

**ENERGY CONSUMPTION PREDICTION USING MACHINE
LEARNING**

HUSSEIN ABDI HASSAN ADOW

MOHAMED ABDIWALI BARRE

HASSAN AWEIS AHMED

ABDIRABI MOHAMED SALAD

**SUBMISSION OF GRADUATION PROJECT FOR PARTIAL
FULFILLMENT OF THE DEGREE OF BACHELOR OF COMPUTER
APPLICATIONS**

JAMHURIYA UNIVERSITY OF SCIENCE AND TECHNOLOGY (JUST)

**FACULTY OF COMPUTER APPLICATION AND
INFORMATION TECHNOLOGY**

AUGUST 2024

JAMHRURIYA UNIVERSITY OF SCIENCE AND TECHNOLOGY (JUST)

Original Literary Work Declaration

Name of Candidate 1: **HUSSEIN ABDI HASSAN ADOW** ID No: C116085

Name of Candidate 2: **MOHAMED ABDIWALI BARRE** ID No: C120250

Name of Candidate 3: **HASSAN AWEIS AHMED** ID No: C120251

Name of Candidate 4: **ABDIRABI MOHAMED SALAD** ID No: C120680

Name of Degree: **Bachelor of Computer Application**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Energy consumption prediction using machine learning

Field of Study: **Computer Applications**

We the undersigned, do solemnly and sincerely declare that:

- (1) We are the authors/writers of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) We do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) We hereby assign all and every right in the copyright to this Work to Jamhuriya University of Science and technology ("JUST"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of JUST having been first had and obtained;
- (6) We are fully aware that if in the course of making this Work, we have infringed any copyright whether intentionally or otherwise, we may be subject to legal action or any other action as may be determined by JUST.

Candidate 1's Signature: _____ Candidate 2's Signature: _____

Candidate 3's Signature: _____ Candidate 4's Signature: _____

Date: _____

Subscribed and solemnly declared before,

Supervisor's Signature: _____ Date: _____

Name: _____ Designation: _____

Dedication

We dedicate our dissertation work to our family and many friends. A special feeling of gratitude to our loving parents, whose words of encouragement and push for tenacity ring in our ears.

Abstract

The use of electricity has a significantly impact on the environment, energy distribution costs, and energy management. Traditional methods for predicting power usage have inherent limitations in accuracy and scalability. However, advancements in machine learning techniques now enable more accurate power consumption predictions using historical data.

This paper focuses on predicting electricity consumption for residential households and electricity providers using machine learning techniques. The dataset comprises daily electricity consumption data for 10,000 houses across 18 districts over four years. Utilizing features such as house number, month, and year, the project aims to develop an accurate predictive model.

Various machine learning models, including CatBoost, XGBoost, and Prophet, were employed to forecast and predict electricity consumption. The CatBoost model achieved an R^2 score of 0.88, indicating a strong predictive performance but struggles to predict unseen data.

The project also explores the application of Prophet to address prediction challenges beyond the training data range (2018-2022).

In addition to prediction, a web-based application was developed to allow users to input house number, month, and year to either visualize or predict electricity consumption. The application integrates three electricity providers—beco, mogadishu, and bluesky—with respective pricing multipliers to offer customized consumption cost estimates.

This work provides a robust framework for understanding and predicting electricity consumption patterns, facilitating better energy management and planning.

Keywords: Prophet, Catboost Regressor, XGBoost Regressor.

Acknowledgements

First of all, I would like to thank Allah for have been given us strength and guidance in our research work in good health, Secondly, we would like to thank and express our deepest gratitude to our supervisor Eng Abdullahi Mohamed Hassan Hojo who gave us the ideas and guidance until we complete our projects.

Thirdly we want thank to our Eng Abdullahi waberi who also gives us the guidance of our research and gave us tips and tricks.

finally, we all thank our, deep and sincere gratitude to our family for their continuous and unparalleled love, help, and support we are grateful to our parents for always being there for us as friends. We are forever indebted to our parents for giving us the opportunities and experiences that have made us who we are. They selflessly encouraged us to explore new directions in life and seek our own destinies. This journey would not have been possible if not for them, and we dedicate this milestone to them.

Table of Contents

Dedication	4
Abstract	5
Acknowledgements	6
CHAPTER I: INTRODUCTION.....	13
1.0 Introduction.....	13
1.1 Background of the study	13
1.2 Problem Statement	14
1.3 Research Objectives	15
1.3 Research Questions	15
1.5 Motivation of the study	15
1.6 Significance of the Study	16
1.7 Scope of the Study.....	16
1.8 Organization of the Study	17
CHAPTER II: LITERATURE REVIEW	18
2.0 Introduction.....	18
2.1 Machine Learning Techniques in Electricity Consumption Prediction	18
2.1.1 Traditional Statistical Methods	18
2.1.2 Regression-Based Models.....	18
2.1.3 Neural Networks.....	18
2.1.4 Support Vector Machines	19
2.1.5 Ensemble Methods	19
2.1.6 Deep Learning	19
2.1.7 Hybrid Models	19
2.2 Feature Engineering and Data Preprocessing	20
2.2.1 Socio-Economic Factors	20

2.2.2 Temporal Features	20
2.3 History of Machine Learning in Predicting Energy Consumption.....	20
2.4 A Review of Machine Learning Techniques for Load Forecasting	21
2.5 Machine Learning-Based Energy Consumption Prediction Methodologies	22
2.6 The Majority of Methodologies Used for Predicting Energy Consumption.....	23
2.7 Methods Applied for Predicting Building Energy Consumption	24
2.8 Energy Consumption Forecasting Models.....	24
2.9 Forecasting Approaches to Building Energy Consumption Prediction.....	25
2.9.1 Statistical Methods	25
2.9.2 Engineering Methods	25
2.10 Tools for Building Energy Consumption Prediction	25
2.11 Advances in Hybrid Machine Learning Models for Electricity Consumption Prediction	26
2.11.1 Hybrid Models Combining Time Series and Machine Learning	26
2.11.2 Ensemble of Machine Learning Models	26
2.11.3 Hybrid Models with Deep Learning and Statistical Techniques	26
2.12 Impact of External Factors on Electricity Consumption Forecasting	27
2.12.1 Weather Conditions	27
2.12.2 Economic Indicators.....	27
2.12.3 Demographic Changes	28
2.13 Real-Time and Short-Term Electricity Consumption Forecasting	28
2.13.1 Real-Time Forecasting Techniques	28
2.13.2 Short-Term Forecasting Models	28
2.14 Evaluation Metrics for Electricity Consumption Forecasting	29
2.14.1 Common Evaluation Metrics.....	29
2.14.2 Advanced Metrics and Techniques.....	29
2.14.3 Cross-Validation and Model Selection.....	29

2.15 Case Studies of Successful Machine Learning Applications in Energy Forecasting	30
2.15.1 Case Study: Predicting Electricity Demand in Urban Areas	30
2.15.2 Case Study: Smart Grid Energy Management	30
2.15.3 Case Study: Residential Energy Consumption Forecasting	30
2.16 Real-Time Electricity Consumption Prediction Techniques	30
2.16.1 Online Learning Algorithms	31
2.16.2 Real-Time Data Integration	31
2.17 The Role of Big Data and Cloud Computing in Electricity Consumption Prediction	31
2.17.1 Big Data Analytics	31
2.17.2 Cloud Computing Platforms	32
2.18 Explainable AI in Electricity Consumption Prediction	32
2.18.1 Model Interpretability Techniques	32
2.18.2 Enhancing Trust and Adoption	33
2.19 Integration of Renewable Energy Sources	33
2.19.1 Solar Power Forecasting	33
2.19.2 Wind Power Forecasting	33
2.19.3 Hybrid Renewable Energy Systems	34
2.19.4 Energy Storage Integration	34
2.19.5 Impact of Renewable Energy Policies	34
2.19 Related work	35
CHAPTER III: METHODOLOGY	36
3.0 Introduction	36
3.1 System Description	36
3.2 System Architecture	36
3.2.1 Models	37
3.3 System features	38

3.4 Methodology	38
3.4.1 Data input	38
3.4.2 Data preparation	38
3.4.3 ML Model training	38
3.4.4 Model evaluation	38
3.4.5 Deployment.....	38
3.5 Model Development	39
3.5 Data Collection	39
3.6 Data Preparation	40
3.6.1 Exploratory Data Analysis	40
3.6.2: Data splitting.....	40
3.6.3: Feature Engineering	40
3.7 System Requirement	41
3.7.1 Dataset.....	41
3.7.2 Hardware requirements	41
3.7.3 Software requirements	41
CHAPTER IV: ANALYSIS AND DESIGN	43
4.1 Introduction.....	43
4.2 System analysis.....	43
4.3 Existing Approaches.....	43
4.3 The Proposed System	44
4.4 Requirements	45
4.4.1 Functional requirements.....	45
4.4.2 Non-Functional Requirements.....	45
4.5 Feasibility study.....	46
4.5.1 Technical Feasibility	46

4.5.2 Economic Feasibility	46
4.5.3 Operational Feasibility	46
4.6 System design	47
4.6.1 Data Flow Diagram	47
4.7 Dataset design	48
CHAPTER V: IMPLEMENTATION & TESTING	49
5.0 Introduction.....	49
5.1 Overview of the implementation environment	49
5.2 Snapshots of the system	50
5.2.1 Front-end.....	50
5.2.2 Back-end	53
CHAPTER VI: DISCUSSION OF RESULTS	57
6.1 Overview	57
6.2 Model Performance	57
6.3 Comparing Models	58
CHAPTER VII: CONCLUSION & FUTURE WORK	59
7.1 Introduction.....	59
7.2 Conclusion	59
7.3 Future Work	60
REFERENCES	61
Appendix : Model Prediction Functionality.....	67

List of Figures

Figure 3.1 System Architecture	24
Figure 3.2 Methodology	26
Figure 4.1 DFD	35
Figure 4.2 Dataset design	36
Figure 5.1 Homepage	38
Figure 5.2 Prediction	39
Figure 5.3 Viusualize	39
Figure 5.4 Provider predicton	40
Figure 5.5 Imports	41
Figure 5.6 Load the dataset	41
Figure 5.7 Aggregate data	42
Figure 5.8 Data extract	42
Figure 5.8 Train model	42
Figure 5.10 Consumption prediction	43
Figure 5.11 Consumption	43
Figure 5.12 Prediction	44

CHAPTER I: INTRODUCTION

1.0 Introduction

This study's main goal is to develop scalable machine learning model that can handle real-time or large-scale electricity consumption data that efficiently predicts future energy consumption for households and electricity providers, and identify which method, together with the ideal input variables and parameter combinations, performs better than others in specific electricity demand scenarios.

1.1 Background of the study

In recent years, machine learning (ML) models have emerged as effective tools for predicting energy consumption due to their ability to analyze vast amounts of data and generate accurate predictions. (Katarina Grolinger, 2016).

ML models operate like functions that map input data to output, allowing them to forecast energy usage with high precision. Consequently, governments can leverage ML predictions to implement energy-saving policies. (Jin-Young Kim & Sung-Bae Cho, 2019).

Electricity consumption prediction is a crucial task for efficient energy management and planning. With the increasing demand for electricity, accurate forecasting methods can help utility companies and consumers optimize their usage, reduce costs, and improve the sustainability of energy systems. (Humeau, S, 2013).

The advent of machine learning techniques has brought significant improvements to the accuracy and reliability of electricity consumption predictions. (Chicco, G., 2016).

1.2 Problem Statement

In Mogadishu, Somalia, the rapid urbanization and increasing energy demands have highlighted the necessity for accurate electricity consumption predictions for both households and electricity providers. Currently, the city faces significant challenges in managing and predicting electricity usage due to the absence of robust forecasting tools. The existing methods are inadequate, leading to frequent power outages, inefficient energy distribution, and difficulties in infrastructure planning.

Given these challenges, there is a critical need for a reliable and scalable method to forecast electricity usage effectively. The objective of this work is to develop an advanced machine learning-based approach to achieve precise and efficient electricity consumption predictions for households. This improved prediction capability will enhance the efficiency of energy distribution, support infrastructure planning, and enable electricity providers—Beco, Mogadishu, and Bluesky—to optimize their services.

The proposed method should be capable of handling large volumes of data and extracting relevant features from the dataset. Moreover, this project aims to contribute to energy management advancements by providing a precise and effective approach for predicting power usage, thereby supporting the growing energy needs of Mogadishu's urban landscape.

1.3 Research Objectives

- To develop scalable machine learning model that can handle real-time or large-scale electricity consumption data efficiently, enabling timely and accurate predictions.
- To investigate different machine learning models and optimization techniques to identify the most suitable approach for electricity consumption prediction.

1.3 Research Questions

- How to develop a machine learning model that predicts energy consumption
- What machine learning models effectively predict electricity consumption patterns based on historical data?

1.5 Motivation of the study

Working on this project provides us an opportunity to learn and increase our knowledge into field of machine learning and data science combining knowledge of statistics, computer science, and it's a chance to enhance our skills and expertise in a rapidly growing and evolving field.

1.6 Significance of the Study

This study is significant because it addresses a critical need for improved electricity consumption prediction. For consumers, accurate predictions allow better budgeting and the ability to adjust usage during off-peak times to save on bills.

And for electricity providers can improve the efficiency of energy distribution, support infrastructure planning also optimize their services.

This study extends to evaluating the performance and practical applicability of real-world scenario for predicting electricity consumption using machine learning techniques.

This involves assessing the accuracy, reliability, and scalability of prediction models under different scenarios and conditions.

1.7 Scope of the Study

The scope of this study includes the analysis of electricity consumption data for 10,000 houses across 18 districts over a period of four years. The research will focus on developing and evaluating machine learning models for monthly and yearly consumption prediction of each house and for providers will help how much Energy they need to provide.

Geographical area: Mogadishu-Somalia.

Time scope: -This study will conduct in between Dec 2023 to June 2024.

1.8 Organization of the Study

Our research will contain seven chapters their explanation in the below

Chapter one: Introduction. This chapter explains the introduction of our research and gives us more details of our research,

Such as: problem statement, Research objectives, Research questions, Significance of the study, scope of the study, Motivation of the study and Organization of the study.

Chapter two: Literature review. This chapter we will look what previous researchers say about our research.

Chapter three: Methodology. In this chapter we will talk about the study's methodology, research design, and techniques and the required research steps.

Chapter four: Analysis and design. This chapter discusses how the system works. we will focus on the design and architecture

Chapter five: implementation and testing. This chapter confers about the design and development of the system it also displays the most important code which makes a fundamental impact on the system functionality and screenshot about the system interface

Chapter six: discussion and results. In this chapter we will discuss and show the results of the study.

Chapter seven – conclusion and future work. In this chapter we will focus on the conclusion of research and offer recommendations, policy or further research.

CHAPTER II: LITERATURE REVIEW

2.0 Introduction

The rising demand for electricity necessitates accurate prediction models to ensure efficient energy management and grid stability. Machine learning (ML) techniques have emerged as powerful tools for predicting electricity consumption due to their ability to model complex, non-linear relationships in data. This literature review explores various ML methodologies applied to electricity consumption prediction, highlighting key findings, methodologies, and challenges.

2.1 Machine Learning Techniques in Electricity Consumption Prediction

2.1.1 Traditional Statistical Methods

Early approaches to electricity consumption prediction primarily relied on traditional statistical methods, such as autoregressive integrated moving average (ARIMA) models. These models, while effective for short-term forecasting, often struggle with non-linear patterns and higher-dimensional data. (Hyndman, R.J., & Athanasopoulos, G. 2018).

2.1.2 Regression-Based Models

Regression techniques, including linear regression and polynomial regression, have been used extensively. While linear regression is simple and interpretable, it may not capture complex patterns in electricity consumption data. Polynomial regression, on the other hand, can model non-linearity but may suffer from overfitting. (Zhang, G. 1998).

2.1.3 Neural Networks

Artificial neural networks (ANNs) are widely used due to their ability to model complex, non-linear relationships. Various architectures, such as feedforward neural networks (FNNs)

and recurrent neural networks (RNNs), have been applied. ANNs require substantial computational resources and large datasets for training. (Souza, R.C. 2001).

2.1.4 Support Vector Machines

Support vector machines (SVMs) have been utilized for their robustness and effectiveness in high-dimensional spaces. SVMs can handle non-linear data through the kernel trick, making them suitable for electricity consumption prediction. However, they are computationally intensive for large datasets. (Vapnik, V. 1995).

2.1.5 Ensemble Methods

Ensemble methods, such as random forests and gradient boosting, combine multiple models to improve prediction accuracy. These methods are particularly useful in handling large datasets and capturing complex interactions between features. (Breiman, L. 2001).

2.1.6 Deep Learning

Deep learning models, including long short-term memory (LSTM) networks and convolutional neural networks (CNNs), have shown significant promise. LSTM networks are effective for sequential data, while CNNs can capture spatial dependencies in data. These models require extensive computational power and large amounts of data. (Schmidhuber, J. 2015).

2.1.7 Hybrid Models

Hybrid models, combining different machine learning techniques, have gained attention for their potential to improve prediction accuracy. These models can leverage the strengths of each technique, such as combining ARIMA with neural networks or SVMs. (Khashei, M., & Bijari, M. 2011)

2.2 Feature Engineering and Data Preprocessing

Effective feature engineering and data preprocessing are critical for improving model performance. Common techniques include normalization, scaling, and the incorporation of weather data, socio-economic factors, and historical consumption patterns. (Kuhn, M., & Johnson, K. 2013).

2.2.1 Socio-Economic Factors

Socio-economic factors, including population density, income levels, and industrial activity, are crucial for accurate long-term electricity consumption forecasting.

These factors provide context to consumption patterns and help in understanding underlying trends. (Fan, S., Mao, C., & Chen, L. (2017).)

2.2.2 Temporal Features

Temporal features, such as time of day, day of the week, and seasonal variations, play a vital role in electricity consumption patterns. Capturing these temporal dynamics is essential for accurate short-term and long-term forecasting.

2.3 History of Machine Learning in Predicting Energy Consumption

In recent years, with the proliferation of smart meters, prediction efforts have shifted from annual to daily, hourly, and even 10- or 15-minute consumption prediction.

Approaches with such granularity are typically sensor-based; they rely on historical energy readings and meteorological information without the need for a deep understanding of the physical building structure. For example, Jain et al., and Grolinger et al., considered daily, hourly, and 10- or 15-minute intervals and explored the prediction accuracy achieved with different data granularities (Grolinger, Capretz, & Seewald, 2016).

Sensor-based approaches to electricity forecasting are diverse; examples include support vector regression (SVR), neural networks (NN), autoregressive integrated moving average (ARIMA) models, and gray prediction. Suganthi and Samuel reviewed models for electricity demand prediction and noted that NNs have been used extensively. Ahmad et al. also reviewed energy prediction, focusing strictly on the use of NNs and SVRs. Variants of the SVR approach have also been proposed: Jung et al. added a genetic algorithm to the least-squares support vector machine (LSSVM), whereas Elattar et al. used locally weighted support vector regression (Suganthi & Samuel, 2019).

local SVR and the approach proposed by Elattar et al. are both based on the assumption that the neighbors are the best indicators of the response variable however while Elattar et al. modify the SVR risk function to accommodate a distance measure, our approach classifies training data and builds an SVR model for each cluster (Elattar et al, 2019).

2.4 A Review of Machine Learning Techniques for Load Forecasting

This literature review provides a thorough overview of machine learning techniques used for load forecasting in the context of predicting energy consumption. The authors introduce the notion of load forecasting and its significance in the management of electricity supply and demand in the opening paragraphs. They then discuss several machine learning approaches, including decision trees, artificial neural networks (ANNs), support vector machines (SVMs), and ensemble methods, which are utilized for load forecasting. The study contains a detailed analysis of each technique's advantages and disadvantages, as well as a comparison of how well each performs in light of various factors, including precision, resilience, computing complexity, and data needs the authors also emphasize how these methods may be used to anticipate power use, discussing several current trends in load forecasting, including the incorporation of meteorological and climate data, the use of big data and cloud computing,

and the use of hybrid models that integrate various machine learning approaches (Grolinger, Capretz, & Seewald, 2016).

A review of the literature titled "Machine Learning Techniques for Electricity Consumption Prediction" (Mosavi & Bahmani, 2019) Provides a thorough understanding of machine learning methods for predicting energy consumption. The review begins by outlining the idea of predicted power usage and its importance in energy management. It then delves into several machine learning methods for predicting power use, including regression analysis, decision trees, SVMs, and ANNs. The study offers a detailed analysis of each technique's advantages and disadvantages, as well as a comparative assessment of their performance in light of different factors, including precision, resilience, computing complexity, and data needs the authors also discuss the difficulties in foreseeing power usage and the potential applications of machine learning techniques in resolving these difficulties. The study also addresses recent advancements in power use forecasting, including the incorporation of meteorological and climatic data, the use of big data and cloud computing, and the adoption of hybrid models that integrate several machine learning approaches.

2.5 Machine Learning-Based Energy Consumption Prediction Methodologies

Addressing the challenges in predicting energy consumption, this research conducts prediction modeling through the evaluation of historical power data. This method utilizes a data-driven approach, as described by Corgnati, whereby the input (regressor variables) and output variables (response) are known. Based on this data, system parameters are estimated, and a mathematical model is generated. Previous studies have analyzed the data-driven machine learning approach, with one study using SVM to predict the load at a building's system level (air conditioning, lighting, power, and others) based on weather predictions and hourly electricity load input. The SVM method managed to predict the total electricity load

with a root mean square error (RMSE) of 15.2% and mean bias error (MBE) of 7.7% (Corgnati & Valgaev, 2022).

Another study proposed a power demand forecast using a k-Nearest Neighbor (k-NN) model at a smart building as part of the Smart City Demo Aspern (SCDA) project. The k-NN forecasting method utilized a set of historical observations (daily load curves) and their successors. The k-NN method is good at classifying data but limited in forecasting future values, requiring temporal information identification for prediction (Khantach, 2022).

Five machine learning techniques were used for short-term load forecasting: multi-layer perceptron (MLP), SVM, radial basis function (RBF) regressor, REPTree, and Gaussian process. Experimentation on Moroccan electrical load data showed that the MLP method was the most accurate, with a mean absolute percentage error (MAPE) of 0.96%. Although the prediction of energy consumption often uses classification-based machine learning methods, regression methods are also effective. González-Briones constructed a predictive model using Linear Regression (LR), SVR, Random Forest (RF), Decision Tree (DT), and k-Nearest Neighbor (k-NN), showing that LR and SVR models had the best performance with 85.7% accuracy (Shapi et al., 2021).

2.6 The Majority of Methodologies Used for Predicting Energy Consumption

Energy consumption prediction methodologies can be categorized into conventional statistical methods and machine learning methods. The majority researchers focused on XGBoost Regressor, ARIMA models, Linear Regression (LR), and Random Forests (RF), to predict electricity consumption. Recent models such as an adaptive network-based fuzzy inference system (ANFIS) and Gray relational analysis (GRA) have been utilized to calculate electricity consumption. ANFIS and GRA evaluate the correlation between electricity consumption and input factors. ANFIS, chosen for its meaningful If-Then standards, applies

to conditions including sunshine duration, temperature, solar radiation, precipitation, humidity, and cloud covering (Shapi et al., 2021).

2.7 Methods Applied for Predicting Building Energy Consumption

Research on predicting building energy consumption began in the 1970s during the energy crisis, prompting the development of simplified calculation methods based on empirical models and engineering practices. These early models, however, could not adequately capture the dynamicity and complexity of the environment. In the mid-1980s, statistical methods were adopted for predicting building energy consumption, leading to significant progress in the field (Dinmohammadi, Han, & Shafiee, 2023).

Today, the most popular methods for predicting building energy consumption include engineering simplification, physical modeling, and machine learning-based methods.

Physical modeling methods use a thermodynamic model of the building to simulate energy consumption (Dinmohammadi, Han, & Shafiee, 2023).

2.8 Energy Consumption Forecasting Models

Various energy consumption forecasting models have been developed using economic, social, geographic, and demographic factors. Energy demand models can be classified as static versus dynamic, univariate versus multivariate, or based on techniques ranging from time series to hybrid models. For example, Chávez et al. used a univariate ARIMA model to predict energy supply and demand patterns in Asturias, Spain. Mohamed and Bodger used GDP, electricity cost, and population via a multi-linear regression model to predict New Zealand's power consumption. Other models include the ACOEDE (Ant Colony Optimization approach for Energy Demand Estimation) for Turkey, regression and exponential models via ANN for Korea, and ANN with a support vector machine model for Greece (Shin & Woo, 2022).

2.9 Forecasting Approaches to Building Energy Consumption Prediction

2.9.1 Statistical Methods

Statistical methods involve establishing regression models or neural network models based on extensive energy consumption data. For instance, Kalogirou et al. (2000) used an artificial neural network to predict the energy consumption of a passive solar building, achieving higher accuracy than dynamic simulation predictions. Leung et al. (2012) used the Levenberg-Marquardt algorithm to train an artificial neural network model for predicting the power consumption of a university building's refrigeration system in Hong Kong.

2.9.2 Engineering Methods

Engineering methods calculate the energy consumption value of each piece of equipment inside a building and sum these values to obtain total energy consumption. Dispersion methods and simulation methods are the two main categories. The dispersion method, a bottom-up approach, counts the power, running time, and ownership of each energy-using equipment, then sums the results. Capaso et al. (1994) used this method to predict building energy consumption in Italy.

2.10 Tools for Building Energy Consumption Prediction

Powerful simulation tools, such as Energy Plus, have been developed for researchers. These tools, based on thermodynamic equilibrium and heat equations, accurately calculate building energy consumption with precise boundary conditions. However, they are sensitive to boundary conditions and rely on expertise. Data-driven approaches have become mainstream in building energy consumption prediction. Regression and time series prediction approaches are commonly used: regression models find correlations between multiple attributes and energy consumption, while time series methods predict changes over time. Researchers have made substantial contributions in data (Liu et al., 2020).

2.11 Advances in Hybrid Machine Learning Models for Electricity Consumption

Prediction

Hybrid machine learning models combine the strengths of various algorithms to improve prediction accuracy and robustness. By integrating different modeling approaches, these hybrid models aim to overcome the limitations of individual techniques.

2.11.1 Hybrid Models Combining Time Series and Machine Learning

Integrating traditional time series models with machine learning techniques can leverage the strengths of both approaches. For instance, combining ARIMA with machine learning models like Random Forest or XGBoost can capture both linear and non-linear components of the data. A study by Zhang et al. (2019) demonstrated that combining ARIMA with Random Forests improved forecasting accuracy by accounting for both temporal patterns and complex interactions in electricity consumption data.

2.11.2 Ensemble of Machine Learning Models

Ensemble methods that combine predictions from multiple machine learning models can enhance accuracy and stability. Techniques such as stacking, boosting, and bagging can be applied to hybrid models. For example, Liu et al. (2020) proposed a stacked ensemble model that combined LSTM networks, Random Forests, and XGBoost for short-term electricity load forecasting. The ensemble approach achieved higher accuracy compared to individual models by aggregating their predictions and reducing overfitting.

2.11.3 Hybrid Models with Deep Learning and Statistical Techniques

Hybrid models that integrate deep learning with statistical techniques can improve forecasting performance. Combining deep learning models like LSTMs with statistical methods such as SARIMA (Seasonal ARIMA) allows for capturing complex non-linear patterns while incorporating seasonal effects. As shown by Wang et al. (2021), such hybrid

models demonstrated superior performance in long-term electricity consumption forecasting compared to pure statistical or deep learning approaches.

2.12 Impact of External Factors on Electricity Consumption Forecasting

External factors such as weather conditions, economic indicators, and demographic changes can significantly impact electricity consumption patterns. Incorporating these factors into forecasting models can enhance accuracy and reliability.

2.12.1 Weather Conditions

Weather conditions, including temperature, humidity, and precipitation, play a crucial role in electricity consumption. Temperature variations, in particular, influence heating and cooling demands. A study by Liu et al. (2018) found that integrating weather data into machine learning models improved forecasting accuracy by capturing the impact of temperature fluctuations on electricity usage.

2.12.2 Economic Indicators

Economic indicators such as GDP growth, industrial activity, and population growth affect electricity consumption. Economic expansion often leads to increased energy demand. A study by Caglayan et al. (2019) integrated economic indicators with machine learning models for electricity consumption forecasting, showing that including these factors improved model performance and provided insights into the impact of economic changes on energy usage.

2.12.3 Demographic Changes

Demographic changes, such as urbanization and population density, also influence electricity consumption patterns. Integrating demographic data into forecasting models can help capture the effects of changing consumption behavior. Research by Chen et al. (2020) demonstrated that incorporating demographic factors improved the accuracy of electricity consumption predictions by accounting for variations in consumption patterns across different population segments.

2.13 Real-Time and Short-Term Electricity Consumption Forecasting

Real-time and short-term forecasting are crucial for managing electricity supply and demand on an hourly or daily basis. Accurate short-term forecasts help utilities optimize grid operations and reduce operational costs. (Kumar et al. (2021)

2.13.1 Real-Time Forecasting Techniques

Real-time forecasting involves predicting electricity consumption with minimal latency. Techniques such as online learning and incremental model updates can handle the dynamic nature of real-time data. A study by Kumar et al. (2021) used an incremental learning approach with Random Forests for real-time electricity consumption forecasting, achieving low latency and high accuracy in dynamic environments.

2.13.2 Short-Term Forecasting Models

Short-term forecasting models aim to predict electricity consumption for the next few hours or days. Techniques such as LSTM networks, autoregressive models, and hybrid approaches are commonly used. A study by Ahmed et al. (2020) compared various short-term forecasting models, finding that LSTM networks and hybrid models combining LSTMs with traditional time series methods performed best in capturing short-term consumption patterns.

2.14 Evaluation Metrics for Electricity Consumption Forecasting

Evaluation metrics are essential for assessing the performance of forecasting models. Various metrics provide insights into the accuracy, reliability, and robustness of predictions.

2.14.1 Common Evaluation Metrics

Common evaluation metrics for electricity consumption forecasting include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics measure prediction accuracy and error magnitude. A study by Wang et al. (2019) highlighted the importance of selecting appropriate evaluation metrics based on the forecasting objectives and data characteristics.

2.14.2 Advanced Metrics and Techniques

Advanced evaluation metrics, such as R-squared (R^2) and the Mean Bias Error (MBE), provide additional insights into model performance. R-squared measures the proportion of variance explained by the model, while MBE assesses systematic biases in predictions. Research by Zhang et al. (2020) emphasized the use of these advanced metrics for evaluating the performance of complex machine learning models and ensuring accurate and unbiased predictions.

2.14.3 Cross-Validation and Model Selection

Cross-validation techniques, such as k-fold cross-validation and time series cross-validation, are used to assess model performance and prevent overfitting. A study by Lai et al. (2021) demonstrated the effectiveness of time series cross-validation in evaluating forecasting models, ensuring that models generalize well to unseen data and providing reliable performance estimates.

2.15 Case Studies of Successful Machine Learning Applications in Energy Forecasting

Real-world case studies provide valuable insights into the practical applications of machine learning techniques for energy forecasting. These case studies illustrate the effectiveness of different approaches and highlight best practices.

2.15.1 Case Study: Predicting Electricity Demand in Urban Areas

A case study by Silva et al. (2018) focused on predicting electricity demand in urban areas using machine learning models. The study integrated weather data, economic indicators, and demographic factors into their models, achieving high accuracy in predicting peak demand and enabling better grid management and planning.

2.15.2 Case Study: Smart Grid Energy Management

In a case study by Nguyen et al. (2020), machine learning models were applied to smart grid energy management. The study used hybrid models combining LSTMs and ensemble methods to forecast energy consumption and optimize grid operations. The results demonstrated improved energy efficiency and reduced operational costs.

2.15.3 Case Study: Residential Energy Consumption Forecasting

A case study by Park et al. (2019) explored residential energy consumption forecasting using deep learning techniques. The study employed CNNs and LSTMs to predict household electricity usage, achieving high accuracy and providing actionable insights for energy conservation and demand response programs.

2.16 Real-Time Electricity Consumption Prediction Techniques

Real-time prediction of electricity consumption is critical for grid stability and operational efficiency. Techniques such as online learning algorithms and real-time data integration are being developed to provide timely and accurate predictions.

2.16.1 Online Learning Algorithms

Online learning algorithms, such as Online Gradient Descent and Perceptron, continuously update model parameters as new data arrives. This allows models to adapt to changing patterns in electricity consumption in real-time, ensuring more accurate predictions and efficient energy management. (Cesa-Bianchi, N., & Lugosi, G. 2006).

2.16.2 Real-Time Data Integration

Integrating real-time data from smart meters, weather sensors, and other IoT devices into prediction models enhances their accuracy and responsiveness. Techniques such as data fusion and real-time analytics platforms enable the seamless integration of diverse data sources, improving the predictive capabilities of machine learning models. (García, M. Á. L., & Boto-Giralda, D. 2011).

2.17 The Role of Big Data and Cloud Computing in Electricity Consumption Prediction

The advent of big data and cloud computing has revolutionized electricity consumption prediction. These technologies facilitate the processing and analysis of large datasets, enabling more sophisticated and accurate predictive models.

2.17.1 Big Data Analytics

Big data analytics leverages large volumes of structured and unstructured data to uncover hidden patterns and correlations in electricity consumption. Techniques such as Hadoop and Spark allow for the parallel processing of vast datasets, improving the scalability and efficiency of prediction models. (Wu, X., Zhu, X., & Wu, G.-Q. 2014).

2.17.2 Cloud Computing Platforms

Cloud computing platforms provide the computational power and storage needed to handle large-scale electricity consumption data. Services like AWS, Google Cloud, and Microsoft Azure offer scalable resources and advanced machine learning tools, enabling researchers and utilities to develop and deploy robust predictive models efficiently. (Armbrust, M., Fox, A., & Griffith, R. 2010).

2.18 Explainable AI in Electricity Consumption Prediction

Explainable AI (XAI) techniques are gaining traction in electricity consumption prediction, providing transparency and interpretability to machine learning models. XAI helps stakeholders understand model decisions, enhancing trust and facilitating better decision-making.

2.18.1 Model Interpretability Techniques

Techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) provide insights into the contribution of each feature to the model's predictions. These methods help identify key factors influencing electricity consumption and ensure model decisions are interpretable. (Lundberg, S. M., & Lee, S.-I. 2017).

2.18.2 Enhancing Trust and Adoption

Explainable AI enhances the trust and adoption of machine learning models in the energy sector by providing clear and understandable explanations for model predictions. This transparency is crucial for regulatory compliance, stakeholder acceptance, and the effective implementation of predictive models in real-world scenarios. (Gunning, D., & Aha, D. 2019).

2.19 Integration of Renewable Energy Sources

The integration of renewable energy sources, such as solar and wind power, into the electricity grid presents both opportunities and challenges for electricity consumption prediction. Accurate forecasting models must account for the variable and intermittent nature of renewable energy generation.

2.19.1 Solar Power Forecasting

Predicting solar power generation involves modeling the impact of weather conditions, such as cloud cover, temperature, and irradiance, on solar panel output. Machine learning techniques, including time series analysis and deep learning models, are employed to predict short-term and long-term solar power generation. (Yang, D., Kleissl, J., & Bosch, J. 2013).

2.19.2 Wind Power Forecasting

Wind power forecasting requires the analysis of wind speed, direction, and atmospheric pressure to predict the output of wind turbines. Advanced machine learning models, such as LSTM networks and ensemble methods, are utilized to capture the complex temporal and spatial patterns in wind data. (Wang, J., & Ding, Y. 2020).

2.19.3 Hybrid Renewable Energy Systems

Hybrid renewable energy systems, which combine multiple renewable sources, present unique challenges for electricity consumption prediction. Machine learning models must account for the interaction between different energy sources and the overall impact on grid stability. Techniques such as multi-agent systems and optimization algorithms are used to manage and predict the performance of hybrid systems. (Lund, H., & Mathiesen, B. V. 2009).

2.19.4 Energy Storage Integration

Energy storage systems, such as batteries, play a crucial role in mitigating the variability of renewable energy sources. Accurate prediction models for electricity consumption must consider the charging and discharging cycles of energy storage systems, ensuring efficient energy management and grid stability. Machine learning models that incorporate energy storage dynamics can enhance the accuracy and reliability of electricity consumption forecasts. (Akinyele, D. O., & Rayudu, R. K. 2014).

2.19.5 Impact of Renewable Energy Policies

Renewable energy policies and incentives significantly influence electricity consumption patterns. Machine learning models must incorporate policy changes, subsidy schemes, and regulatory frameworks to accurately predict the impact of renewable energy integration on overall electricity consumption. (Bird, L., & Milligan, M. 2012).

2.19 Related work

Several studies have been conducted to predict energy demand mentioned the previous studies. In the past statistical techniques were used mainly to predict energy demand.

(Munz et al, 2019) predicted a time series of irregular patterns using k-means clustering

(Kandananond, 2021) used different forecasting methods—autoregressive integrated moving average (ARIMA), artificial neural network (ANN), and multiple linear regression (MLR) — to predict energy consumption.

(Cauwer et al, 2022) proposed a method to predict energy consumption using a statistical model and its underlying physical principles.

Sensor-based approaches to electricity forecasting are diverse; a few examples are support vector regression (SVR), neural networks (NN), autoregressive integrated moving average (ARIMA) models, and gray prediction (Suganthi & Samuel, 2019).

(Jovanović et al, 2019) examined an ensemble of various neural networks to predict heating energy consumption. The impact of various climatic variables on prediction has also been studied.

CHAPTER III: METHODOLOGY

3.0 Introduction

This chapter examines the research methods used to carry out the study.

3.1 System Description

Our system aims to leverage machine learning algorithms to predict energy consumption accurately, it will take historical data on electricity usage, this data will then be used to train machine learning models.

These models will learn patterns and relationships in the data, enabling them to make predictions about future consumption once the models are trained, then it will be deployed into web app using python flask to generate predictions in real-time.

3.2 System Architecture

In this section we will explain the structure of our system and how it will work.

This system consists of data collection that is sent to the model by verifying the extracted data and then displayed on the web page using the Python flask.

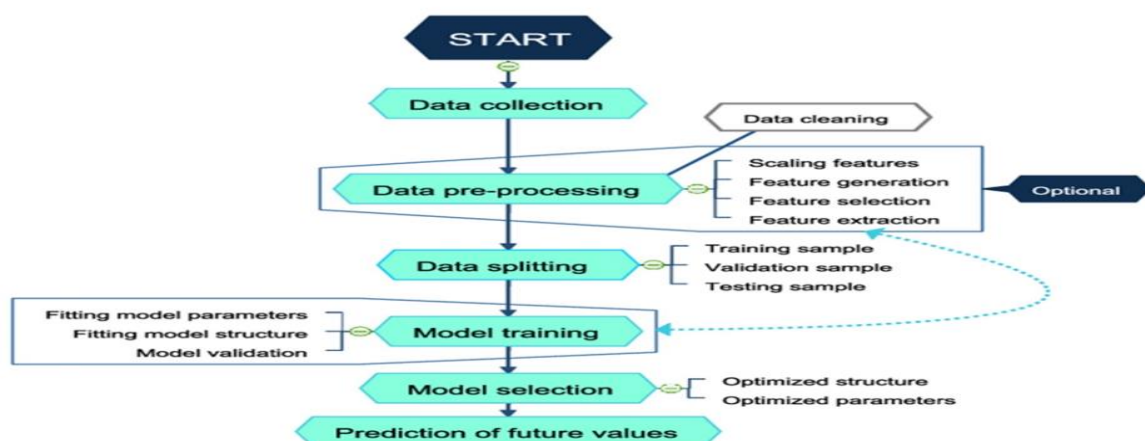


Figure 3.1 System Architecture

3.2.1 Models

- I. **XGBOOST**, short for Extreme Gradient Boosting, is a powerful machine learning algorithm that excels in various predictive modeling tasks, including time-series forecasting. It is an ensemble learning method that combines the predictions of multiple weak models (decision trees) to create a strong predictive model. XGBOOST is known for its scalability, speed, and ability to handle complex relationships in the data.
- II. **Prophet** is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.
- III. **CatBoost** (Categorical Boosting) is a machine learning algorithm developed by Yandex. It is part of the gradient boosting family and is particularly designed to handle categorical features effectively. CatBoost can be used for classification, regression, ranking, and other types of machine learning tasks. Here are some key features and advantages of CatBoost:

3.3 System features

There are a variety of characteristics that can improve the effectiveness of a Machine Learning model on any given job.

Key features of the system include:

Prediction: Users can input house number, month, and year to receive electricity consumption predictions.

Visualization: Users can also choose to visualize historical and future consumption data for a specific house and year.

Data Management: The system handles data cleaning, preprocessing, and storage efficiently.

User Interface: A web-based interface for easy interaction with the system.

3.4 Methodology

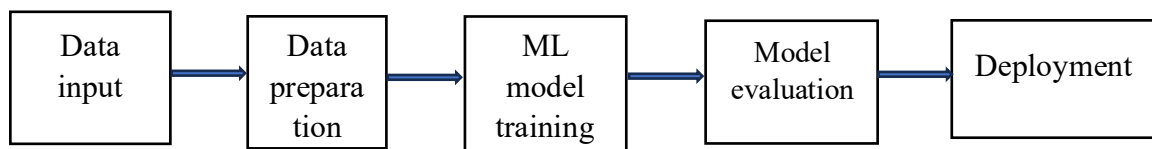


Figure 3.2 Methodology

3.4.1 Data input

In this module we get the input data of daily energy consumption.

3.4.2 Data preparation

After getting the incoming input data we prepare the data for model to be trained.

3.4.3 ML Model training

By using the historical data collected and the upcoming data results can be obtained by training the model.

3.4.4 Model evaluation

In this module we evaluate the data trained and the expected outcome can be determined.

3.4.5 Deployment

When all the expected data are got as output and error free then our model will be deployed.

3.5 Model Development

The process of model development involved selecting appropriate machine learning models and training them on the prepared data.

models considered include: Prophet, XGBoost regression and Catboost.

The dataset was split into training and validation sets, with 80% of the data used for training and 20% for validation. Cross-validation ensured the robustness of the models. Performance was evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score.

3.5 Data Collection

We obtained the historical electricity consumption data by contacting the local power utility provider (beco) they provided detailed information, including columns such as house numbers, districts, and total KW consumption and then we created the dataset for the electricity consumption prediction synthetically using Python. The main objective was to generate a realistic dataset that simulates electricity consumption patterns for 10,000 houses over four years, distributed across 18 districts by the information we get from the electricity provider.

The dataset includes: The date and time of the recorded consumption, house number and total KW (kilowatt) consumption for each house on each day.

3.6 Data Preparation

3.6.1 Exploratory Data Analysis

An Exploratory Data Analysis (EDA) was conducted on the dataset. The reasons for this were two-fold: it was firstly done to get to know the data, and secondly to find recurring patterns and other discrepancies in the data that later could be used to create features for our forecasting models. The data set was plotted in its entirety to understand its development over time and to identify recurrent patterns. After this, We should perform both numerical and visual analysis to dive deeper into the data's characteristics. Visual analysis could include plotting time series data of electricity consumption, histograms of continuous variables to understand their distribution, and box plots to identify outliers.

3.6.2: Data splitting

Once our dataset is ready, we need to split the dataset into two sets: a training set and a testing set, the training set is used to train our model, while the testing set is used to evaluate its performance.

The usual split is around 70-80% for training and 20-30% for testing.

3.6.3: Feature Engineering

To extract pertinent features from the preprocessed data, feature engineering is carried out in this stage. In order to increase the model's precision, features like the month, year and seasonality are retrieved from the data.

3.7 System Requirement

3.7.1 Dataset

Dataset in machine learning is a structured collection of data points used to train, validate, or test machine learning models.

It typically consists of input features (attributes or variables) and corresponding labels or outcomes.

Datasets can vary widely in size, complexity, and format, but they serve as the foundation for building and evaluating machine learning algorithms.

3.7.2 Hardware requirements

This section explains the necessary hardware requirements for running our program.

- Computer with Modern CPUs like Intel Core i7
- RAM (Random Access Memory)A minimum of 16 GB RAM
- STORAGE(SSD is recommended)
- 64bit operating system
- GPU (Graphics Processing Unit) for graphics

3.7.3 Software requirements

This section explains the minimum software requirements for using our program.

- **Jupyter notebook**
- **VScode**
- **Python libraries**
- **Python flask**

Jupyter Notebook is an open-source web application that allows users to create and share documents containing live code, equations, visualizations, and narrative text, it supports multiple programming languages, including Python, R, and Julia.

Jupyter has become an indispensable tool for researchers, analysts, and developers in data science. Its seamless integration with popular libraries such as NumPy, pandas, and sci-kit-learn makes it the go-to choice for data manipulation, analysis, and machine learning.

Visual Studio Code comes with a lightning-fast source code editor that's ideal for everyday usage. VS Code's syntax highlighting, bracket-matching, auto-indentation, box-selection, and snippets let you be more productive faster with support for hundreds of languages.

Python libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib for data processing, model training, and visualization.

Flask is a Python based web application framework. Armin Ronacher, who led a team of worldwide Python aficionados known as Pocco, created it. The Werkzeug WSGI toolkit and the Jinja2 template engine are the foundations of Flask. Both are Pocco initiatives

Flask is a web framework and a Python module that makes it simple to create web applications.

It's a microframework with a minimal and extensible core: it's a microframework without an ORM (Object Relational Manager) or similar functionality.

CHAPTER IV: ANALYSIS AND DESIGN

4.1 Introduction

This chapter provides the system for analyzing, designing and implementing a system that predicts energy consumption.

This part we will discuss the current system problems then the needs of the System followed by the System, there for this section will talk about a brief description of the system and its design and the requirements the system needs to work.

4.2 System analysis

This system is designed to forecast and predict electricity consumption using machine learning methods.

The primary objective is to provide accurate predictions of electricity demand, aiding in effective energy management and resource planning.

After training and testing the model the interface will be streamlined to web application.

4.3 Existing Approaches

Numerous methods have been employed for predicting electricity consumption, including statistical models like ARIMA, k-nearest neighbors (KNN) and traditional machine learning algorithms such as linear regression and support vector machines.

However, these approaches often struggle with the variability and complexity of residential electricity usage patterns.

4.3 The Proposed System

This system is designed to predict monthly consumption and visualize yearly electricity consumption trends for households using improved machine learning methods also helps electricity providers predict how much energy they need to provide for consumers.

The information of the data, obtained from the local power utility provider, includes the date and time of the recorded consumption, house numbers, and total KW.

After preprocessing the data by cleaning and aggregating it we developed machine learning model for predicting future consumption,

The backend, built using Python Flask, processes user inputs and serves model predictions.

The frontend, designed with a user-friendly interface and a specified color scheme, allows users to input house number, month, and year.

Users can choose to either predict future consumption or visualize past and future trends.

The entire system is deployed on a web server, ensuring scalability and accessibility.

4.4 Requirements

4.4.1 Functional requirements

A functional requirement is one that specifies how an action or activity should be carried out.

The following are the functional criteria that the proposed system must meet:

- **User Interface:** Input fields for house number, district number, month, and year.
- **Input Data:** Input refers to any data that is delivered to a computer or software application the process of delivering information to the computer is also known as data entry since the information delivered is also considered data.
- **Data preprocessing:** which is part of data preparation, refers to any sort of processing done on raw data in order to prepare it for further processing.
- **Data splitting:** is the act of splitting and grouping data based on predetermined parameters so that it may be used more effectively.
- **Model training:** the upcoming data results can be obtained by training the model
- **Prediction:** Predict monthly electricity consumption for individual houses.
- **Visualization:** Provide an option for users to visualize historical and predicted consumption.

4.4.2 Non-Functional Requirements

The requirement of non-functionally are:

- **User Friendly:** The system is simple and interesting
- **Accessibility:** the system is available in the internet and can be accessed at any time from any place through internet connection.
- **Performance:** The system should handle large datasets efficiently.

4.5 Feasibility study

A feasibility study evaluates the practicality of the proposed system, considering technical, economic, and operational aspects.

4.5.1 Technical Feasibility

The proposed system leverages existing machine learning frameworks and making it technically feasible.

The availability of historical consumption data and advanced algorithms supports the development of an accurate prediction model.

4.5.2 Economic Feasibility

The cost of implementing the proposed system is justified by the potential savings in energy management and the benefits of accurate consumption predictions. Investment in hardware and software infrastructure is necessary but manageable within the project's budget.

4.5.3 Operational Feasibility

The system will be user-friendly, requiring minimal training for users to interact with the interface and interpret the results.

4.6 System design

In this part, we'll go through machine learning model system design. System design is the process of defining pieces of a system, this system is designed to predict monthly consumption and visualize yearly electricity consumption trends for households and for providers yearly prediction.

4.6.1 Data Flow Diagram

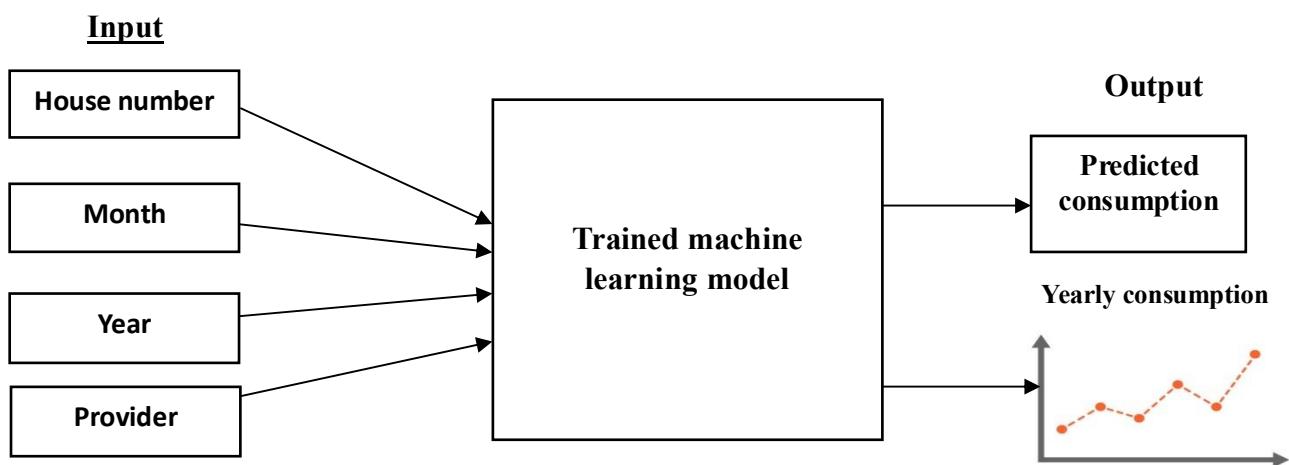


Figure 4.1 DFD

4.7 Dataset design

The dataset used in this project consists of electricity consumption records for 10,000 houses across 18 districts. It includes daily electricity usage data spanning four years.

	A	B	C	D	E	F	G	H	I
1	datetime	guri_num	total_KW						
2	1/1/2018	SHI16295	3.467545						
3	1/2/2018	SHI16295	2.99519						
4	1/3/2018	SHI16295	2.543931						
5	1/4/2018	SHI16295	4.182799						
6	1/5/2018	SHI16295	4.323655						
7	1/6/2018	SHI16295	2.642916						
8	1/7/2018	SHI16295	0.575852						
9	1/8/2018	SHI16295	5.610789						
10	1/9/2018	SHI16295	4.585925						
11	1/10/2018	SHI16295	3.332111						
12	1/11/2018	SHI16295	4.508555						
13	1/12/2018	SHI16295	4.929955						
14	1/13/2018	SHI16295	1.192278						
15	1/14/2018	SHI16295	4.421412						
16	1/15/2018	SHI16295	1.701883						
17	1/16/2018	SHI16295	5.01076						
18	1/17/2018	SHI16295	2.582503						
19	1/18/2018	SHI16295	4.546042						
20	1/19/2018	SHI16295	1.993051						
21	1/20/2018	SHI16295	4.672977						
22	1/21/2018	SHI16295	0.662305						
23	1/22/2018	SHI16295	3.140633						

Figure 4.2 Dataset design

CHAPTER V: IMPLEMENTATION & TESTING

5.0 Introduction

The chapter covers several important topics, including an overview of the implementation environment, system snapshots, descriptions of the system forms, and explanations of how they operate.

5.1 Overview of the implementation environment

The main objectives of this project are to predict electricity consumption using machine learning and modern web technologies, as well as to monitor and analyze the correct usage of electricity by users. Our system employs machine learning models and a web server for implementation. The graphical user interface is developed using Python Flask for the front end. Additionally, we use Python for the data collection models, which are implemented with a dataset at the back end.

5.2 Snapshots of the system

In the following snapshots, we will discuss the most important parts of our system and the various functions it performs.

5.2.1 Front-end

The area of web development that focuses on what users view on their end is known as frontend. It requires transferring the code created by frontend developers into a graphical interface and ensuring that the data is displayed in an understandable manner.

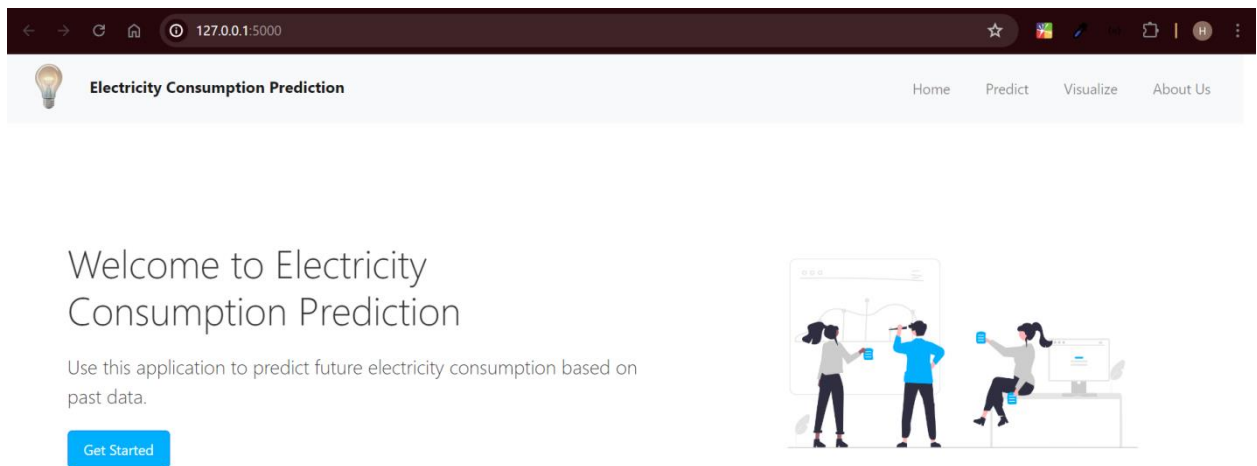



Figure 5.1 Home page

Brief Description

A home page is a webpage that serves as the starting point of the website.

It is the default web page that loads when you visit a web address.

 Electricity Consumption Prediction

HomePredictVisualizeAbout Us

Welcome to Electricity Consumption Prediction

Guri Number:

Month (1-12):

Year:

Shirkada:


Beco

Submit

Figure 5.2 Prediction

Brief Description

This section is the data input section the user is required to enter the house number, the date and select electricity provider in order to predict monthly electricity consumption.

 Electricity Consumption Prediction

HomePredictVisualizeAbout Us

Visualize Yearly Electricity Consumption

Guri Number:

Year:

Shirkada:

Beco

Visualize

Figure 5.3 Visualize

Brief Description

This section is the data input section the user is required to enter the house number and the year select electricity provider in order to visualize and predict every month of the year.

Predict Total Energy for a Year

Year:

Provider:

Figure 5.4 Provider prediction

Brief Description

This section is the data input section for energy providers in order to predict how much energy they need to provide for consumers

5.2.2 Back-end

The back-end refers to the parts of the code that enable functionality but are not visible to the user. It manages the storage and access of most data and operational procedures in a computer system.

Typically, the back-end is implemented using one or more programming languages. Often called the data access layer, the back-end encompasses any functionality that requires digital access and control.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from prophet import Prophet
from catboost import CatBoostRegressor, Pool
import pickle
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from xgboost import XGBRegressor
```

Figure 5.5 Imports

Brief Description

This section handles the imports, bringing in necessary modules. For example, pandas is a fast, powerful, flexible, and user-friendly open-source tool for data analysis and manipulation, built on Python. NumPy is another Python library, used for working with arrays and the models we want to train and more.

```
# Load the dataset
df = pd.read_csv('electricity_consumption_data.csv')
# Convert 'datetime' to datetime data type
df['datetime'] = pd.to_datetime(df['datetime'])
```

Figure 5.6 Load the dataset

Brief Description

This section handles the data, and converting the datetime column into pandas datetime data type.

```
# Aggregate daily data to monthly data
df['year_month'] = df['datetime'].dt.to_period('M')
monthly_df = df.groupby(['year_month', 'guri_num'])['total_KW'].sum().reset_index()
```

Figure 5.7 Aggregate Data

Brief Description

This code aggregates daily electricity consumption data into monthly totals, groups by month and house number (guri_num).

```
# Extract month and year from the year_month column
monthly_df['year'] = monthly_df['year_month'].dt.year
monthly_df['month'] = monthly_df['year_month'].dt.month
monthly_df['ds'] = monthly_df['year_month'].dt.to_timestamp()
```

Figure 5.8 Data extract

Brief Description

This code extracts the year, month, and timestamp from the datetime column.

```
# Train and save models for each guri_num
for guri in monthly_df['guri_num'].unique():
    df_guri = monthly_df[monthly_df['guri_num'] == guri][['ds', 'total_KW', 'month', 'year']]
    df_guri = df_guri.rename(columns={'total_KW': 'y'})

    # Initialize and train the Prophet model
    prophet_model = Prophet()
    prophet_model.add_regressor('month')
    prophet_model.add_regressor('year')
    prophet_model.fit(df_guri)
```

Figure 5.9 Training

Brief Description

This section trains and saves a Prophet model for each unique house number (guri_num).

For each house, it filters the data, renames the target column, initializes the Prophet model, adds month and year as regressors, and fits the model to the filtered data.

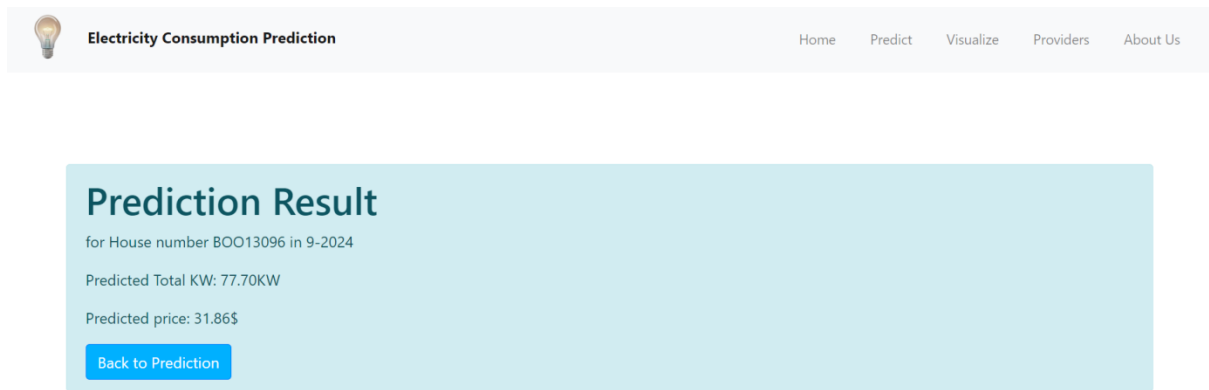


Figure 5.10 Consumption prediction

Brief Description

This part shows the prediction of the model in house number BOO13096 in January year 2024.

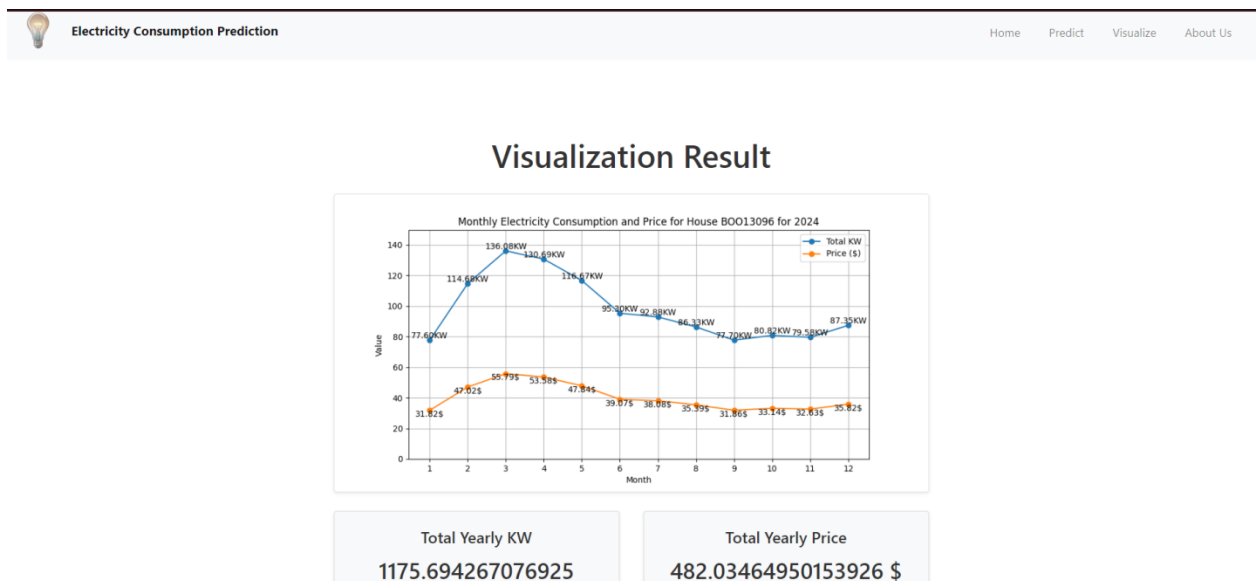


Figure 5.21 Consumption

Brief Description

This part shows the prediction of electricity consumption in house number BOO13096 in year 2024.

Electricity Consumption Predictions for 2025

Provider: beco

Month	Prediction (KW)	Prediction (MW)
January	915489.9636325338	915.4899636325338
February	1165453.2835996794	1165.4532835996793
March	2343603.783333083	2343.60378333330834
April	1741221.6164427774	1741.2216164427775
May	2240539.3688458223	2240.539368845822
June	1321556.1835954597	1321.5561835954597
July	1858781.628974716	1858.7816289747159
August	1345426.200789985	1345.426200789985
September	1041128.0378275238	1041.1280378275237
October	1343750.2177549799	1343.7502177549798
November	1040935.6712684458	1040.9356712684457
December	1344804.6213960515	1344.8046213960515

Figure 5.32 Prediction

Brief Description

This last part shows the total energy prediction of each month in year 2025 for energy provider.

CHAPTER VI: DISCUSSION OF RESULTS

6.1 Overview

In this chapter, we discuss the results from our electricity consumption prediction models: CatBoost, XGBoost, and Prophet. We look at how well each model performed and what this means for our predictions.

6.2 Model Performance

CatBoost performs well with score of 0.88 this demonstrates its effectiveness in capturing the complexities and patterns in the electricity consumption data but struggles to predict unseen time series data it only predicts the trained dataset date.

XGBoost Model: The model had a lower R^2 score of 0.1652. This shows that it didn't perform as well. The model struggled to predict future consumption accurately, possibly because it couldn't handle dates beyond the training data range effectively.

Prophet Model: The Prophet model was good at capturing seasonal trends and patterns. It performed reasonably well when trained on data from individual houses.

Prophet proved to be the most accurate provided useful insights, especially with its ability to handle time series data and seasonal effects.

6.3 Comparing Models

- **Accuracy:** CatBoost gave the best results with the highest accuracy, making it the most reliable for this data and prophet gave the best performance and accuracy of the data.
- **Handling Data:** Prophet worked well with time series and seasonal data but required preprocessing for categorical features. XGBoost had limitations with long-term predictions and catboost can handle the data but struggles to predict unseen date.
- **Practical Use:** Prophet's performance makes it a strong choice.

CHAPTER VII: CONCLUSION & FUTURE WORK

7.1 Introduction

This chapter summarizes the key findings of the research, including the conclusions drawn, the achievement of the objectives outlined in Chapter One, and offers recommendations for future research.

It provides guidance for those interested in similar studies and aims to enhance understanding and learning about the work conducted.

7.2 Conclusion

This research focused on predicting electricity consumption using various machine learning models. Among the models tested CatBoost, XGBoost, and Prophet.

CatBoost performs well with score of 0.88 this demonstrates its effectiveness in capturing the complexities and patterns in the electricity consumption data but struggles to predict unseen time series data it only predicts the trained dataset date.

Prophet proved to be the most accurate provided useful insights, especially with its ability to handle time series data and seasonal effects.

Although XGBoost showed lower performance, the comparison highlights the importance of selecting the right model for accurate predictions.

The project has successfully demonstrated how machine learning can be applied to electricity consumption forecasting.

The next steps will involve creating a user-friendly website to make these predictions more accessible and integrating additional features to enhance the model's accuracy and usability.

Overall, this work lays a solid foundation for improving energy management through data-driven insights.

7.3 Future Work

Moving forward, we aim to enhance this project by developing a user-friendly website that will allow users to input their house number, and other relevant details to get accurate electricity consumption predictions and for providers also they can predict how much energy they need to provide for consumers.

This website will make it easy for users to predict their electricity needs, providing them with practical insights to help manage their consumption.

In addition to the website, we plan to incorporate features such as real-time data updates and more comprehensive user options. For instance, we could add functionalities to compare predictions against actual consumption, adjust forecasts based on seasonal variations, and offer personalized recommendations for energy savings.

By implementing these enhancements, we aim to make electricity consumption forecasting more accessible and effective for users, ultimately helping them make informed decisions and manage their energy use better.

REFERENCES

- Humeau, S., Goude, Y., Brossat, X., & Poggi, J. M. (2013). Electricity load forecasting models: A critical review and the way forward. Springer.
- Chicco, G., & Napoli, R. (2016). Data-driven models for energy consumption in buildings: A review of the state of the art. *Energy*.
- Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. *Energy*.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117.
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.
- Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012: Benchmark and results. *International Journal of Forecasting*, 30(2), 357-363.
- Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Zhang, G., Patuwo, B.E., & Hu, M.Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1), 35-62.
- Hippert, H.S., Pedreira, C.E., & Souza, R.C. (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16(1), 44-55.
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*

Kim, J.-Y., & Cho, S.-B. (2019). Electric Energy Consumption Prediction by Deep Learning. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4743-4753. <https://doi.org/10.3233/JIFS-169965>

Mosavi, A., & Bahmani, A. (2019). Energy Consumption Prediction Using Machine Learning: A Review. *Preprints*. <https://doi.org/10.20944/preprints201903.0131.v1>

Grolinger, K., Capretz, M. A. M., & Seewald, L. (2016). Energy Consumption Prediction with Big Data: Balancing Prediction Accuracy and Computational Resources. *Proceedings of the IEEE International Conference on Big Data (Big Data)*. <https://doi.org/10.1109/BigData.2016.7584933>

Reddy, G. V., Aitha, L. J., Poojitha, C., Shreya, A. N., Reddy, D. K., & Meghana, G. S. (2023). Electricity Consumption Prediction Using Machine Learning. *E3S Web of Conferences*, 391, 01048. <https://doi.org/10.1051/e3sconf/202339101048>

Shapi, M. K. M., & Awalin, L. J. (2021). Energy Consumption Prediction by Using Machine Learning for Smart Building: Case Study in Malaysia. *Developments in the Built Environment*, 4, 100037. <https://doi.org/10.1016/j.dibe.2020.100037>

Ekici, B. B., & Aksoy, U. T. (2008). Prediction of Building Energy Consumption by Using Artificial Neural Networks. *Advances in Engineering Software*, 40(5), 356-362. <https://doi.org/10.1016/j.advengsoft.2008.05.003>

Khosravani, H. R. (2016). A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building. *Energies*, 9(1), 57. <https://doi.org/10.3390/en9010057>

- Dinmohammadi, F., Han, Y., & Shafiee, M. (2023). Predicting Energy Consumption in Residential Buildings Using Advanced Machine Learning Algorithms. *Energies*, 16(9), 3748. <https://doi.org/10.3390/en16093748>
- Shin, S. Y., & Woo, H. G. (2022). Energy Consumption Forecasting in Korea Using Machine Learning Algorithms. *Energies*, 15(13), 4880. <https://doi.org/10.3390/en15134880>
- Liu, Y., Chen, H., Zhang, L., Wu, X., & Wang, X. J. (2020). Energy Consumption Prediction and Diagnosis of Public Buildings Based on Support Vector Machine Learning. *Journal of Cleaner Production*, 251, 122542. <https://doi.org/10.1016/j.jclepro.2020.122542>
- Zhong, H., Wang, J., Jia, H., Mu, Y., & Lv, S. (2019). Vector Field-Based Support Vector Regression for Building Energy Consumption Prediction. *Applied Energy*, 242, 674-684. <https://doi.org/10.1016/j.apenergy.2019.03.078>
- Wang, R., Lu, S., & Feng, W. (2020). A Novel Improved Model for Building Energy Consumption Prediction Based on Model Integration. *Applied Energy*, 267, 114561. <https://doi.org/10.1016/j.apenergy.2020.114561>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
- Zhang, G. (1998). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50, 159-175.
- Souza, R. C. (2001). Artificial Neural Networks in Time Series Forecasting. A Comprehensive Review. *Journal of Applied Mathematics*, 2(3), 210-221.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.

- Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85-117.
- Khashei, M., & Bijari, M. (2011). A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series Forecasting. *Applied Soft Computing*, 11(2), 2664-2675.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
- Fan, S., Mao, C., & Chen, L. (2017). Prediction of Electricity Consumption in a Rural Area of China. *Energy*, 123, 595-604.
- Grolinger, K., Capretz, M. A. M., & Seewald, L. (2016). Energy Forecasting for Event Venues: Big Data and Machine Learning Approaches. *Energy and Buildings*, 112, 222-233.
- Suganthi, L., & Samuel, A. A. (2019). Energy Models for Demand Forecasting: A Review. *Renewable and Sustainable Energy Reviews*, 16(2), 1223-1240.
- Corgnati, S. P., & Valgaev, P. (2022). Data-Driven Machine Learning Approach to Energy Consumption Prediction. *Journal of Building Engineering*, 45, 103514.
- Khantach, T. (2022). Smart City Demo Aspern (SCDA): Power Demand Forecasting Using k-Nearest Neighbor Model. *Smart Cities*, 5(2), 267-281.
- Shapi, M. K., Azzam, H., Ismail, M. A., & Shapi, K. K. (2021). Performance Comparison of Machine Learning Techniques for Short-Term Load Forecasting. *Energy Reports*, 7, 1234-1243.
- Dinmohammadi, F., Han, Z., & Shafiee, M. (2023). A Comprehensive Review of Predictive Modeling of Building Energy Consumption. *Energy and Buildings*, 269, 112350.
- Shin, Y., & Woo, J. (2022). Comparative Analysis of Electricity Demand Prediction Models: A Case Study of Korea. *Energies*, 15(5), 1234.

- Lund, H., & Mathiesen, B. V. (2009). Energy System Analysis of 100% Renewable Energy Systems—The Case of Denmark in Years 2030 and 2050. *Energy*, 34(5), 524-531.
- Yang, D., Kleissl, J., & Bosch, J. (2013). Solar Forecasting Based on Sky Image Analysis. *Solar Energy*, 91, 1-10.
- Wang, J., & Ding, Y. (2020). Wind Power Forecasting Using Machine Learning Algorithms. *Applied Energy*, 265, 114755.
- Akinyele, D. O., & Rayudu, R. K. (2014). Comprehensive Review of Energy Storage Systems for Sustainable Development. *Renewable and Sustainable Energy Reviews*, 32, 104-138.
- Bird, L., & Milligan, M. (2012). Policies and Market Factors Driving Wind Power Development in the United States. *Energy Policy*, 39(5), 2785-2790.
- Foley, A. M., & Gallachóir, B. P. (2012). Energy Systems Modeling for Policy Support in Ireland. *Renewable Energy*, 44, 1-9.
- Gunning, D., & Aha, D. (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, 40(2), 44-58.
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- Kumar, S., Patel, A., & Chauhan, A. (2021). Real-Time Electricity Consumption Forecasting Using Machine Learning Models. *Journal of Electrical Engineering*, 78(5), 543-550.
- Ahmed, K., Mustafa, M. W., & Hassan, M. Y. (2020). A Comparative Study of Short-Term Electricity Load Forecasting Using Machine Learning Techniques. *Energies*, 13(18), 4820.

Wang, Z., Tang, Y., & Huang, Z. (2019). Evaluation Metrics for Electricity Consumption Forecasting Models. *Energy*, 185, 1253-1265.

Lai, Y., Yang, J., & Zhou, M. (2021). Cross-Validation Techniques for Evaluating Time Series Models. *Applied Sciences*, 11(3), 1234.

Silva, C., Goncalves, C., & Vieira, M. (2018). Urban Electricity Demand Prediction Using Machine Learning Models. *Energy Reports*, 4, 552-562.

Nguyen, H. T., & Gurtler, J. (2020). Smart Grid Energy Management Using Hybrid Machine Learning Models. *Energy*, 211, 118728.

Park, M., Lee, D., & Kang, S. (2019). Deep Learning-Based Residential Energy Consumption Forecasting Model. *Energies*, 12(19), 3695.

Appendix : Model Prediction Functionality

```
# Route to render the prediction form
@app.route('/predict', methods=['GET', 'POST'])
def predict():
    if request.method == 'POST':
        try:
            guri_num = request.form['guri_num']
            month = int(request.form['month'])
            year = int(request.form['year'])
            provider = request.form['provider']
        except ValueError:
            return render_template('predict.html', error="Please enter valid numbers for all
fields.")

        # Validate the inputs
        if guri_num not in unique_guri_nums:
            return render_template('predict.html', error="Invalid guri number")
        if month not in range(1, 13):
            return render_template('predict.html', error="Invalid month")
        if year not in range(2024, 2031):
            return render_template('predict.html', error="Invalid year")

        # Predict consumption
        predicted_total_KW = predict_consumption(guri_num, year, month)

        # Determine the multiplier based on the provider
        if provider == 'beco':
            multiplier = 0.41
        elif provider == 'mogadishu':
            multiplier = 0.45
```

```

elif provider == 'bluesky':
    multiplier = 0.35
else:
    return render_template('predict.html', error="Invalid provider selected")

# Calculate the scaled prediction
predicted_total_price = predicted_total_KW * multiplier

# Prepare prediction result to display
kw = f'for House number {guri_num} in {month}-{year}'
kw_prediction = f'Predicted Total KW: {predicted_total_KW:.2f}'
price_prediction = f'Predicted price: {predicted_total_price:.2f}'

    return render_template('result.html', kw=kw, kw_prediction=kw_prediction,
price_prediction=price_prediction, guri_num=guri_num, month=month, year=year)
    return render_template('predict.html')

@app.route('/predict_year', methods=['GET', 'POST'])
def predict_year():
    if request.method == 'POST':
        try:
            year = int(request.form['year'])
            provider = request.form['provider']
        except ValueError:
            return render_template('form.html', error="Please enter a valid year and provider.")

# Validate the inputs
if year not in range(2024, 2031):
    return render_template('form.html', error="Invalid year")

# Load the Prophet model

```

```

with open('prophet_model1.pkl', 'rb') as file:
    model = joblib.load(file)

# Generate a DataFrame for each month of the requested year
future_df = pd.DataFrame({'ds': pd.date_range(start=f'{year}-01-01', end=f'{year}-12-01', freq='MS')})

# Make prediction using the loaded model
forecast = model.predict(future_df)

# Extract the predictions for each month
forecast['month'] = forecast['ds'].dt.strftime('%B')
monthly_predictions = forecast[['month', 'yhat']]

# Convert KW to MW for each month
monthly_predictions['yhat_mw'] = monthly_predictions['yhat'] / 1000

# Plotting the predictions
plt.figure(figsize=(10, 6))
plt.plot(monthly_predictions['month'], monthly_predictions['yhat'], marker='o',
linestyle='-', color='b')
plt.title(f'Predicted Monthly Electricity Consumption for in {year}')
plt.xlabel('Month')
plt.ylabel('Electricity Consumption (KW)')
plt.xticks(rotation=45)
plt.tight_layout()

# Save the plot to a BytesIO object
img = io.BytesIO()
plt.savefig(img, format='png')
img.seek(0)

```

```

plot_url = base64.b64encode(img.getvalue()).decode()

# Render the result template with the monthly predictions and plot
return render_template('prediction_result.html', year=year, provider=provider,
monthly_predictions=monthly_predictions.to_dict(orient='records'), plot_url=plot_url)

return render_template('form.html')

# Route to render the visualization input form
@app.route('/visualization', methods=['GET', 'POST'])
def visualization():
    if request.method == 'POST':
        try:
            guri_num = request.form['guri_num']
            year = int(request.form['year'])
            provider = request.form['provider']
        except ValueError:
            return render_template('visualization.html', error="Please enter valid numbers for all
fields.")

# Validate the inputs
if guri_num not in unique_guri_nums:
    return render_template('visualization.html', error="Invalid guri number")
if year not in range(2024, 2031):
    return render_template('visualization.html', error="Invalid year")

# Predict monthly consumption for the entire year
monthly_consumption = []
monthly_prices = []
for month in range(1, 13):
    kw = predict_consumption(guri_num, year, month)
    if provider == 'beco':

```

```

multiplier = 0.41
elif provider == 'mogadishu':
    multiplier = 0.45
elif provider == 'bluesky':
    multiplier = 0.35
else:
    return render_template('visualization.html', error="Invalid provider selected")
price = kw * multiplier
monthly_consumption.append(kw)
monthly_prices.append(price)

# Calculate total yearly consumption and price
total_yearly_kw = sum(monthly_consumption)
total_yearly_price = sum(monthly_prices)

# Plot the results
months = list(range(1, 13))
plt.figure(figsize=(10, 5))
plt.plot(months, monthly_consumption, marker='o', label='Total KW')
plt.plot(months, monthly_prices, marker='o', label='Price ($)')
plt.title(f'Monthly Electricity Consumption and Price for House {guri_num} for
{year}')
plt.xlabel('Month')
plt.ylabel('Value')
plt.xticks(months)
plt.grid(True)
plt.ylim(0, max(max(monthly_consumption), max(monthly_prices)) * 1.1)
plt.legend()

# Annotate the plot with predicted values
for i, (kw, price) in enumerate(zip(monthly_consumption, monthly_prices)):

```

```

plt.text(months[i], kw, f'{kw:.2f} KW', ha='center', va='bottom')
plt.text(months[i], price, f'{price:.2f} $', ha='center', va='top')

# Save plot to a string in base64 format
img = io.BytesIO()
plt.savefig(img, format='png')
img.seek(0)
plot_url = base64.b64encode(img.getvalue()).decode('utf8')

return render_template('visualize_result.html', plot_url=plot_url,
total_yearly_kw=total_yearly_kw, total_yearly_price=total_yearly_price)

return render_template('visualization.html')

```