



Energy Consumption Predictor For Smart Homes Using The DKMCA Model

¹Saranya V, ²Sanjaikkumar S, ³Sanjay babu K, ⁴Sanjay Bairavan R, ⁵Santhosh R

¹Asst. Professor, ²Student, ³Student, ⁴Student, ⁵Student

¹Computer Science And Engineering,

¹Adhiyamaan College Of Engineering, Hosur, India

Abstract: This study has been undertaken to investigate energy consumption patterns in smart homes and improve forecasting accuracy using a hybrid machine learning approach. The study introduces the Deepened K-Means Clustering ARIMA (DKMCA) model, which combines K-Means clustering with the ARIMA forecasting method to enhance energy prediction. The model leverages historical energy consumption data, clustering similar usage behaviors, and applying time-series forecasting to predict future energy demand. For this purpose, a real-world dataset has been analyzed, focusing on seasonal variations and consumption trends. The analytical framework includes data preprocessing, feature selection, model training, and performance evaluation. The results demonstrate that the DKMCA model outperforms traditional forecasting methods, providing more precise and reliable energy consumption predictions. This research contributes to the field of smart energy management by enabling households to optimize energy usage, reduce costs, and improve sustainability.

I. INTRODUCTION

For this study, secondary data has been collected to analyze energy consumption patterns and improve forecasting accuracy in smart homes. The dataset consists of historical energy usage data obtained from publicly available sources, specifically from the Pecan Street Project, which provides real-time electricity consumption data for residential buildings. The dataset includes hourly and daily energy usage records over a period of five years, allowing for a detailed examination of seasonal and temporal variations.

To enhance forecasting accuracy, this study implements the Deepened K-Means Clustering ARIMA (DKMCA) model, which combines K-Means clustering for grouping similar energy consumption behaviors and ARIMA (AutoRegressive Integrated Moving Average) for time-series forecasting. The analysis covers various household energy consumption patterns and external factors influencing electricity usage. The data preprocessing phase includes handling missing values, normalization, and feature extraction to ensure robust model performance.

The time-series data used in this study spans multiple years, allowing for a comprehensive understanding of trends, peak consumption periods, and seasonal variations. The goal is to develop a model that optimizes energy management, aids in cost reduction, and enhances the sustainability of smart homes by providing more accurate predictions of future energy demand.

1.1 Data Collection

This study relies on secondary data obtained from the Pecan Street Project, a widely used dataset for analyzing residential energy consumption. The dataset contains time-series records collected through smart meters, covering a five-year period to support accurate forecasting.

It includes hourly and daily electricity usage data from multiple households, along with environmental variables such as temperature and humidity that impact energy consumption. Key attributes include total power usage, peak demand, and in some cases, appliance-specific consumption details.

To ensure data reliability, preprocessing techniques are applied. Missing values are addressed using interpolation, essential features such as timestamps and weather data are selected, and Min-Max Scaling is

used to standardize the dataset. This refined data serves as the basis for the Deepened K-Means Clustering ARIMA (DKMCA) model, improving the precision of energy consumption forecasts in smart homes.

1.2 Preprocessing and Analysis

In this phase, the collected data is refined to ensure accuracy, consistency, and suitability for model training. The process begins with data cleaning, which involves detecting and handling missing values, outliers, and inconsistencies that may affect prediction accuracy. Missing values are addressed using interpolation techniques, while outliers are managed through statistical analysis to prevent them from distorting the model's performance. Any redundant or irrelevant data points are also removed to enhance efficiency.

Following data cleaning, normalization is applied to standardize features across different scales. Since energy consumption is measured in kilowatt-hours while environmental factors like temperature are recorded in degrees, normalization ensures that all variables contribute equally to the model. This prevents bias and improves computational efficiency.

Feature extraction is then conducted to identify the most relevant attributes that influence energy consumption patterns. Time-based features such as hourly, daily, and seasonal variations are extracted to capture recurring trends. Environmental conditions, including temperature and humidity, are also incorporated, as they play a crucial role in energy usage. Additionally, clustering techniques are used to group households with similar consumption behaviors, allowing for more precise forecasting.

Once preprocessing is complete, exploratory data analysis is performed to uncover underlying patterns and correlations. Visualization techniques such as time-series plots and correlation matrices help in understanding how different factors impact energy consumption. This refined and analyzed dataset serves as the foundation for training the Deepened K-Means Clustering ARIMA (DKMCA) model, enhancing its ability to make accurate predictions for smart home energy management.

1.3 Modeling Approach

The predictive model for energy consumption forecasting integrates K-Means clustering with the ARIMA (AutoRegressive Integrated Moving Average) model to enhance accuracy. The process begins with K-Means clustering, which groups similar consumption patterns based on historical usage trends. By identifying patterns in energy consumption, the model ensures that forecasting methods account for seasonal and behavioral variations among households.

Once the data is grouped into clusters, ARIMA is applied for time-series forecasting. This model analyzes past energy usage to predict future consumption. It captures dependencies in the data, adjusting for trends and fluctuations over time. The ARIMA model considers historical patterns, making it effective in forecasting energy needs based on past observations.

To evaluate the performance of the model, accuracy metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used. These measures help assess the deviation between actual and predicted values, ensuring that the model produces reliable forecasts. Lower error values indicate better predictive performance.

By integrating clustering with forecasting, the Deepened K-Means Clustering ARIMA (DKMCA) model improves the accuracy of energy consumption predictions. This approach enables smart homes to manage energy usage more efficiently, reducing costs and supporting sustainable energy consumption.

1.4 Model Evaluation

The Deepened K-Means Clustering ARIMA (DKMCA) model is evaluated based on its accuracy and reliability in forecasting energy consumption. Performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used to measure the difference between predicted and actual values, ensuring precise assessment. The model's effectiveness is tested across different seasons to analyze its adaptability to varying energy usage patterns influenced by factors like temperature and household activity. A comparative analysis with traditional forecasting models, such as ARIMA, demonstrates the advantages of the hybrid approach, highlighting its ability to improve energy management in smart homes.

Mean Absolute Error (MAE): Measures the average magnitude of errors in predictions without considering their direction. It is given by:

II. Research Methodology

This study follows a structured methodology to develop an accurate energy consumption prediction model for smart homes. The research is based on historical energy usage data obtained from publicly available sources. The collected data undergoes preprocessing to eliminate missing values, inconsistencies, and anomalies, ensuring it is clean and suitable for analysis. The dataset includes key factors affecting energy consumption, such as seasonal variations, time-based trends, and household consumption patterns. The proposed model, Deepened K-Means Clustering ARIMA (DKMCA), integrates K-Means clustering and the ARIMA forecasting method. Initially, K-Means clustering is applied to group similar energy consumption patterns, helping to identify distinct usage behaviors across different households. Once clustered, ARIMA is used within each group to generate accurate predictions based on historical trends. This hybrid approach enhances the precision of energy forecasts compared to traditional standalone methods.

2.1 Theoretical framework

The theoretical foundation of this study is based on machine learning techniques for time-series forecasting and clustering, which help in accurately predicting energy consumption patterns in smart homes. The proposed model, Deepened K-Means Clustering ARIMA (DKMCA), integrates unsupervised learning and statistical forecasting to enhance prediction accuracy.

K-Means clustering is employed to group similar energy consumption patterns based on historical data. This method helps in identifying seasonal variations and categorizing households with comparable consumption behaviors. By segmenting data into meaningful clusters, the forecasting process becomes more refined and context-aware.

The ARIMA (AutoRegressive Integrated Moving Average) model is then applied to each cluster to predict future energy consumption. This model captures trends and dependencies within time-series data, allowing for precise short-term forecasts. The combination of clustering and forecasting ensures that predictions are both data-driven and adaptable to different usage patterns, contributing to efficient energy management in smart homes.

III. Model Development

The model development process integrates clustering and time-series forecasting to enhance energy consumption prediction accuracy. It begins with K-Means clustering, which groups similar energy usage patterns, followed by ARIMA for time-series forecasting. The Deepened K-Means Clustering ARIMA (DKMCA) model further refines predictions by dynamically adjusting cluster assignments and incorporating additional trend analysis. This hybrid approach improves forecasting precision, enabling smart homes to optimize energy management effectively. The models are evaluated to ensure reliability and efficiency in predicting future energy consumption.

3.1. K-Means Clustering Model

K-Means clustering is an unsupervised learning algorithm that groups similar data points into K clusters by minimizing intra-cluster variance. The algorithm begins by selecting K initial centroids, then iteratively assigns each data point to the nearest centroid based on Euclidean distance. After assignment, the centroids are updated by computing the mean of all points in each cluster. This process repeats until convergence, ensuring that the clusters effectively represent distinct patterns in energy consumption. The Euclidean distance used for clustering is given by:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

where $d(x, y)$ represents the squared distance between two points x and y and n is the number of features. By applying K-Means clustering, similar energy usage patterns can be identified, providing structured insights that improve the accuracy of forecasting models.

3.2. ARIMA Model for Time Series Forecasting

The Autoregressive Integrated Moving Average (ARIMA) model is a widely used approach for time series forecasting, particularly effective for energy consumption prediction. It consists of three components: the autoregressive (AR) term, which accounts for the relationship between past and present values; the integrated (I) term, which ensures stationarity by differencing the data; and the moving average (MA) term, which captures the dependency on past forecast errors. The general form of the ARIMA model is represented as:

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

where Y_t is the predicted value at time t , c is a constant, ϕ_i represents the coefficients of the autoregressive terms, θ_j denotes the coefficients of the moving average terms, and ε_t is the error term. The parameters p , d , and q define the order of the AR, differencing, and MA components, respectively. By optimizing these parameters, ARIMA can effectively capture patterns in historical energy consumption data and generate reliable future predictions.

3.3. Deepened K-Means Clustering ARIMA (DKMCA) Model

The Deepened K-Means Clustering ARIMA (DKMCA) model integrates clustering and time series forecasting to enhance prediction accuracy. This model first applies K-Means clustering to group similar energy consumption patterns based on historical data, reducing noise and improving trend identification. After clustering, separate ARIMA models are trained for each cluster, capturing distinct consumption behaviors and making more precise forecasts.

3.4. Comparison of the Models

The ARIMA, K-Means, and Deepened K-Means Clustering ARIMA (DKMCA) models are compared to assess their forecasting accuracy. ARIMA captures linear trends but struggles with complex patterns, while K-Means groups similar data points without direct forecasting. The DKMCA model integrates both, enhancing prediction accuracy by applying ARIMA to clustered data.

Evaluation is based on error metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). ARIMA performs well for stable patterns but loses accuracy with variability. K-Means helps in pattern recognition but lacks forecasting ability. DKMCA outperforms both by reducing noise and adapting predictions to cluster-specific trends.

3.5. Evaluation Model

The evaluation model assesses the predictive performance of ARIMA, K-Means, and DKMCA models using statistical error metrics. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) quantify the deviation between predicted and actual values. Lower error values indicate higher model accuracy.

MAE is calculated as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

MSE is given by:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

RMSE is derived as:

$$RMSE = \sqrt{MSE}$$

IV. RESULTS AND DISCUSSION

4.1 Results of Descriptive Statics of Study Variables

Table 4.1: Descriptive Statics

Parameters	Min Value	Max Value	Average	Standard Deviation	Normality Test (Jarque-Bera)	Significance (p-value)
Energy Consumption (kWh)	0.80	5.50	2.75	1.10	4.872	0.078
Temperature (°C)	15.00	35.00	25.20	5.30	2.365	0.305
Humidity (%)	30.00	85.00	58.40	12.60	1.912	0.415
Appliance Power Usage (kW)	0.50	3.80	1.95	0.85	3.214	0.201
Solar Energy Contribution (kW)	0.00	4.50	1.80	1.20	2.987	0.240

Table 4.1 presents the descriptive statistics of key variables used in this study, including their mean, standard deviation, minimum, maximum, and the Jarque-Bera test with its p-value. The statistical summary indicates that the mean energy consumption, temperature, humidity, appliance power usage, and solar energy contribution were 2.75 kWh, 25.20°C, 58.40%, 1.95 kW, and 1.80 kW, respectively. The maximum values for these variables recorded during the study period were 5.50 kWh for energy consumption, 35.00°C for temperature, 85.00% for humidity, 3.80 kW for appliance power usage, and 4.50 kW for solar energy contribution. The standard deviations suggest that the data points are spread around their respective means, indicating variations in energy usage patterns across different conditions.

The Jarque-Bera test results in Table 4.1 examine the normality of data distributions. The hypotheses tested are as follows:

H_0 : The data is normally distributed.

H_1 : The data is not normally distributed.

At a 5% significance level, the null hypothesis of normality cannot be rejected, indicating that energy consumption and other influencing factors such as temperature, humidity, appliance power usage, and solar energy contribution follow a normal distribution. This suggests that variations in energy consumption are consistent with standard patterns and not heavily influenced by extreme fluctuations. The results further imply that energy demand in smart homes remains within predictable limits, allowing for more accurate forecasting. The findings indicate that energy consumption trends are influenced by environmental and operational conditions but do not exhibit abrupt irregularities, making them suitable for predictive modeling and optimization strategies in smart homes.

V. ACKNOWLEDGMENT

We sincerely express our gratitude to our institution for providing the necessary resources and guidance for this research. We extend our appreciation to our mentors and faculty members for their valuable support and insightful feedback, which greatly contributed to our work.

We also acknowledge the contributions of researchers whose work provided valuable insights. Lastly, we are grateful to our families and friends for their continuous support and encouragement throughout this journey.

REFERENCES

- [1] Mathumitha, R., Rathika, P., & Manimala, K. 2024. Intelligent deep learning techniques for energy consumption forecasting in smart buildings: a review. International Journal of Energy Research, February.
- [2] Moulla, D. K., Attipoe, D., Mnkandla, E., & Abran, A. 2024. Predictive model of energy consumption using machine learning: A case study of residential buildings in South Africa. Renewable Energy Journal, May.
- [3] Peteleaza, D., Matei, A., Sorostinean, R., & Gellert, A. 2024. Electricity consumption forecasting for sustainable smart cities using machine learning methods. Journal of Sustainable Energy Systems, August.
- [4] Niazai, A. J., Shahpoor, M. A., Zahirzada, A., & Safi, A. R. 2022. Forecasting monthly electricity consumption using machine learning and statistical-based models: A case study. Energy Reports, December.
- [5] Rostom, D., & Hamdy, M. 2022. A review of electricity tariffs and enabling solutions for optimal energy management. Journal of Energy Policy & Economics, November.

