strong links between overall house metrics, peak hours, and wind speed. Moreover, sophisticated methods like hyperparameter adjustment and cross-validation noticeably improved models' performance. The results also correspond to the main idea, which is utilizing historical information, weather conditions, and patterns for the usage of appliances to optimize energy management in smart homes.

# CHAPTER 6: REQUIREMENT EVALUATION

## 

## 6.1 - Summary of Requirements

An analysis of the completion of the requirements described in the Project Requirements is presented in this section.

|  |  |  |
| --- | --- | --- |
| ***Requirement*** | ***Priority*** | ***Evaluation*** |
| Machine Learning Environment Setup | Must Have | Achieved |
| High-Quality Dataset | Must Have | Achieved |
| Data Pre-processing | Must Have | Achieved |
| Model Development | Must Have | Achieved |
| Model Evaluation | Must Have | Achieved |
| Model Optimization | Should Have | Achieved |
| Model Deployment | Should have | Partially Implemented |
| Extended Data Features | Should Have | Achieved |
| Extended Device Integration | Could Have | Not Implemented |
| Advanced Visualization Tools | Could Have | Achieved |
| User Customization Features | Could Have | Not Implemented |
| Full Automation of Appliance Control | Won’t Have | Not Implemented |
| Internationalization | Won’t Have | Not Implemented |
| Comprehensive User Training Tools | Won’t Have | Not Implemented |

Table

## 6.2 Analysis of Requirement Fulfillment

All “must -have” requirements were met to ensure the comprehensive treatment of the project core elements. These included efficient optimization and assessment, robust model performance, and thorough data processing.

All "should have" requirements were met to their fullest potential, enhancing the project's flexibility and usability during implementation. However, the “could have" requirements of the project have not been examined to their full potential due to data limitations and time restrictions that were set for the project's implementation. None of the "won’t have" requirements were met.

## 6.3 Recommendations for Future Work

Following the project results, the recommendations could involve Further integration with more home devices, which will result in better information collection. The development of models to handle the real-time data stream and provide continuous prediction updates should also be considered, as it allows users to retrieve and update their smart home devices accordingly. Additional personalized features related to energy usage reports and alerts could also be included.

## 6.4 Conclusion

The requirement evaluation shows that the project succeeded in implementing the main objective. Implementing machine learning models to predict the energy consumption in smart homes. Analytical performance, model design, and optimization are key success factors that have led to project achievement goals. In the future, it is possible to enhance the framework's applicability and user engagement further where more efficient and sustainable energy management in residential settings will be promoted.

# CHAPTER 7: CONCLUSION

“Using machine learning models to predict the total energy consumption in smart homes based on historical data, weather conditions, and appliance usage patterns?”

## 7.1 Research Aims, Objectives, and Questions

The primary goal of the research is to determine how machine learning models, using historical data and weather conditions, can predict total energy consumption in smart homes based on power usage patterns of individual appliances. The study's goals were to develop a forecasting model using machine learning techniques to assess the influence of weather on domestic energy consumption and appliance loading and discover the correlation using time-series data to understand the patterns of energy consumption.

The specific objectives are:

1. Develop a predictive framework using machine learning.
2. Evaluate the impact of weather conditions and appliance usage on energy consumption.
3. Identify patterns and trends in energy consumption over time.

## 7.2 Key Findings

Developing machine learning models, especially neural networks, and decision trees, helped increase the precision of energy consumption predictions. Consequently, it was discovered that weather strongly contributes to total energy output. Influential features include temperature, humidity, and wind speed. Additionally, HVAC systems and refrigerators are shown to be the leading causes of total energy usage. Integrating IoT and data in real-time increased the system's prediction accuracy.

## 7.3 Relation to Research Aims and Objectives

This research achieved a machine learning-based model utilizing historical and real-time data to predict energy consumption, accomplishing the target goal. The study showed that weather conditions are significant factors controlling the amount of energy spent, which correlates with the second objective. By achieving the third objective, time-series analysis and feature engineering helped visualize standout usage patterns.

## 7.4 Limitations of the Study

Although complete, the dataset covers only one week (January 2016). A more extensive dataset would undoubtedly result in a more precise model. The observations are based on data from a specific spatial location and may not be widely applicable. Even with its success, the model could be enhanced with more intricate algorithms and further parameter tweaking.

## 7.5 Suggestions for Improvement

***Larger Datasets:*** Future research should employ datasets encompassing multiple seasons and years to increase model robustness.

***Advanced Algorithms***: Employing advanced algorithms, such as deep learning or hybrid models, could improve prediction accuracy.

***Real-time Data Integration:*** Incorporating real-time data from various smart devices could lead to more dynamic and adaptive models.

## 7.6 Implications and Recommendations

### 7.6.1 Practical Applications:

The research has developed a framework that practitioners could adopt to create an efficient home energy management system. This system could lead to considerable savings and environmental benefits.

### 7.6.2 Future Research Directions:

***Dynamic Models:***  Research in this field should concentrate on creating new models capable of dynamically adjusting to the variations in weather conditions and usage patterns.

***Integration with Renewable Energy***: Exploring the possibility of renewable energy feedstock usage in the predictive framework lies at the heart of the environmental sustainability issue.

***User Behavior Analysis:*** More studies could uncover the effect of people’s behaviors on energy, eventually leading to the development of user-oriented energy management solutions.

Such results manifest the possibilities of machine learning in increasing the energy efficiency and management of smart homes, thus establishing a direction for future improvements in this field.

# REFERENCES

Xu, H., He, Y., Sun, X., He, J., & Xu, Q. (2020). Prediction of thermal energy inside smart homes using IoT and classifier ensemble techniques. Computer Communications, 151, 581-589. <https://www.sciencedirect.com/science/article/abs/pii/S0140366419313416>

Zainab, A., Refaat, S. S., & Bouhali, O. (2020). Ensemble-based spam detection in smart home IoT devices time series data using machine learning techniques. Information, 11(7), 344. <https://www.mdpi.com/2078-2489/11/7/344>

Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the internet of things using big data analytics. Computer Networks, 101, 63-80. <https://www.sciencedirect.com/science/article/abs/pii/S1389128616000086>

Iqbal, F., Altaf, A., Waris, Z., Aray, D. G., Flores, M. A. L., Díez, I. D. L. T., & Ashraf, I. (2023). Blockchain-Modeled Edge-Computing-Based Smart Home Monitoring System with Energy Usage Prediction. Sensors, 23(11), 5263. <https://www.mdpi.com/1424-8220/23/11/5263>

Balta-Ozkan, N., Davidson, R., Bicket, M., & Whitmarsh, L. (2013). Social barriers to the adoption of smart homes. Energy Policy, 63, 363-374. <https://www.sciencedirect.com/science/article/abs/pii/S0301421513008471>

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable and Sustainable Energy Reviews, 33, 102-109. <https://www.sciencedirect.com/science/article/abs/pii/S1364032114000914>

Edwards, R. E., New, J., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings, 49, 591-603. <https://www.sciencedirect.com/science/article/abs/pii/S0378778812001582>

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. Frontiers of Computer Science, 14, 241-258. <https://link.springer.com/article/10.1007/s11704-019-8208-z>

Ahmad, W. D., & Bakar, A. A. (2020). Ensemble machine learning model for higher learning scholarship award decisions. International Journal of Advanced Computer Science and Applications, 11(5). <https://www.proquest.com/docview/2655154141?fromopenview=true&pq-origsite=gscholar&sourcetype=Scholarly%20Journals>

El-Gohary, N., & Amasyali, K. (2016). Machine Learning-Based Building Energy Consumption Prediction. <https://www.researchgate.net/publication/334790540_Machine_Learning-Based_Building_Energy_Consumption_Prediction>

Tischer, H., & Verbic, G. (2011, November). Towards a smart home energy management system-a dynamic programming approach. In 2011 IEEE PES Innovative Smart Grid Technologies (pp. 1-7). IEEE. <https://ieeexplore.ieee.org/abstract/document/6167090>

Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. Renewable and Sustainable Energy Reviews, 56, 215-225. <https://www.sciencedirect.com/science/article/abs/pii/S1364032115013179>

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. Energy and Buildings, 40(3), 394-398.

Xu, H., He, Y., Sun, X., He, J., & Xu, Q. (2020). Prediction of thermal energy inside smart homes using IoT and classifier ensemble techniques. Computer Communications, 151, 581-589. <https://www.sciencedirect.com/science/article/abs/pii/S0140366419313416>

Zainab, A., Refaat, S. S., & Bouhali, O. (2020). Ensemble-based spam detection in smart home IoT devices time series data using machine learning techniques. Information, 11(7), 344. <https://www.mdpi.com/2078-2489/11/7/344>

Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the internet of things using big data analytics. Computer Networks, 101, 63-80. <https://www.sciencedirect.com/science/article/abs/pii/S1389128616000086>

Iqbal, F., Altaf, A., Waris, Z., Aray, D. G., Flores, M. A. L., Díez, I. D. L. T., & Ashraf, I. (2023). Blockchain-Modeled Edge-Computing-Based Smart Home Monitoring System with Energy Usage Prediction. Sensors, 23(11), 5263. <https://www.mdpi.com/1424-8220/23/11/5263>

Balta-Ozkan, N., Davidson, R., Bicket, M., & Whitmarsh, L. (2013). Social barriers to the adoption of smart homes. Energy Policy, 63, 363-374. <https://www.sciencedirect.com/science/article/abs/pii/S0301421513008471>

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable and Sustainable Energy Reviews, 33, 102-109. <https://www.sciencedirect.com/science/article/abs/pii/S1364032114000914>

Edwards, R. E., New, J., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings, 49, 591-603. <https://www.sciencedirect.com/science/article/abs/pii/S0378778812001582>

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. Frontiers of Computer Science, 14, 241-258. <https://link.springer.com/article/10.1007/s11704-019-8208-z>

Ahmad, W. D., & Bakar, A. A. (2020). Ensemble machine learning model for higher learning scholarship award decisions. International Journal of Advanced Computer Science and Applications, 11(5). <https://www.proquest.com/docview/2655154141?pq-origsite=gscholar&fromopenview=true&sourcetype=Scholarly%20Journals>

El-Gohary, N., & Amasyali, K. (2016). Machine Learning-Based Building Energy Consumption Prediction <https://www.researchgate.net/publication/334790540_Machine_Learning-Based_Building_Energy_Consumption_Prediction>.

Tischer, H., & Verbic, G. (2011, November). Towards a smart home energy management system-a dynamic programming approach. In 2011 IEEE PES Innovative Smart Grid Technologies (pp. 1-7). <https://ieeexplore.ieee.org/abstract/document/6167090>

Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. Renewable and Sustainable Energy Reviews, 56, 215-225. <https://www.sciencedirect.com/science/article/abs/pii/S1364032115013179>

Iram, S., Al-Aqrabi, H., Shakeel, H., Farid, H., Riaz, M., Hill, R., Vethathir, P., & Alsboui, T. (2023). An Innovative Machine Learning Technique for the Prediction of Weather Based Smart Home Energy Consumption. IEEE Access, 11, 76300-76320. <https://doi.org/10.1109/ACCESS.2023.3287145>.

Rambabu, M., Ramakrishna, N., & Polamarasetty, P. (2022). Prediction and Analysis of Household Energy Consumption by Machine Learning Algorithms in Energy Management. E3S Web of Conferences. <https://doi.org/10.1051/e3sconf/202235002002>

Bandyopadhyay, A., & Bhattacharya, A. (2021). Residential Appliance Usage Patterns From Overall Energy Consumption Data: A Statistical Machine Learning Approach. Volume 8A: Energy. <https://doi.org/10.1115/imece2021-70122>.

Manivannan, M., Najafi, B., & Rinaldi, F. (2017). Machine Learning based Short-term Prediction of Air-conditioning Load through Smart Meter Analytics. Energies, 10, 1905. <https://doi.org/10.3390/EN10111905>.

Hameed, M., Yassen, E., & Jasim, W. (2023). Enhancement Methods for Energy Consumption Prediction in Smart House based on Machine Learning. Iraqi Journal for Computer Science and Mathematics. <https://doi.org/10.52866/ijcsm.2023.04.04.008>.

Lekidis, A., & Papageorgiou, E. (2023). Edge-Based Short-Term Energy Demand Prediction. Energies. <https://doi.org/10.3390/en16145435>.

Almughram, O., Zafar, B., & Slama, S. (2022). Home Energy Management Machine Learning Prediction Algorithms: A Review. Advances in Intelligent Systems Research. <https://doi.org/10.2991/aisr.k.220201.008>.

Chou, J., & Tran, D. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy. <https://doi.org/10.1016/J.ENERGY.2018.09.144>.

Jain, R., Smith, K., Culligan, P., & Taylor, J. (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Applied Energy, 123, 168-178. <https://doi.org/10.1016/J.APENERGY.2014.02.057>.

Dan, T., & Phuc, P. (2018). Application of Machine Learning in Forecasting Energy Usage of Building Design. 2018 4th International Conference on Green Technology and Sustainable Development (GTSD), 53-59. <https://doi.org/10.1109/GTSD.2018.8595595>.

Ribeiro, M., Grolinger, K., ElYamany, H., Higashino, W., & Capretz, M. (2018). Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. Energy and Buildings, 165, 352-363. <https://doi.org/10.1016/J.ENBUILD.2018.01.034>.

Shorfuzzaman, M., & Hossain, M. S. (2021). Predictive Analytics of Energy Usage by IoT-Based Smart Home Appliances for Green Urban Development. ACM Transactions on Internet Technology (TOIT), 22(2), 1-26. <https://dl.acm.org/doi/full/10.1145/3426970?casa_token=nRR3FfDg_IkAAAAA%3A35sTqxoqYyCGbGnMWz-wyW_Manzzndt5M9xM2wxwywycp8pi3QNi87dPSAjt5-E1LjomQlQe5z8d76w>

McGrane, M. (2021). The Prediction and Optimisation of Smart Energy Usage through Machine Learning Recommendations (Doctoral dissertation, Dublin, National College of Ireland). <https://norma.ncirl.ie/5189/1/markmcgrane.pdf>

S. Iram et al., "An Innovative Machine Learning Technique for the Prediction of Weather Based Smart Home Energy Consumption," in IEEE Access, vol. 11, pp. 76300-76320, 2023, doi: 10.1109/ACCESS.2023.3287145. <https://pure.hud.ac.uk/en/publications/an-innovative-machine-learning-technique-for-the-prediction-of-we>

Balaji, S., & Karthik, S. (2023, January). Comparative Study of Various Machine Learning and Deep Learning Techniques for Energy Prediction and Consumption Using IoT Modules. In Proceedings of the International Conference on Cognitive and Intelligent Computing: ICCIC 2021, Volume 2 (pp. 99-105). Singapore: Springer Nature Singapore. <https://link.springer.com/chapter/10.1007/978-981-19-2358-6_10>

Huda, N. U., Ahmed, I., Adnan, M., Ali, M., & Naeem, F. (2024). Experts and intelligent systems for smart homes’ Transformation to Sustainable Smart Cities: A comprehensive review. Expert Systems with Applications, 238, 122380. <https://www.researchgate.net/publication/375122032_Experts_and_intelligent_systems_for_smart_homes'_Transformation_to_Sustainable_Smart_Cities_A_comprehensive_review>

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. <https://www.nature.com/articles/nature14539>

Ahmad, T., Chen, H., Wang, J., & Guo, Y. (2014). Application of ANN and SVM for building energy prediction. Studies in Computational Intelligence, 485, 69-79. <https://www.researchgate.net/publication/260315410_A_review_on_applications_of_ANN_and_SVM_for_building_electrical_energy_consumption_forecasting>

*What is granularity in data analysis and why is it important?* (n.d.). <https://www.talon.one/glossary/granularity>

Kwak SK, Kim JH. Statistical data preparation: management of missing values and outliers. Korean J Anesthesiol. 2017 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5548942/>