Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

CLASSIFICATION AND CLUSTERING OF RURAL AREA ELECTRICITY CONSUMERS USING MACHINE LEARNING AND GEOGRAPHICAL INFORMATION SYSTEMS

[1]Rahul R. Patil, [2]Dr. Mukesh Kumar Gupta, [3]Dr. Kishor V. Bhadane

[1]Research Scholar, Department of Electrical Engineering,Suresh Gyan Vihar University, Jaipur, India

[2]Professor, Department of Electrical Engineering, Suresh Gyan Vihar University,
Jaipur, India

Department of Electrical Engineering, Amrutvahini College of Engineering, Sangamner
Maharashtra, India

rahulpatil2050@gmail.com, mukeshkr.gupta@mygyanvihar.com, mkgupta72@gmail.com, kishorbhadane31@gmail.com

Abstract

Sustainable developments and reliable power system operations require effective controls on the electricity consumption behavior which is very essential to achieve the saving target and control. In the energy sector, practically machine learning (ML) techniques are being used to study the consumption patterns of the customers which are helpful in effective demand response management, energy efficiency policy. In this article, study carried out on cluster analysis for classification and evaluation of energy consumption (EC) patterns of domestic consumers with single phase low voltage level in Sinnar tehsil under Maharashtra State. Dataset consists of consumption data, geographical data and weather conditions. There are many clustering techniques available but here K-means (KM) clustering technique used to classify collected information into two and three classes. Basically, in this article clustering form based on the Energy Consumption pattern as well as spatial location of consumers using QGIS tool. To achieve optimal number of clusters, we apply elbow method and to cross check Silhouette score technique used to distinguish the clusters based on average monthly consumption. Different classification models are developed by the results obtained from K-means clustering to assess the consumers EC patterns. GIS is used to visualize, analyze and interpret these clusters spatially. The results analyzed that the K-means clustering reveals minimum overlap of 3.13 % and 1.48% for two classes and three classes respectively mostly living in gavthan colonies and wadi vasti domestic area having kutcha, pucca houses as per GIS visualization and analysis. The classification techniques exhibit that the XGBoost

Received: July 18, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

technique achieves a higher accuracy for cluster classification of this kind of information data.

Keywords— Machine Learning, QGIS, Domestic Consumers, Clustering, Classification techniques, Energy Consumption, Spatial Data.

I. INTRODUCTION

THE NEED FOR GRANULAR ENERGY CONSUMPTION ANALYSIS

An electrical power system includes the generation, transmission, and distribution of electricity. Power is produced at a certain voltage in different regions and is then transmitted and distributed to densely populate metropolitan areas, scatter rural areas. The supplier supply electrical power to the industrial, residential, commercial, traction, agriculture and many other sectors which is received from substation of distribution and transmission system The analysis of energy consumption (EC) is significant to both developed and developing countries, as it strongly influences their socio-economic development and power grid stability. EC is growing dramatically around the globe with the increase in population. Monthly Energy Consumption pattern (MECP) changes with the time slot of the day, week, circumstantial weather changes, climatic shifts and several other factors [3]. In industrial zone, the MECP is mostly constant with the exception of a limited number of industrial sectors. Agricultural consumption requires more discipline. In the household category of consumers, cloud cover and temperature dictate the consumption of EC, contributing to its dependency as a meteorological variable [2]. The demand behavior for domestic consumers changes with variables such as consumer seasonal festivals, economic status, geography of consumer dwelling, lifestyle, building design, along with many other factors.

India is the fourth largest energy consumer in the world and energy demand in rural areas is on a steady rise. Worldwide, Per capita energy consumption is critically important as it currently stands at 1.9 tons oil equivalent per year. The same applies to India, which has lower per capita energy consumption than world average but is experiencing rapid growth in energy demand. A significant portion of India's rural regions have a sizeable population as well. In India, the vast amount of domestic energy consumption in the rural regions is aimed primarily at household tasks such as cooking and lighting up the household.

Efficient management and allocation of electrical energy resources are essential in light of increasing energy demands and environmental concerns. Understanding the intricacies of energy consumption patterns is critical for developing effective energy policies and implementing demand-side management strategies. The rise of smart meters has created an era of unprecedented data availability, providing detailed insights into residential and commercial energy usage that were previously unattainable. However, this abundance of data presents significant analytical challenges. Traditional methods for analyzing energy consumption often fail to capture the diversity of consumer behavior, as they do not account for the various factors influencing energy use across different consumer groups. Therefore, there is a pressing

Received: July 18, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

need to develop advanced analytical techniques capable of extracting meaningful insights from this complex data landscape.

This study presents a novel approach to classifying and clustering electrical energy consumers by combining machine learning (ML) with Geographic Information Systems (GIS). By integrating these powerful tools, we aim to gain a deeper understanding of energy consumption patterns at a detailed level. This will enable us to identify distinct consumer groups based on their energy use profiles, spatial distribution, and associated socioeconomic factors. The insights obtained from this analysis can guide the creation of targeted interventions to improve energy efficiency and optimize resource allocation, thereby contributing to a more sustainable and resilient energy future. The main goal is to demonstrate the effectiveness of this integrated ML-GIS approach in enhancing energy consumption analysis and to inform the development of more effective energy management strategies.

The information of the dataset is obtained from the electricity data of Maharashtra State Electricity Distribution Company Ltd. It contains the monthly electricity consumption records (ECP) for low voltage (LV) domestic consumers around 2500 residential category consumers belong to villages of Sinnar tehsil, Nasik, Maharashtra, India is consider for study analysis. Here, K-means clustering algorithm is used to form the group of consumers. Different classification techniques, namely RF, SVM, XGBoost, KNN, LR, DT, NB are used to identify the consumer group. Accuracy and precision are some of the metrics employed to evaluate performance of the classifiers. The five main aspects of this study are as follows:

A. Apply clustering algorithm to categorize domestic consumers EC pattern and show location using GIS tool.

B. The clustering outcome of K-means is used as a preliminary stage for classification.

C. Based on the spatial data prepared consumers cluster and visualize and analyse EC pattern clusters in it.

D. Apply Classification algorithms to classify EC Pattern.

E. The performance of LR classifier is outstanding among the other classifiers as per the experimental results.

Now the rest of this paper is structured as follows. In Section 2, we provide a review from other researchers. In Section 3 we provide methodology along with description of the dataset. In section 4, we examine results and discussion of classification and clustering. In Section 5, we conclude the article with implication and future direction.

II. Literature Review

The application of machine learning (ML) and Geographic Information Systems (GIS) has significantly advanced the field of energy consumption analysis. Numerous studies have explored the use of various ML techniques for energy forecasting [4], [2], demonstrating their ability to model complex temporal dependencies and predict future energy demand with improved accuracy compared to traditional methods [5], [6]. These models range from simple

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

linear regression [7] to sophisticated deep learning architectures like Long Short-Term Memory (LSTM) networks [8] and recurrent neural networks [9], each tailored to specific data characteristics and forecasting horizons. In addition to forecasting, ML has proven valuable in modeling consumer behavior [10], [11], enabling the identification of factors influencing energy use and the development of personalized energy management strategies [12], [13]. Studies have explored the use of various ML algorithms, including support vector machines (SVMs) [14], artificial neural networks (ANNs) [14], and Bayesian networks [15], to understand the complex interplay between consumer choices and energy consumption. The use of machine learning as a service (MLaaS) [16] has also emerged as a promising approach for providing scalable and flexible solutions for energy consumption analysis.

Furthermore, GIS plays a crucial role in integrating spatial information into energy consumption analysis. GIS provides the tools to visualize and analyze the spatial distribution of energy consumption data [17], [18], [19], revealing patterns and correlations that might otherwise be missed. Studies have used GIS to map energy consumption across different geographic areas [20], [21], identify hotspots of high energy use, and investigate the influence of geographic factors on energy consumption [22]. The integration of GIS with socioeconomic data allows researchers to explore the relationships between energy consumption and demographic characteristics [11], [10], enabling a more nuanced understanding of the drivers of energy use. The combination of GIS with clustering algorithms [23], [24] allows for a spatial segmentation of consumers based on their energy consumption patterns. This spatial analysis, combined with the temporal analysis offered by ML, provides a powerful means of identifying distinct consumer groups and understanding the spatial dynamics of energy consumption. The use of techniques such as k-means clustering [2], [23] has been particularly effective in identifying groups of consumers with similar consumption profiles. However, this review also highlights the need for exploring alternative clustering methods [23] to address potential limitations and improve the accuracy of consumer segmentation. Table 1 shows the highlighted bullet points on the literature review.

III. Methodology: Data Acquision, Preprocessing And Model Development

This study employs a multi-stage methodology that integrates data acquisition, preprocessing, and model development using both ML and GIS techniques.

The first stage involves acquiring energy consumption data from a reliable source. In study, we utilize monthly electricity consumption data for individual consumers based on monthly frequently readings on meters. These data are complemented by geographic information including location coordinates of consumers, land use information, weather condition like mean temperature, humidity and other relevant spatial variables. Socioeconomic data, such as household income, size, and occupancy, are also collected to enrich the analysis and provide a more comprehensive understanding of the factors influencing energy consumption.

Here, information of dataset is collected from Maharashtra State Electricity Distribution Co ltd (MSEDCL), a public sector undertaking by Government of Maharashtra. Information set

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

consists of 2500 domestic consumers monthly electricity consumption (EC) data obtained from rural villages mostly lies in west side of Sinnar tehsil at the parallels of 19.85^0 N and meridians of 74.00^0 E geographical location, Maharashtra.

Table 1 Literature Review Summary

Category	Technique/Method	Application	Description	Relevant Citations
Machine Learning (ML)	KM	Energy Demand Forecasting	Energy Forecasting: Demonstrated ability to model complex temporal dependencies and predict future energy demand with improved accuracy.	[4], [2]
	LSTM, RNN	Energy Demand Forecasting	ML models for forecasting range from simple linear regression to sophisticated deep learning architectures.	[7], [8], [9]
	Review	Socioeconomic Integration	Consumer Behavior Modeling: Enables identification of factors influencing energy use and development of personalized energy management strategies.	[10], [11], [12], [13]
	SVM, ANN, Bayesian Network	Consumer Behavior Modeling	Various ML algorithms used to understand the interplay between consumer choices and energy consumption.	[14], [15]
	MLaaS	Scalable Solutions	Machine Learning as a Service (MLaaS): Emerged as a promising approach for scalable and flexible solutions.	[16]
Geographic Information Systems (GIS)	GIS	Spatial Analysis	Provides tools to visualize and analyze the spatial distribution of energy consumption data, revealing patterns and correlations.	[17], [18], [19]
	GIS	Mapping Energy Consumption	Used to map energy consumption across different geographic areas and identify hotspots of high energy use.	[20], [21]
	GIS	Mapping Energy Consumption	Investigates the influence of geographic factors on energy consumption.	[22]
	GIS	Socioeconomic Integration	Integration with socioeconomic data allows exploration of relationships between energy consumption and demographic characteristics.	[11], [10]

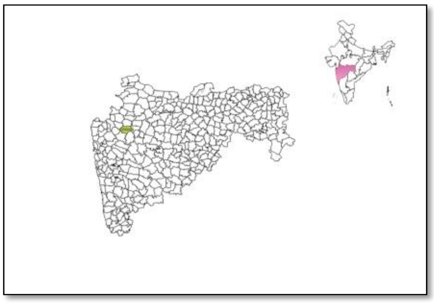
Received: July 18, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

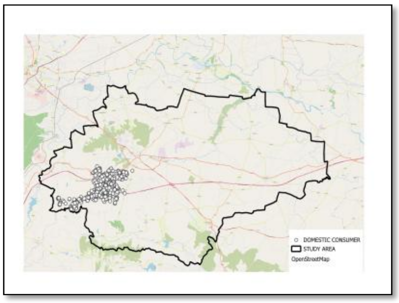
GIS & ML Integration	GIS, ML	Consumer Grouping	Combination with clustering algorithms allows for spatial segmentation of consumers based on energy consumption patterns.	[23], [24]
	KM	Spatial Segmentation	Segment consumers based on energy consumption patterns	[2], [23]

Fig 1 a) indicates the geographical location of Sinnar tehsil in Maharashtra, India map. Some consumer having improper information in consumption due to change of connectivity from single phase to three phase, shifted to net meter, faulty consumption and disconnection. So, only 1887 consumers' domestic EC data consider after removing these consumers information. Fig. 1 b) indicates the geographical information and location of the consumers to be consider for analysis, further Fig. 1 c) exhibits the analysis of information collected for categorization in proportion of consumer on the basis of their sanction load (kW). The analysis shows that around 64 % consumers fall below 0.5 kW, 31 % consumers come under the category of 0.5-1 kW sanction load and almost 5 % consumers under the category of above 1kW of load.

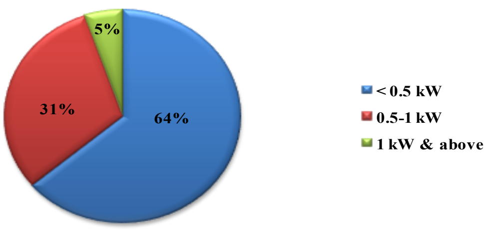
Energy Consumption (EC) pattern of consumer varies with changing weather and seasons conditions. Fig 2 a) and b) shows the change in EC patterns with different months and seasons as per weather and temperature conditions. As usual maximum average consumption of consumers is identified in the summer season whereas low consumption found in winter season. The average temperature and humidity data are gathered from website https://www.timeanddate.com/weather/india/nashik/climate.



a) Study Location Map of Sinnar Tehsil, Maharashtra State, India



b) Study Area with domestic consumers locations



c) Proportion of Consumers Sanction Load

Fig. 1. Study Location Map, Domestic Consumers locations and Proportion of Consumers Sanction Load

Received: July 18, 2025

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

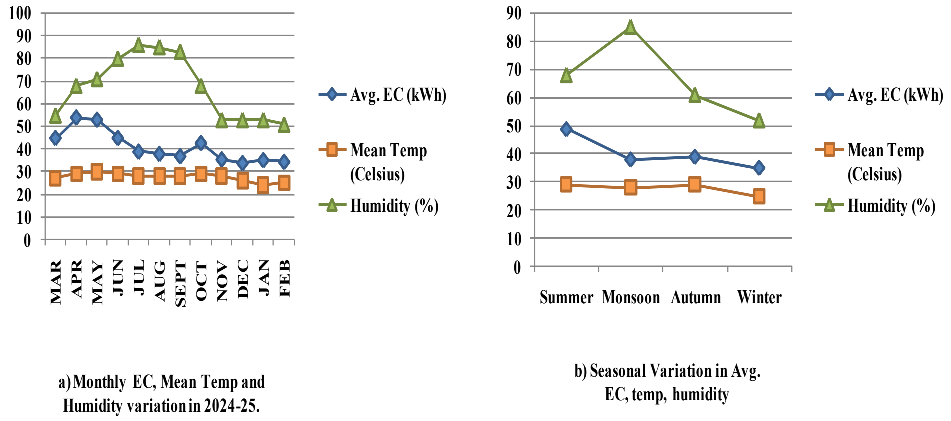


Fig. 2. Variation in Avg. EC, Mean Temp and Humidity as per Month and Season in 2024-25

The second stage focuses on data preprocessing, a critical step to ensure the quality and reliability of the analysis. This involves handling missing values, which may arise due to meter malfunctions or data transmission issues. Several techniques can be employed; including imputation methods that estimate missing values based on existing data patterns [25]. Outlier detection and treatment are also essential, as outliers can skew the results and negatively impact the performance of ML algorithms [26]. Robust statistical methods are used to identify and manage outliers, either by removing them or replacing them with more representative values. Data normalization is employed to scale the data to a consistent range, ensuring that variables with different scales do not disproportionately influence the ML algorithms [25]. Finally, feature engineering may be necessary to create new variables from existing ones, potentially improving the performance of the ML models.

The third stage involves the development and application of ML models for classification and clustering. The choice of algorithms depends on the specific research questions and data characteristics.

Clustering is an unsupervised learning technique that group's data objects into two or more clusters. Each cluster shows a common property among its members. Here, the KM clustering algorithm is used to form group of consumers on the basis of their monthly energy consumption pattern characteristics. Aim of K-means clustering is to minimize the within class sum of square error (WCSS) and it calculated using following equation (1).

$$WCSS = \sum k(\sum j(||xj-ck||)^2) \qquad (1)$$

Where:

k – No of clusters

j- No of data points within cluster k.

xj – data point.

ck – centroid of cluster k.

Received: July 18, 2025

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$(\|xj-ck\|)^2$ – squared Euclidean distance between data point and centroid .

Fig 3 show the evaluation methods for finding optimal value of K for K-Means clustering fig. a) and b) respectively represents the pictorially results of Elbow method and Silhouette score to determine the number of clusters for clustering consumers category. From a) it seems that graph smoothly descent at elbow point and in this case it is seen at point 3rd. To verify the obtained results, we explore another Silhouette Score method. From b) Cluster number 2 gives highest score whereas 2 is too small and second highest score being on cluster number 3.

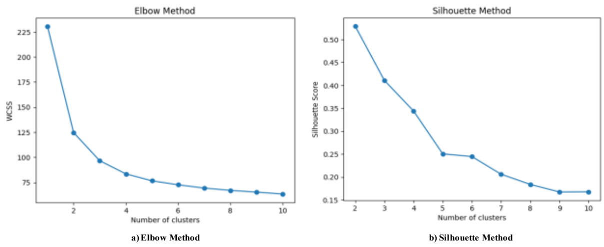


Fig. 3. Evaluation Methods for Optimal Value of K for Clustering

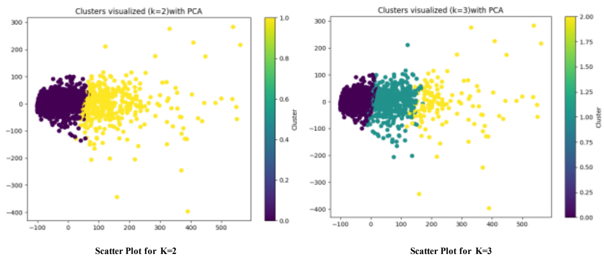


Fig. 4 Scatter plot of clustering with PCA

At the time of clustering process, Principal Component Analysis (PCA) is integrated to enable the efficient categorisation of consumption patterns on the basis of similarity scores obtained from PCA during the fitting process. The PCA algorithm enhances the performance of the clustering results as shown in fig 4.

For classification, algorithms such as random forests (RF), support vector machine (SVM), Xgboost, K-Nearest Neighbors (KNN), Linear Regression (LR), decision trees (DT), Naive Bayes (NB) can be considered. For clustering, k-means, hierarchical clustering, and density-based clustering algorithms are potential candidates [23], [2]. The performance of these models is evaluated using appropriate metrics, such as accuracy and precision to select the best-performing model for the specific task. Accuracy determines the overall correctness of model's predictions whereas Precision measures the accuracy of positive predictions. These performances parameters calculated using following equation (2) and (3).

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \qquad (2)$$

$$\text{Precision} = TP / (TP + FP) \qquad (3)$$

Where:

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

TP – True Positive (Rightly predicted positive instances)

TN – True Negative (Rightly predicted negative instances)

FP – False Positive (Wrongly predicted positive instances)

FN – False Negative (Wrongly predicted negative instances)

The selected model is then used to classify and cluster consumers based on their energy consumption profiles, spatial distribution, and socioeconomic characteristics.

The final stage integrates GIS tools to visualize and analyze the spatial patterns of energy consumption. The results of the ML clustering are mapped onto a GIS platform, allowing for the visualization of the spatial distribution of different consumer groups. Spatial analysis techniques, such as spatial autocorrelation analysis and hotspot analysis, are employed to identify clusters of consumers with similar energy consumption patterns and investigate potential correlations with geographic features or socioeconomic factors. The integration of GIS and ML provides a powerful framework for understanding the spatial dynamics of energy consumption and identifying areas for targeted energy efficiency interventions.

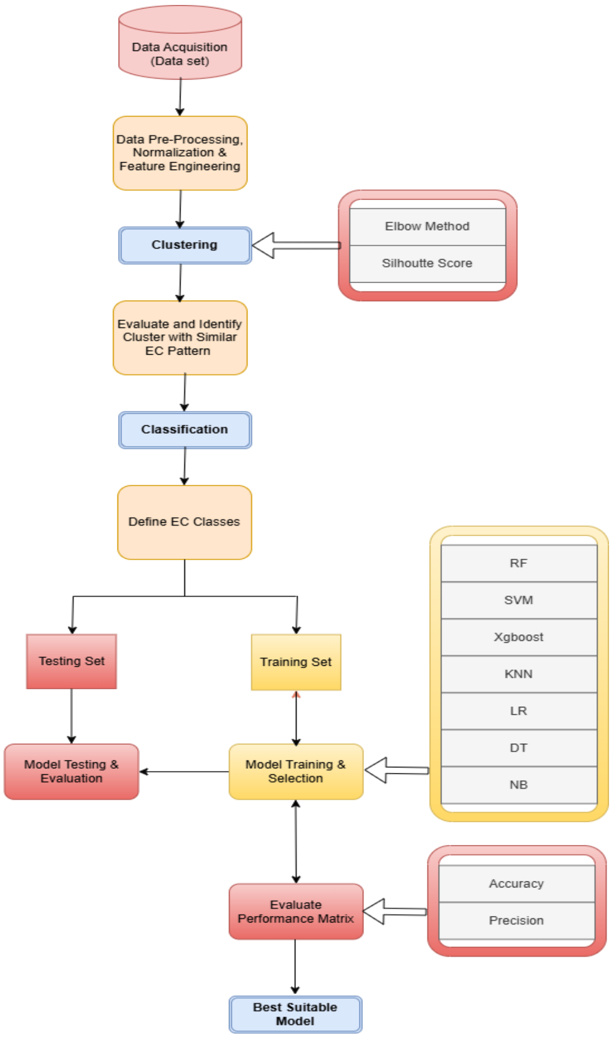


Fig.5) Proposed Workflow Diagram

Received: July 18, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

IV. RESULTS AND DISCUSIION: ANALYSIS OF CONSUMER GROUPS AND SPATIAL PATTERNS

Fig. 5 shows the proposed workflow for the study analysis. This section discusses the performance of the algorithms designed on the basis of accuracy and precision. Consumers are categorized according to their Energy Consumption Pattern using the (K=2,3) KM clustering method and classification algorithms, namely RF, SVM, XGBoost, KNN, LR, DT, NB approaches. The classifiers are then used to externally validate the classifications.

A. Clustering Results and Analysis

Consumers were classified based on their Electricity Consumption Patterns (ECP) into specific categories using the Monthly Electricity Consumption Patterns (MECP) approach. The gather Energy Consumption Pattern data of 1887 customers are analyzed using the K-means technique.

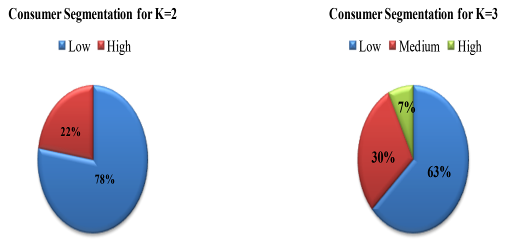
Table 2 Average consumption Range of consumer

Range of Avg. Consumption (in kWh)	No of Consumers	Percentage
0-30	819	43.40
31-60	723	38.31
61-90	247	13.09
91-120	62	3.29
121-150	22	1.17
151-180	10	0.53
Above 180	4	0.21

The analysed results for two and three class categorizations of consumers, on the basis of their monthly EC Pattern over a one year period, are depicted in Figures 4. Figure 7 illustrates a minimal overlap of 3.13%, involving in the two class categorization of two clusters. A total of 1463 consumers were identified as having low-level electricity consumption patterns, ranging from 14 to 61 kWh per month, while the remaining 424 consumers were classified under the high-level consumption category, with usage between 52 and 193 kWh per month. Figure 6(a) indicates that 78% of consumers fall into the low-level category, whereas 22% are in the high-level consumption category. Concurrently, the K-means algorithm results for the three-class classification reveal that 1195 consumers with an electricity consumption of 14–44 kWh/month are categorized into category-1, 555 consumers with consumption between 39 and 90 kWh/month are placed in category-2, 137 consumers with a consumption range of 78–193 kWh/month are assigned to category-3. Figure 6(b) demonstrates that the three consumer classes exhibit a consumption percentage of 63% for low-level, 30% for medium and 7% for high-level consumption. Table 3 shows that increasing the number of clusters leads to an unequal distribution of consumption patterns across groups. Furthermore, increasing the number of clusters results in an insufficient number of consumers in one or more clusters, raising questions about its proportionality. Table 4 reveals that the two and

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

three class contains an overlapped cluster with similar data points in both classes. Additionally, in the three-class clustering analysis, there is a very low overlap between the medium and high categories, thereby enhancing clustering performance. Therefore, consumer clustering should aim for greater similarity within clusters and greater divergence between clusters.



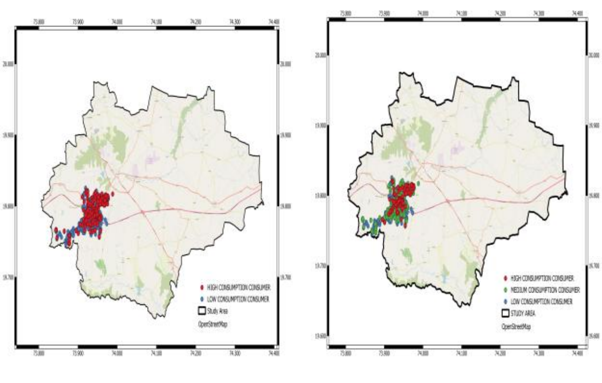
a) Percentage of Consumers for K = 2 b) Percentage of Consumers for K = 3

Fig. 6) Proportion of Consumers in Clustering

Table 4 No of Consumer Overlap in Clustering

Cluster	Overlap label	Overlapping consumers	% overlap
2	Low-High	59	3.13
3	Low-Medium	70	3.71
	Medium-High	28	1.48

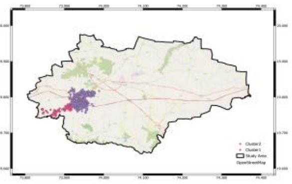
The efficacy of the clustering algorithm is assessed using Table 2. This table reveals that the distribution of consumer consumption patterns indicates the K-means algorithm's ability to effectively differentiate consumption patterns into distinct categories. This capability is further corroborated by the overlap of consumption patterns for each clustering algorithm, as depicted in Figures 7, which illustrate two and three-class consumption categorizations, respectively. Consequently, the results of the K-means algorithm are considered for the segregation of customers on the basis of the EC Pattern.



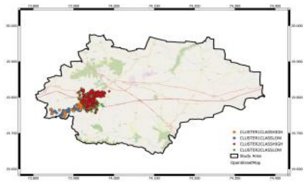
a) Spatial Location of Consumer Class for K = 2 a) Spatial Location of Consumer Class for K = 3

Fig.8) Based on the Energy Consumption Pattern Spatial Location of Cluster Consumer

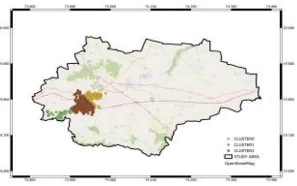
International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



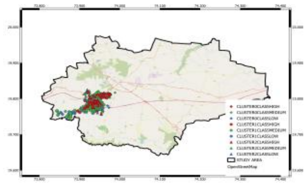
a) Spatial Location Cluster for K = 2



b) Spatial Location of Consumer Class



c) Spatial Location Cluster for K = 3



d) Spatial Location of Consumer Class

Fig.9) Clustering and Consumer Class based on Spatial Data

Figure 8 shows the spatial location of consumers in both cluster forms based on the energy consumption pattern where the information and statistical data of consumers as per geospatial information indicates that the consumers belong to villages mostly located in the west direction of sinnar tehsil. Basically most of consumers who has been consuming low and medium pattern mostly lives in hilly areas of Sahyadri mountain range.

Based on the spatial data, K Means clustering were carried out with the help of processing tool available in QGIS. Fig 9 a) and c) indicates the cluster formation based on the location using k =2 and 3. The results and class formed based on the energy consumption pattern were used to indicate in the cluster formed on the basis of geospatial location as shown in fig 9 b) and d).

Table 3 Consumer clustering by K-Means Technique

K-Means Clustering for Consumer Segmentation						
Cluster	Cluster	Label	No of Consumers	Percentage of Consumers	Min (Avg. kWh)	Max (Avg. kWh)
K = 2	0	Low	1463	77.53	14	61
	1	High	424	22.47	52	193
K = 3	0	Low	1195	63.33	14	44
	1	Medium	555	29.41	39	90
	2	High	137	7.26	78	193

International Journal of Applied Mathematics

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

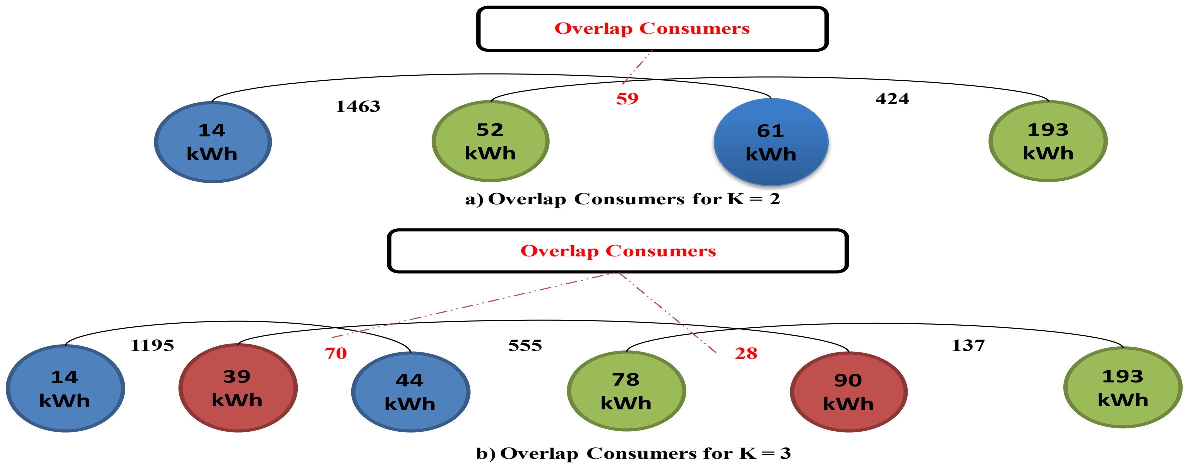


Fig. 7) Overlap Consumers

The results of the hypothetical study reveal three distinct consumer groups with unique energy consumption profiles. Group 1, representing approximately 7.26% of the consumers, exhibits consistently high energy consumption throughout the year, with a significant peak during the summer months.

This group is predominantly comprised of larger households with high incomes, located in Bungalows, Farmhouse with single family with large members. Their high energy consumption is likely driven by cooling, heating and other energy-intensive appliances.

Group 2, comprising about 29.41% of the consumers, show medium energy consumption with a noticeable peak during the summer months. This group is primarily composed of medium-sized households like kutcha, pucca houses with moderate incomes, residing in gavthan colonies or smaller single-family homes with some energy-efficient features. Their energy consumption is largely driven by cooling and other household appliances.

Group 3, accounting for 63.33% of the consumers, display consistently low energy consumption throughout the year. This group consists mainly of smaller households such as kutcha, huts with lower incomes, residing in wadi vasti, gavthan colonies with low connected load. Their low energy consumption reflects reduced energy demand. Further in this group most of consumers overlapped due to highly variable energy consumption with seasonal use pattern and diverse household sizes and incomes, and their consumption pattern.

The spatial analysis using GIS reveals interesting spatial patterns. Group 1 (high consumption) is largely concentrated in certain neighborhoods characterized by larger homes and higher property values. Group 3 (low consumption) is basically belong to BPL consumers predominantly located in areas with higher population density and a greater prevalence of wadi vasti, gavthan colonies. Group 2 (Medium consumption) is more evenly distributed across the study area. These spatial patterns highlight the influence of both socioeconomic factors and geographic location on energy consumption.

International Journal of Applied Mathematics

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

B. Classification Results and Analysis

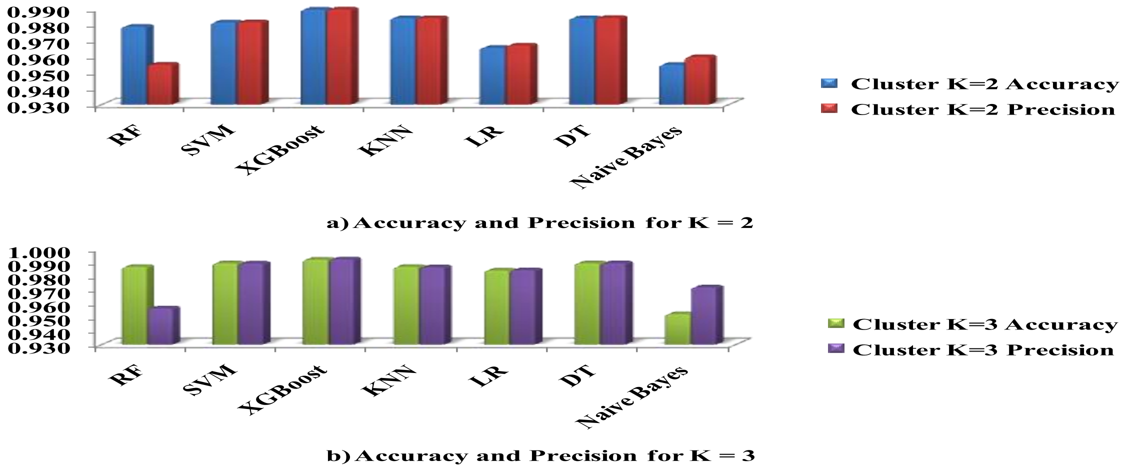


Fig. 10) Accuracy and Precision of different classifiers for K = 2 and K = 3

Table 5 Performance parameters of different classifiers

CV Technique Results

	Cluster K=2		Cluster K=3	
Classifier	Accuracy	Precision	Accuracy	Precision
RF	0.979	0.955	0.987	0.957
SVM	0.981	0.982	0.989	0.989
XGBoost	0.989	0.990	0.992	0.992
KNN	0.984	0.984	0.987	0.986
LR	0.966	0.967	0.984	0.984
DT	0.984	0.984	0.989	0.989
Naive Bayes	0.955	0.960	0.952	0.972

In this section, the effectiveness of the designed classifiers in consumer classification using the ECP is discussed. RF, SVM, XGBoost, KNN, LR, DT, NB techniques are the classification algorithms that are designed to separate the consumer consumption classes. In the present study, the significance of consumer EC patterns is substantiated by assessing the performance of seven algorithms based on their classification capacity for electricity consumption data. This research employs two class and three class categorization of EC Pattern for study using classification techniques.

The classifier performance has been evaluated using 10-fold Cross-Validation techniques. Here 10 fold Cross Validation technique employed because of small data set and to maximize the data set for training and evaluation. Table 5 indicates the performance of different classifiers for both two and three class clustering. The XGBoost has the highest 98.9 % accuracy and 99 % precision for two class clustering whereas from fig 10 it is clear that the same classifier meanwhile XGBoost has achieved the 99.2 % accuracy and precision for three class classification.

Received: July 18, 2025

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

It is important to acknowledge the limitations of this hypothetical study. The accuracy of the results depends on the quality and completeness of the data, and the selection of ML algorithms and parameters can influence the outcome. Furthermore, the hypothetical nature of the data limits the generalizability of the findings to other geographic areas or contexts. However, this study serves as a proof of concept, demonstrating the potential of the integrated ML-GIS approach for analyzing energy consumption patterns and identifying distinct consumer groups with unique characteristics.

V. Conclusion: Implications And Future Research Directions

This study demonstrates the significant potential of integrating machine learning (ML) and Geographic Information Systems (GIS) for granular analysis of electrical energy consumption. The electricity consumption patterns of the Low Voltage customers of the villages belong to Sinnar Tehsil are categorized. In this work, the EC Pattern is appropriately splits into two and three clusters using the K-means clustering technique. To determine which cluster would be best for evaluation, the study used the elbow and silhouette score method. The K-means algorithm results are taken into account for additional classification because they show little overlap. With the help of QGIS tool and spatial data; consumer clusters are formed based on the geographical location. Observing the geographical information; it is clear that the EC pattern depends on the socioeconomic and geospatial factor. Seven classification algorithms are constructed using the clustering outcomes of the K-means algorithm after the EC Pattern of consumers has been categorized. The analysis of the performance of the classifier on EC data showed that the XGBoost approach achieved 98.2 % accuracy for two class and 99.2 % accuracy for three class classification, which demonstrate that XGBoost performed better than other classifiers.

The benefits from this work is to identify distinct domestic consumer groups with unique energy consumption profiles allows for the development of targeted energy efficiency interventions and demand-side management (DSM) strategies. By tailoring interventions to the specific needs and characteristics of each group, policymakers and energy providers can maximize the effectiveness of their programs and achieve greater energy savings.

Future research should focus on several key areas. The exploration of more sophisticated ML algorithms, such as deep learning models or ensemble methods could further improve the accuracy and interpretability of the results. The integration of additional data sources, such as appliance-level consumption data, and House characteristics, can enhance the understanding of the factors influencing energy consumption. The application of the methodology to other geographic areas and energy consumption contexts will help to assess the generalizability of the findings. Furthermore, research should investigate the ethical considerations of using smart meter data for consumer profiling and the importance of protecting consumer privacy. Finally, the development of robust and scalable platforms for implementing the proposed methodology is essential for widespread adoption and application.

The integrated ML-GIS approach presented in this study offers a powerful tool for enhancing energy consumption analysis and informing the development of more effective energy

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

management strategies. By providing a more granular understanding of consumer behavior and the spatial dynamics of energy consumption, this approach can play a significant role in achieving greater energy efficiency, reducing greenhouse gas emissions, and promoting a more sustainable energy future.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support from Maharashtra State Electricity Distribution Company (MSEDCL) for the important inputs and actual consumption data to proceed the analysis on domestic consumers in study area.

REFERENCES

[1] Shin, Changho, Lee, Eunjung, Han, Jeongyun, Yim, Jaeryun, Rhee, Wonjong, and Lee, Hyoseop. 2019. "The ENERTALK dataset, 15 Hz electricity consumption data from 22 houses in Korea". Nature Portfolio. https://doi.org/10.1038/s41597-019-0212-5

[2] Khan, A., Iqbal, Naeem, Rizwan, Atif, Ahmad, Rashid, and Kim, Do-Hyeun. 2021. "An Ensemble Energy Consumption Forecasting Model Based on Spatial-Temporal Clustering Analysis in Residential Buildings". Energies. https://doi.org/10.3390/EN14113020

[3] Staffell I, Pfenninger S. The increasing impact of weather on electricity supply and demand. Energy 2018;145:65–78. http://dx.doi.org/10.1016/j.energy.2017. 12.051

[4] Mohapatra, Sunil Kumar, Mishra, Sushruta, Tripathy, Hrudaya Kumar, Bhoi, Akash Kumar, and Barsocchi, Paolo. 2021. "A Pragmatic Investigation of Energy Consumption and Utilization Models in the Urban Sector Using Predictive Intelligence Approaches". Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/en14133900

[5] Peng, Jieyang, Kimmig, Andreas, Wang, Jiahai, Liu, Xiufeng, Niu, Zhibin, and Ovtcharova, Jivka. 2021. "Dual-stage attention-based long-short-term memory neural networks for energy demand prediction". Elsevier BV. https://doi.org/10.1016/j.enbuild.2021.111211

[6] R. R. Patil, M. K. Gupta and K. V. Bhadane, "Electrical Energy Forecasting of Industrial Feeder from Rural Area Using Different Regression Techniques," 2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET), Pune, India, 2025, pp. 1-5, doi: 10.1109/ICAET63349.2025.10932244.

[7] Lee, Yujin, Tay, K.G., and Choy, Y. Y.. 2018. "Forecasting Electricity Consumption Using Time Series Model". None. https://doi.org/10.14419/ijet.v7i4.30.22124

[8] Javaid, Ali, Javaid, Umer, Sajid, Muhammad, Rashid, Muhammad, Uddin, E., Ayaz, Y., and Waqas, A.. 2022. "Forecasting Hydrogen Production from Wind Energy in a Suburban Environment Using Machine Learning". Energies. https://doi.org/10.3390/en15238901

[9] Jamil, Faisal, Iqbal, Naeem, Imran, Imran, Ahmad, Shabir, and Kim, DoHyeun. 2021. "Peer-to-Peer Energy Trading Mechanism Based on Blockchain and Machine Learning

Received: July 18, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

for Sustainable Electrical Power Supply in Smart Grid". Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/access.2021.3060457

[10] Frederiks, Elisha R., Stenner, Karen, and Hobman, Elizabeth V.. 2015. "The Socio-Demographic and Psychological Predictors of Residential Energy Consumption: A Comprehensive Review". Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/en8010573

[11] Ali, Sharif Shofirun Sharif, Razman, Muhammad Rizal, Awang, Azahan, Asyraf, M. R. M., Ishak, M. R., Ilyas, R.A., and Lawrence, Roderick J.. 2021. "Critical Determinants of Household Electricity Consumption in a Rapidly Growing City". Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/su13084441

[12] Apipuchayakul, Nuttakrij and Vassanadumrongdee, Sujitra. 2020. "Factors Affecting the Consumption of Energy-Efficient Lighting Products: Exploring Purchase Behaviors of Thai Consumers". Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/su12124887

[13] Wall, W. P., Khalid, Bilal, Urbaski, M., and Kot, M.. 2021. "Factors Influencing Consumers Adoption of Renewable Energy". Energies. https://doi.org/10.3390/en14175420

[14] Ardabili, S., Abdolalizadeh, Leila, Mako, C., Torok, Bernat, and Amir, Mosavi. 2022. "Systematic Review of Deep Learning and Machine Learning for Building Energy". Frontiers in Energy Research. https://doi.org/10.3389/fenrg.2022.786027

[15] Singh, Shailendra and Yassine, Abdulsalam. 2018. "Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting". Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/en11020452

[16] Ribeiro, Mauro C. C., Grolinger, Katarina, and Capretz, Miriam A. M.. 2015. "MLaaS: Machine Learning as a Service". None. https://doi.org/10.1109/icmla.2015.152

[17] Martn, Jos Manuel Snchez, Gallego, Juan Ignacio Rengifo, and Blas-Morato, Roco. 2019. "Hot Spot Analysis versus Cluster and Outlier Analysis: An Enquiry into the Grouping of Rural Accommodation in Extremadura (Spain)". Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/ijgi8040176

[18] Brownstein, John S., Rosen, Hilary, Purdy, Dianne, Miller, James R., Merlino, Mario, Mostashari, Farzad, and Fish, Durland. 2002. "Spatial Analysis of West Nile Virus: Rapid Risk Assessment of an Introduced Vector-Borne Zoonosis". Mary Ann Liebert, Inc.. https://doi.org/10.1089/15303660260613729

[19] Corkeron, Maree, Norton, Robert, and Nelson, Paul N.. 2010. "Spatial analysis of melioidosis distribution in a suburban area". Cambridge University Press. https://doi.org/10.1017/s0950268809991634

[20] Rajasekhar, M., Raju, G. Sudarsana, Raju, R. Siddi, and Basha, U. Imran. 2018. "Data on artificial recharge sites identified by geospatial tools in semi-arid region of Anantapur District, Andhra Pradesh, India". Elsevier BV. https://doi.org/10.1016/j.dib.2018.04.050

Received: July 18, 2025

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[21] Camargo, Simone Cristina, et al.. 2014. "Implementation of a geographic information system (GIS) for the planning of beekeeping in the west region of Paran". Brazilian Academy of Sciences. https://doi.org/10.1590/0001-3765201420130278

[22] Eisen, Rebecca J., Eisen, Lars, and Lane, Robert S.. 2006. "PREDICTING DENSITY OF IXODES PACIFICUS NYMPHS IN DENSE WOODLANDS IN MENDOCINO COUNTY, CALIFORNIA, BASED ON GEOGRAPHIC INFORMATION SYSTEMS AND REMOTE SENSING VERSUS FIELD-DERIVED DATA". American Society of Tropical Medicine and Hygiene. https://doi.org/10.4269/ajtmh.2006.74.632

[23] Miraftabzadeh, Seyed Mahdi, Colombo, Cristian Giovanni, Longo, Michela, and Foiadelli, Federica. 2023. "K-Means and Alternative Clustering Methods in Modern Power Systems". Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/access.2023.3327640

[24] Zaane, Osmar R. and Lee, Chi-Hoon. 2003. "Clustering spatial data in the presence of obstacles: a density-based approach". Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ideas.2002.1029674

[25] Khan, Inam Ullah, Javeid, Nadeem, Taylor, C. James, Gamage, Kelum A. A., and Ma, Xiandong. 2021. "A Stacked Machine and Deep Learning-Based Approach for Analysing Electricity Theft in Smart Grids". Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/tsg.2021.3134018

[26] Sun, Li, Zhou, Kaile, Zhang, Xiaoling, and Yang, Shanlin. 2018. "Outlier Data Treatment Methods Toward Smart Grid Applications". Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/access.2018.2852759

[27] Gul, Hira, Javaid, Nadeem, Ullah, Ibrar, Qamar, Ali Mustafa, Afzal, Muhammad Khalil, and Joshi, Gyanendra Prasad. 2020. "Detection of Non-Technical Losses Using SOSTLink and Bidirectional Gated Recurrent Unit to Secure Smart Meters". Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/app10093151

[28] Wang, Yi, Chen, Qixin, Kang, Chongqing, Zhang, Mingming, Wang, Ke, and Zhao, Yun. 2015. "Load profiling and its application to demand response: A review". Tsinghua University Press. https://doi.org/10.1109/tst.2015.7085625

[29] Latifi, Milad, Khalili, Azam, Rastegarnia, Amir, Zandi, Sajad, and Bazzi, Wael M.. 2017. "A distributed algorithm for demand-side management: Selling back to the grid". Elsevier BV. https://doi.org/10.1016/j.heliyon.2017.e00457

[30] Paterakis, Nikolaos G., Erdin, Ozan, Bakirtzis, Anastasios G., and Catalo, Joo P. S.. 2015. "Optimal Household Appliances Scheduling Under Day-Ahead Pricing and Load-Shaping Demand Response Strategies". Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/tii.2015.2438534

[31] Chin, JunXing, Rubira, Toms Tinoco De, and Hug, Gabriela. 2017. "Privacy-Protecting Energy Management Unit Through Model-Distribution Predictive Control". Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/tsg.2017.2703158.

Received: July 18, 2025