

INTELLIGENT CLASSIFICATION OF RURAL AREA ELECTRICITY CONSUMPTION MANAGEMENT PROFILES USING FUZZY CLUSTERING AND MACHINE LEARNING

Durgadevi Velusamy¹, Ishwarya S^{2*}, Rajesh Kumar B³, Karthikeyan Ramasamy⁴

¹Department of Information Technology, SSN School of Engineering, Shiv Nadar University, Chennai, Tamil Nadu, 603110, India

^{2,3,4}Department of Electrical and Electronics Engineering, M. Kumarasamy College of Engineering, Karur, Tamil Nadu, 639113, India

Corresponding author(s): ISHWARYA S
ishwarya.s23@gmail.com

Abstract: The rising variability in household electricity usage behaviour needs robust data-driven consumer categorisation systems. This paper suggests using actual monthly electricity consumption data to develop a Fuzzy C-Means (FCM)-based framework for categorising residential electricity consumers. The dataset includes 48 months of power usage data of 1110 residential consumers from Puduppatti Village, Namakkal District, Tamil Nadu, India. The preprocessed and standardised consumption profiles were grouped with FCM for $K = 2$ to $K = 10$; clustering quality was evaluated based on the Silhouette Coefficient (SCI) and Davies–Bouldin Index (DBI). The results showed that $K = 2$ is the most appropriate clustering configuration with SCI of 0.4239 and DBI of 0.9520. The $K = 2$ solution classified consumers into Low Consumption (LC) and High Consumption (HC) groups. Subsequently, independent test-set evaluation and 10-fold stratified cross-validation were implemented to evaluate five machine-learning classifiers: K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), and Logistic Regression (LR). SVM attained the best cross-validation accuracy of 98.56% and independent test-set accuracy of 98.65%, with an AUC of 1.000. The results show that the combination of FCM and supervised learning is successful for categorisation of household power consumption and provides a platform for consumer profiling and tailored energy-management methods.

Keywords: Residential electricity consumption; Fuzzy C-Means; clustering; Silhouette Coefficient; Davies–Bouldin Index; machine learning; consumer categorisation.

1.INTRODUCTION:

The rise in residential electricity demand has highlighted the necessity for a comprehensive understanding of household electricity usage patterns. Population expansion, urbanisation, increasing electrification and the widespread use of electrical appliances have changed domestic energy consumption behaviour a lot. Growing use of electricity for heating, cooling, transportation and other energy intensive devices also adds to variances in residential electricity use. These evolving characteristics emphasise the importance of data-driven techniques to identify and classify different home electricity usage patterns [1,2].

In addition to the wide-area deployment of smart meters, enormous amounts of electricity consumption data may be collected at varied time intervals. These data give useful information about household load profiles and consumption behaviour and have motivated the application of data mining algorithms to discover distinct consumer patterns [3,4]. Demand side management, demand-response programs, energy-efficiency planning, load control and targeted energy saving methods can be facilitated by consumer segmentation based on comparable consumption characteristics.

Electricity consumption in homes varies greatly depending on the number of occupants, the lifestyle of the residents, how often appliances are used, work schedules and the time of year. Thus, customers in the same geographic location can have rather distinct purchasing characteristics. The average consumption-based study may not sufficiently represent these differences, and the time-dependent nature of power usage adds to the difficulty of finding relevant consumer groupings [5].

Clustering, an unsupervised data-mining technique, offers an efficient solution to discover groups of consumers with similar features of power usage, without the need of pre-existing labels. Residential power consumption data have been clustered by different clustering approaches such as K-Means, K-Medoids, Fuzzy C-Means, hierarchical clustering and density-based algorithms [6–9]. For example, [10] used smart-meter and socioeconomic data of 433 families to study household electricity-use

behaviour. The load curves for weekdays and weekends were classified using K-Medoids, and the Silhouette Coefficient was used to calculate the optimum number of clusters. [11] have suggested a similarity learning method based on DLDA and Affinity Propagation clustering, using power use data from 5566 London households. The obtained findings for ARI and AMI show the usefulness of the approach for the detection of different patterns of household electricity usage.

Among the clustering algorithms, the fuzzy c-means (FCM) is especially useful for the analysis of residential consumption since it provides a membership degree of each consumer for each cluster. As opposed to hard clustering, FCM can show users whose spending characteristics might overlap somewhat between groups. Therefore, the membership values provide additional information on the degree of linkage of consumers with different consumption categories and the highest membership value can be utilised to derive the final cluster assignment [9,12].

The representation of electricity usage data is also significant in the clustering analysis. In long-term consumption datasets, there may be several monthly variables with temporal fluctuations and connections. Dimensionality reduction techniques such as PCA and other non-linear methods have been examined to model home electricity consumption patterns [13]. However, the dimensionality reduction may alter the original consumption characteristics and may influence the resulting cluster structure. The focus of this work is therefore direct use of FCM on the original pre-processed and standardised consumption profiles.

As clustering is an unsupervised process, it is required to evaluate the clusters quality. The Silhouette Coefficient (SCI) assesses the compactness and separation of clusters, where the greater value means the better-defined clusters. The Davies-Bouldin Index (DBI) measures the similarity among the clusters. The lower the DBI value, the better the clustering outcome. Widely used measures for evaluation of clustering solutions include [6,9,12]. For example, [6] combined LTTB, DTW and K-Medoids on home electricity usage data and obtained a Silhouette Score of 0.5095, a Calinski–Harabasz Index of 991.45 and a Davies–Bouldin Index of 0.6976. But they only used data from one month and from one geographical location.

In addition to finding consumer groups, it is also necessary to see whether the groups obtained can be effectively separable by supervised learning. Machine-learning classifiers may be an efficient tool for the allocation of customers to pre-identified categories. To test the separability of consumer groups created from FCM, algorithms such as K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), and Naive Bayes (NB) can be utilised. The classification performance is evaluated by accuracy, precision, recall and F1-score in cross-validation and Independent Test-Set Evaluation.

Therefore, this study presents a home power consumption categorisation framework by applying FCM on original consumption data and machine-learning classification. The preprocessed and standardised consumption profiles are clustered by using FCM for $K=2$ to $K=10$ and the obtained solutions are evaluated by utilising Silhouette Coefficient (SCI) and Davies–Bouldin Index (DBI). The $K=2$ and $K=3$ solutions are further analysed to understand the resulting customer groups and the $K=2$ solution is then used to generate class labels for supervised classification. Classification performance is evaluated using five machine learning algorithms: KNN, SVM, LR, DT and NB, using cross-validation and Independent Test-Set Evaluation. The proposed framework combines fuzzy clustering and supervised classification for a data-driven approach to home electricity usage categorisation.

2. REVIEW OF LITERATURE

More recently, there has been a growing use of data-driven and machine-learning techniques to evaluate trends of home electricity consumption. Clustering has been routinely employed to distinguish groups of consumers with similar consumption characteristics and to analyse differences in household energy use behaviour.

In this study [1] high-resolution household power consumption data from Austria, Germany and the United Kingdom were investigated, and data-driven load profiles were created based on time-series method and stochastic modelling. The research improved the portrayal of the dynamics and volatility of home consumption, but consumer categorisation was not the focus of the work.

Author [3] identified the residential consumer characteristics utilising a federated learning framework with the electricity use and socio-demographic information. PCA and an ANN classifier were used. The

study was more on predicting consumer attributes rather than classifying consumers according to their electricity usage habits.

In this research [4] employed K-Means and K-Medoids on home electricity usage data to determine daily and annual load profiles. The detected profiles were afterward predicted by machine learning algorithms, showing the potential of the combination of clustering and classification for the analysis of home consumption.

Author [5] has proposed an LTTB-DTW-K-Medoids framework for home electricity usage pattern identification. The third study derived three clusters and employed metrics of clustering validation to assess the derived patterns of consumption. The work showed the importance of associating adequate similarity measures with clustering and validation. However, the analysis was performed for a relatively short consumption period.

The authors of study [10] examined the household electricity-use behaviour based on the smart-meter and socioeconomic data of 433 households. Consumer groups were identified using the K-Medoids algorithm based on weekday and weekend load curves, and a deep learning-based load behaviour model was developed. The study showed the utility of clustering combined with supervised learning in household energy-consumption analysis.

In [12], an enhanced Fuzzy C-Means (FCM) technique was developed for the recognition of household power use patterns. The study has showed that FCM can identify homes with varying consumption levels and daily usage patterns. The work supports the use of fuzzy clustering for the representation of the variations of the home power consumption behaviour.

In [9], K-Means, K-Medoids, Fuzzy C-Means and Gaussian Mixture Models were examined to profile home power demand. We assessed the resulting solutions using multiple clustering-validation indices. The study emphasised the need for comparison of hard and soft clustering methodologies and rigorous evaluation of cluster quality.

This research [8] proposed a personalised demand-side management architecture where the FCM-based clustering was paired with an LSTM model for customer-group prediction. This study shows the potential of fuzzy clustering integrated with supervised learning for residential energy applications.

The data of 4,438 homes in London were used to compare K-Means, K-Medoids, HAC and DBSCAN for residential load-profile clustering by author [7]. The clustering solutions were assessed by Silhouette, Davies–Bouldin and Calinski–Harabasz indices showing the significance of thorough cluster validation.

In this [13] smart-meter data from 5038 consumers in Slovenia was examined for power usage patterns. The consumer groups were identified via GMM, and the clustering structure was assessed by many validation methods such as the Silhouette and Davies–Bouldin indices.

To conclude, the reviewed studies show that clustering can effectively reveal heterogeneity in home power usage patterns. However, most studies relied on hard clustering approaches, particular data formats or restricted consumption periods. There has been comparably little attention paid to the direct application of FCM to long-term original residential consumption profiles. Despite the existence of several works that integrate clustering with supervised learning [19], the application of consumer groups formed from FCM as classification goals and their subsequent assessment using different machine learning methods remains less investigated. These deficiencies motivated the present work which examines FCM-based consumer categorisation based on original consumption profiles, cluster-validation measurements and subsequent machine-learning classification.

3. DATASET DESCRIPTION

The data set used in this study contains household power consumption records of 1110 consumers of Puduppatti Village, Namakkal District, Tamil Nadu, India. The dataset contains monthly electricity usage records for 48 months, starting in March 2020 and ending in February 2024. Each consumer is represented with a 48-month power usage profile. The monthly usage data is in kilowatt-hours (kWh) and the profile is utilised for each consumer. The dataset contains 48 monthly consumption variables, which are used as input features for the subsequent clustering analysis.

Table 1. Characteristics of the Residential Electricity Consumption Dataset

Parameter	Description
Study Area	Puduppatti Village, Namakkal District, Tamil Nadu, India
Consumer Category	Residential Electricity Consumers
Number of Consumers	1,110
Observation Period	March 2020 – February 2024
Duration of Data Collection	48 Months
Number of Consumption Features	48 Monthly Electricity Consumption Attributes
Data Type	Monthly Electricity Consumption (kWh)

The dataset offers a long-term picture of the behaviour of residential consumers regarding their usage of energy, as seen in Table 1. Figure 1 shows the distribution of users according to their sanctioned load. There are 707 clients who have a sanctioned load of 1 kW, followed by 264 consumers who have a load of 2 kW, and 99 consumers who have a load of 3 kW of electricity. Few clients are sanctioned loads of 4 kW and beyond. The sanctioned-load information is used simply to characterise the customer population and not as an input feature for the clustering study.

4. METHODOLOGICAL APPROACH

The suggested methodology involves data pre-processing, Fuzzy C-Means (FCM) clustering, cluster validation, supervised machine-learning classification with Support Vector Machine (SVM), and performance evaluation.

4.1 DATA PREPROCESSING

Monthly home electricity usage statistics were checked for missing and contradictory observations. The consumption variables that were retained were standardised using Z-score standardisation prior to clustering, and incomplete records were eliminated.

The standardised value is calculated by:

$$z = (x - \mu) / \sigma \quad 1$$

where x is the original consumption value, μ is the mean, and σ is the standard deviation.

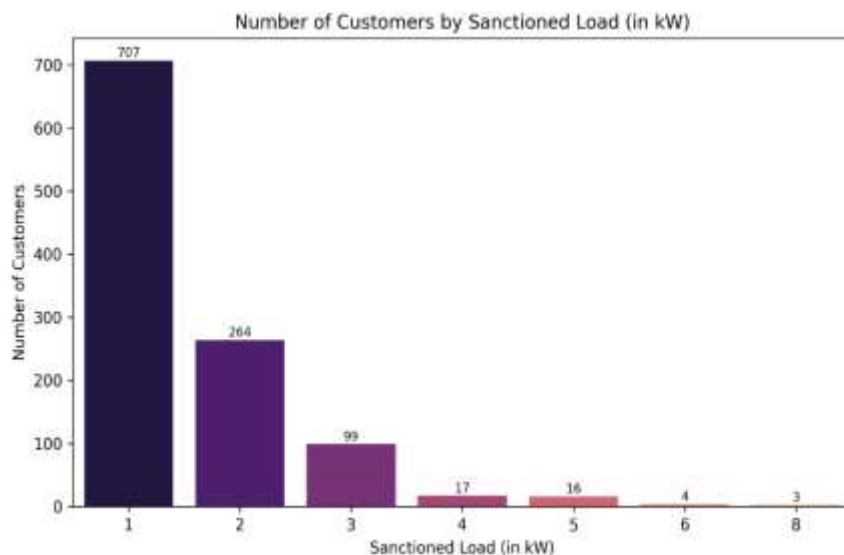


Figure 1. Number of Residential Consumers According to Sanctioned Load

4.2 FUZZY C-MEANS CLUSTERING

Fuzzy C-Means is a soft clustering approach where an observation might have varying degrees of membership in several clusters. This feature is helpful for residential power consumption data because the consumption behaviour of different consumers may not necessarily belong to completely different groups.

Let the input dataset be represented by $X = \{x_1, x_2, \dots, x_n\}$, where N denotes the total number of consumers and x_i represents the 48-month consumption profile of the i^{th} consumer. For a selected number of clusters C , FCM calculates a membership value u_{ij} , representing the degree to which consumer i is associated with cluster j .

$$\begin{aligned} 0 &\leq u_{ij} \leq 1 & 2 \\ \sum_{j=1}^C u_{ij} &= 1 & 3 \end{aligned}$$

The FCM optimisation process minimises the following objective function:

$$J = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m ||x_i - v_j||^2 \quad 4$$

where v_j is the centre of the j^{th} cluster, m determines the degree of fuzziness, and the distance term is the dissimilarity measure between a consumer profile and its associated cluster center.

In the iterative procedure, the membership matrix and the cluster centres are updated until the given convergence condition is satisfied. The output is a membership matrix that indicates the degree of membership of each consumer in each cluster. If required to obtain a single cluster label, the cluster with the highest membership value to a consumer is obtained as its final assignment. In this study, FCM is independently applied for $C = 2, 3, \dots, 10$ using the original pre-processed and standardised consumption profiles.

4.3 CLUSTER-QUALITY ASSESSMENT

Because the FCM procedure does not use predefined class labels, the quality of each clustering solution is assessed using internal validation criteria. Two complementary measures, namely SCI and DBI, are employed.

4.3.1 Silhouette Coefficient

The Silhouette Coefficient measures whether a customer is closer to consumers in its own cluster than to consumers in other clusters. The Silhouette value for the specific customer i is determined as:

$$S_i = (b_i - a_i) / \max(a_i, b_i) \quad 5$$

where a_i denotes the average dissimilarity of consumer i from the other consumers within its assigned cluster, and b_i represents the lowest average dissimilarity between consumer i and the consumers in any alternative cluster.

$$SCI = (1/N) \sum_{i=1}^N S_i \quad 6$$

The value of SCI lies between -1 and $+1$. A value approaching $+1$ indicates that the observations are well matched with their own cluster and clearly separated from neighbouring clusters. Therefore, a larger SCI represents a more desirable clustering structure. SCI is calculated for $K = 2$ to $K = 10$, and the resulting values are plotted to examine the variation in clustering quality with the number of clusters.

4.3.2 Davies–Bouldin Index

The Davies–Bouldin Index is a measure for assessing the quality of clustering that considers the compactness of individual clusters and their separation from neighbouring clusters. For two clusters i and j the relative similarity can be written as:

$$R_{ij} = (S_i + S_j) / M_{ij} \quad 7$$

where S_i and S_j represent the within-cluster dispersion measures, while M_{ij} denotes the distance between the corresponding cluster centres.

$$DBI = \left(\frac{1}{C}\right) \sum_{i=1}^C \max_{j \neq i} (R_{ij}) \quad 8$$

Hence, the smaller the DBI value, the better the clustering structure. This is because smaller DBI value means reduced within-cluster scatter and larger separation between clusters. In this study, DBI is calculated for $K = 2$ to $K = 10$ and compared with the corresponding SCI values.

4.4 SUPERVISED MACHINE-LEARNING CLASSIFICATION

The cluster labels obtained from FCM were used as target classes for supervised classification. The examined models were K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT) and Logistic Regression (LR). The evaluation was on an 80:20 stratified independent test-set. In addition, stratified 10-fold cross-validation was performed. The performance is determined by Accuracy, Precision, Recall and F1-Score.

4.4.1 Support vector machine learning

SVM is a supervised learning algorithm that is applied to classification and regression analysis. It is a statistical learning method that creates a high-dimensional feature space from input points. The categorization of the data points considers the extreme data points, known as support vectors, lying on the hyperplane, which are mapped using a nonlinear function to determine the decision boundary. If the distance between the two support vectors on each side, i.e., the ideal hyperplane or margin, is maximized, then the two classes are linearly separable. The SVM method has better generalization and reduces error by increasing the margin of the hyperplane.

The dataset D consists of N records and M features, where x represents the independent feature $x \in \mathbb{R}^d$, and y represents the dependent feature $y \in \{0,1\}$.

The decision boundary that linearly separates the data points is represented as:

$$xw^T + b = 0 \quad 9$$

where $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$ are and w is weight b is bias, respectively.

The decision function, which is applied to the unseen test data to predict the class of the given sample, is defined as:

$$f(x) = \text{sign}(xw^T + b) \quad 10$$

The margin of the SVM depends on the linear separability of the data. If the data is not linearly separable, the geometric margin is maximized and is represented as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^N \xi_i \quad 11$$

subject to

$$y_i(x_i^T w + b) \geq 1 - \xi_i, \forall i \in \{1, \dots, N\} \quad 12$$

where $\xi_i \geq 0$ is the penalty factor and C is the regularization parameter For non-linearly separable data, the soft-margin SVM is represented as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad 13$$

subject to

$$y_i(x_i^T w + b) \geq 1 \forall i \in \{1, \dots, N\} \quad 14$$

4.4.2 Classification Performance Measures

Classification performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC.

$$\text{Accuracy} = \frac{C_{TP} + C_{TN}}{C_{TP} + C_{TN} + C_{FP} + C_{FN}} \quad 15$$

$$\text{Precision} = \frac{C_{TP}}{C_{TP} + C_{FP}} \quad 16$$

$$\text{Recall} = \frac{C_{TP}}{C_{TP} + C_{FN}} \quad 17$$

$$\text{F1 score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad 18$$

Here, C_{TP} , C_{TN} , C_{FP} and C_{FN} denote the correct positive, correct negative, incorrect positive, and incorrect negative classifications, respectively.

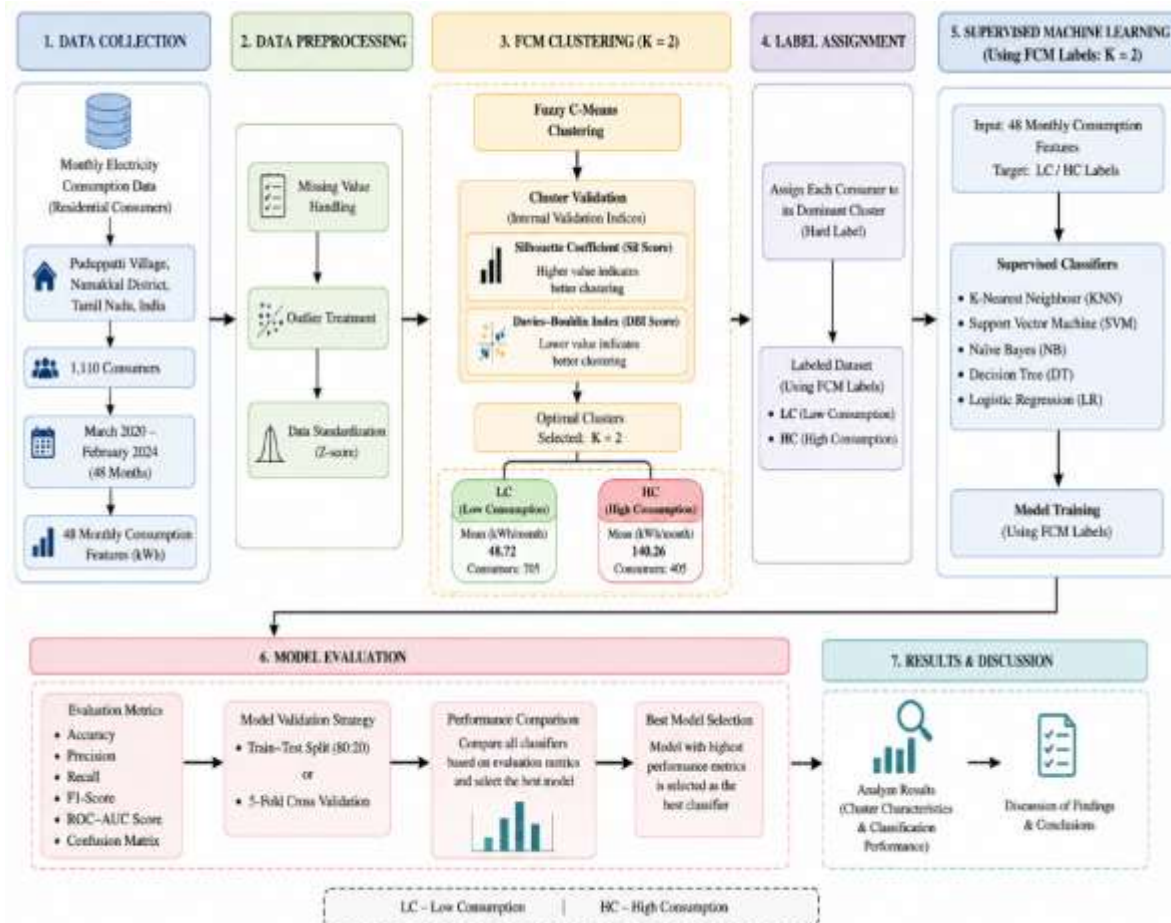


Figure 2. Overall methodology for residential electricity consumption profile classification

5. RESULTS AND DISCUSSION

In this section the results derived from the clustering and supervised classification of home electricity usage profiles are presented. The internal cluster validation indices are initially applied to evaluate the performance of the Fuzzy C-Means (FCM) algorithm in clustering. These cluster assignments are then used as target labels for a supervised machine learning classification. Subsequently, the performance of classification models is evaluated with the chosen evaluation metrics and the cross-validation findings.

5.1 FCM CLUSTERING PERFORMANCE

The clustering performance of the Fuzzy C-Means (FCM) algorithm was tested for varied number of clusters K ($K = 2$ to $K = 10$) by Silhouette Coefficient (SCI) and Davies–Bouldin Index (DBI). The acquired values are tabulated in Table 2 and the variations of SCI and DBI in relation to the number of clusters are shown in Figures 3 and 4 accordingly. The two-validation metrics provide complementary information on the quality of the clustering structure, where a greater SCI and lower DBI suggest better defined clusters.

The maximum Silhouette Score produced by the FCM algorithm was 0.4239 for $K = 2$ (Table 2), which reflects the best separation and internal compactness among the explored clustering options. The Silhouette Score is gradually decreasing from 0.4239 at $K = 2$ to 0.1066 at $K = 10$ as shown in Figure 3. This downward trend shows that by increasing the number of clusters, the separation of the consumer groups in the original consumption is not improved. For Davies-Bouldin Index, a similar pattern is found. The minimum value of DBI (0.9520) is attained at $K=2$ and then the value of DBI increases steadily with the number of clusters reaching the value of 2.5386 at $K=10$. The increase in DBI (Figure 4) demonstrates that the more the number of clusters increases, the poorer the cluster compactness and separation are.

Table 2. FCM Clustering Validation Results for Different Numbers of Clusters

Clusters (K)	Sil Score	DBI
2	0.4239	0.9520
3	0.3264	1.1338
4	0.2572	1.2661
5	0.2158	1.4513
6	0.1789	1.6821
7	0.1490	2.0152
8	0.1306	2.2787
9	0.1155	2.4965
10	0.1066	2.5386

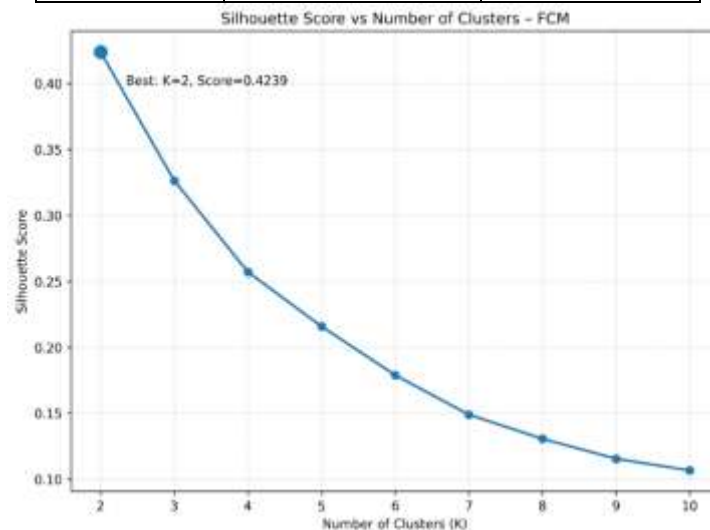


Figure 3. Silhouette Score versus number of clusters for FCM

The concordance between both validation metrics strongly supports the selection of K=2 as the best clustering configuration for the FCM study. The solution of K = 2 has the largest value of SCI (0.4239) and the smallest value of DBI (0.9520) among all the numbers of clusters under evaluation. As the main clustering solution, the K = 2 is picked for the ensuing classification study, while the K = 3 solution is also considered to explore whether a finer classification of residential consumers may give relevant consumption groupings.

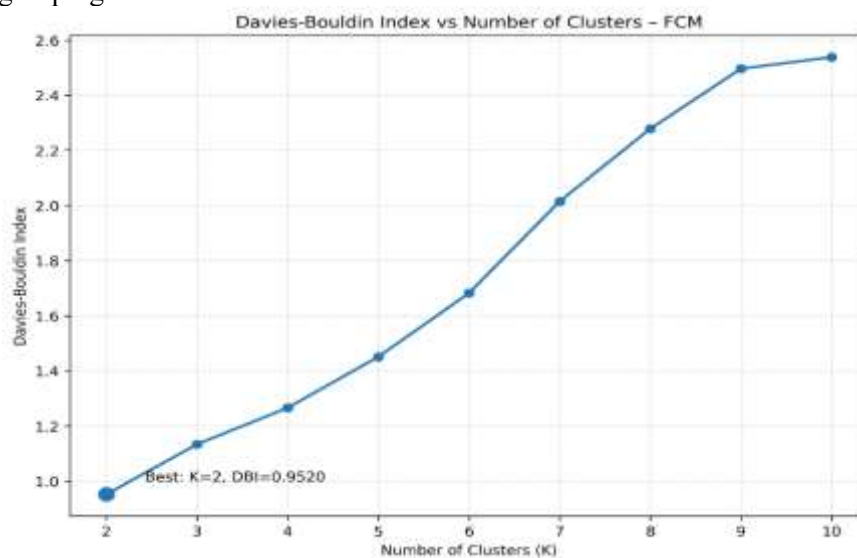


Figure 4. Davies–Bouldin Index versus number of clusters for FCM

5.2 ANALYSIS OF K = 2 AND K = 3 CLUSTERS

The results of the FCM clustering for K=2 and K=3 are displayed in Table 3. The consumption categories were based on the mean monthly electricity use of the obtained clusters, not their cluster names. For K = 2, the first cluster with an average consumption of 48.72 kWh/month, was classified as Low Consumption (LC) with 705 users, the second with a higher average consumption of 140.26 kWh/month, was classified as High Consumption (HC) with 405 consumers

Table 3. Characteristics of Consumer Clusters Identified by FCM for K = 2 and K = 3

FCM Configuration	Consumption Category	Mean Consumption (kWh/month)	Number of Consumers
K = 2	LC – Low Consumption	48.72	705
	HC – High Consumption	140.26	405
K = 3	LC – Low Consumption	37.64	517
	MC – Medium Consumption	96.61	424
	HC – High Consumption	181.81	169

By increasing the number of clusters to K = 3, the consumption profiles might be further separated into three categories. The LC group had the lowest mean consumption of 37.64 kWh/month and consisted of 517 people. The MC group was composed of 424 users with a mean consumption of 96.61 kWh/month, and the HC group consisted of 169 consumers with the greatest mean consumption of 181.81 kWh/month. The K = 3 method thus gives a finer classification of residential customers by separating the intermediate consumption group from the low and high consumption groups. However, the clustering validation results in Section 5.1 indicate that K = 2 was picked as the main FCM solution as it attained the maximum SCI and lowest DBI at the same time. The cluster labels for K = 2 are then employed in the machine-learning classification analysis.

5.3 MACHINE-LEARNING CLASSIFICATION RESULTS

The class labels used for the supervised classification analysis were obtained from the K = 2 FCM clustering solution. The two consumer groups, Low Consumption (LC) and High Consumption (HC) had 705 and 405 people, respectively. Five machine-learning classifiers, namely K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT) and Logistic Regression (LR) were evaluated using 10-fold stratified cross-validation and independent test-set evaluation. The classification performance was evaluated based on the Accuracy, Precision, Recall, and F1-Score.

Table 4. Performance Comparison of Machine-Learning Classifiers for K = 2 FCM-Derived Consumer Categories

Classifier	Evaluation	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
KNN	10-Fold CV	96.85	98.5	92.85	95.54
	Test Set	97.3	98.7	93.83	96.2
SVM	10-Fold CV	98.56	98.12	98.03	98.04
	Test Set	98.65	100	96.3	98.11
NB	10-Fold CV	97.12	93.96	98.77	96.24
	Test Set	97.75	95.24	98.77	96.97
DT	10-Fold CV	93.87	92.46	93.08	92.17
	Test Set	91.89	90.91	86.42	88.61

LR	10-Fold CV	97.12	95.89	96.3	96.07
	Test Set	98.65	100	96.3	98.11

Table 4 shows the comparative performance of classification. SVM gave the best accuracy of 98.56% followed by NB and LR with 97.12% and KNN with 96.85% under 10-fold stratified cross validation. DT had the lowest accuracy of 93.87%. The SVM also got the greatest F1-Score of 98.04% which shows a good balance between precision and recall. NB has the highest recall value of 98.77% but relatively low precision value of 93.96%.

For independent test-set evaluation SVM and LR achieved the greatest accuracy of 98.65% followed by NB (97.75%) and KNN (97.30%). DT again achieved the lowest performance with 91.89% accuracy. SVM and LR obtained F1-Score of 98.11%, but SVM acquired precision of 100 %. NB had the maximum recall at 98.77%, which shows its capacity to correctly identify a large majority of consumers belonging to the positive class. For both assessment methods, SVM exhibited the most stable classification results in general.

The confusion matrix for the SVM classifier is depicted in Figure 5. A total of 141 LC consumers and 78 HC consumers were successfully categorised out of 222 independent test samples, with only 3 HC consumers misclassified as LC. No LC users were misclassified as HC. This demonstrates that the SVM classifier is performing well in classifying the two consumption categories obtained by FCM.

Receiver Operating Characteristic (ROC) curves for the five classifiers are shown in Figure 6. SVM and LR had an AUC of 1.000 showing good discriminating between LC and HC categories. NB also showed a very high AUC of 0.998 followed by KNN with 0.986. DT showed AUC value of 0.907 which was somewhat lower but still exhibited high discrimination ability. Moreover, the ROC analysis confirms the high ability of SVM and LR to classify the consumer groups produced from FCM.

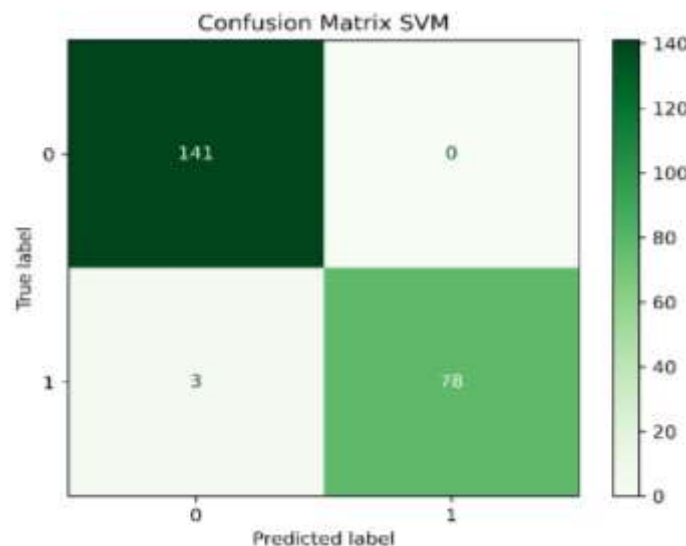


Figure 5. Confusion Matrix of the SVM Classifier for K = 2 FCM-Derived Consumer Categories

Overall, the classification findings suggest that the consumer groups obtained from the K=2 FCM are easily separable by supervised machine learning techniques. SVM is the most stable classifier with the highest accuracy, precision, recall, F1-Score, and AUC of 1.000 among the classifiers tested. The results show that the consumption patterns found by FCM clustering may be well described by two different household power consumption groups and then classified by supervised learning.

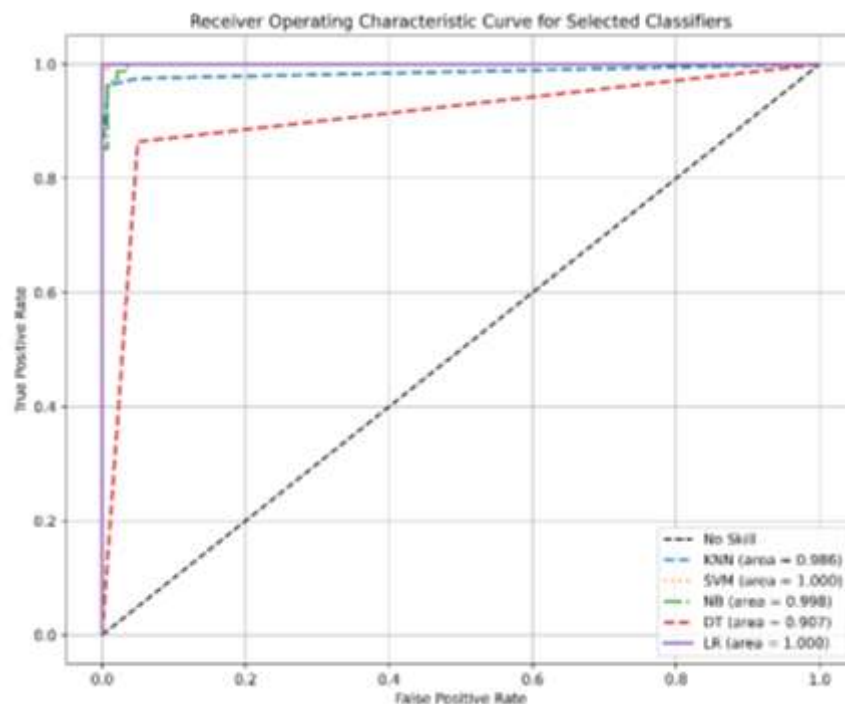


Figure 6. ROC Curves and AUC Values of the Machine-Learning Classifiers

6.CONCLUSION

In this study, a framework based on FCM was proposed for the categorisation of household power consumption utilising original monthly electricity consumption data of 1,110 domestic consumers over a period of 48 months. The FCM clustering was performed for cluster numbers $K = 2$ to $K = 10$ and the clusters obtained were evaluated using the Silhouette Coefficient (SCI) and Davies–Bouldin Index (DBI).

The clustering analysis suggested that $K=2$ was the most appropriate clustering option, with the highest SCI of 0.4239 and the lowest DBI of 0.9520. The customers that resulted from this were grouped into two groups, low consumption (LC) and high consumption (HC), with 705 and 405 consumers, respectively. The $K = 3$ approach further split customers into LC, MC and HC groups, offering a finer classification of consumption habits.

To further investigate the separability of the $K = 2$ categories obtained from the FCM, five machine-learning classifiers (KNN, SVM, NB, DT and LR) were tested with 10-fold stratified cross-validation and independent test-set evaluation. Among the assessed approaches, SVM displayed the most stable overall performance with 98.56% cross-validation accuracy and 98.65% independent test-set accuracy. The SVM classifier also had an AUC of 1.000 indicating excellent discrimination between the two consumption categories.

The results show that FCM can successfully categorise residential electricity consumers according to their consumption characteristics, and the consumer groups identified by FCM may be reliably classified using supervised machine-learning algorithms. The suggested framework provides a simple data-driven strategy that can assist residential consumer profiling, targeted energy-efficiency measures and demand-side energy management.

7. FUTURE WORK:

Future study will focus on extending the suggested framework with larger and more diversified residential power consumption information obtained from different geographic regions and consumer categories. The stability and transferability of the consumption clusters obtained from the FCM can be further explored utilising clustering and validation methods. Also, the classification framework can be enhanced by implementing advanced ensemble and deep learning models for better prediction of customer consumption categories. Furthermore, the consumption patterns obtained can be combined

with renewable energy generation, battery energy storage and demand-response strategies to provide personalised energy-efficiency and energy-management suggestions for residential consumers. The real-time smart-meter data can also be used to provide dynamic consumer profiling and adaptive energy management systems.

7.REFERENCE:

- [1]. M. Anvari, E. Proedrou, B. Schäfer, C. Beck, H. Kantz, M. Timme, Data-driven load profiles and the dynamics of residential electricity consumption, *Nature Communications* 13 (2022) 4593.
- [2]. E. Proedrou, A comprehensive review of residential electricity load profile models, *IEEE Access* 9 (2021) 12114–12133.
- [3]. Y. Wang, I.L. Bennani, X. Liu, M. Sun, Y. Zhou, Electricity consumer characteristics identification: A federated learning approach, *IEEE Transactions on Smart Grid* 12 (4) (2021) 3637–3647.
- [4]. Z. Guo, J.R. O'Hanley, S. Gibson, predicting residential electricity consumption patterns based on smart meter and household data: A case study from the Republic of Ireland, *Utilities Policy* 79 (2022) 101446.
- [5]. J. Zhou, Y. He, Recognition of household electricity consumption pattern in smart grid, *Journal of Building Engineering* 105 (2025) 112467.
- [6]. Y. Sun, M. Li, I. Valiyeva, Z. Xakimov, uncovering residential electricity consumption patterns using an integrated time-series clustering approach, *Energy Strategy Reviews* 65 (2026) 102214.
- [7]. V. Michalakopoulos, E. Sarvas, I. Papias, P. Skaloumpakas, V. Marinakis, H. Doukas, A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs, *Applied Energy* 361 (2024) 122943.
- [8]. H. Wen et al., A novel approach for identifying customer groups for personalized demand-side management services using household socio-demographic data, *Energy* 286 (2024) 129593.
- [9]. S. Kallel, M. Amayri, N. Bouguila, Clustering and interpretability of residential electricity demand profiles, *Sensors* 25 (7) (2025) 2026.
- [10]. W. Tang, H. Wang, X.-L. Lee, H.-T. Yang, Machine learning approach to uncovering residential energy consumption patterns based on socioeconomic and smart meter data, *Energy* 240 (2022) 122500.
- [11]. X. Zhang, J.L. Ramírez-Mendiola, M. Li, L. Guo, Electricity consumption pattern analysis beyond traditional clustering methods: A novel self-adapting semi-supervised clustering method and application case study, *Applied Energy* 308 (2022) 118335.
- [12]. K. Zhou, S. Yang, Z. Shao, Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study, *Journal of Cleaner Production* 141 (2017) 900–908.
- [13]. R. Kaur, D. Gabrijelčič, Behavior segmentation of electricity consumption patterns: A cluster analytical approach, *Knowledge-Based Systems* 251 (2022) 109236.
- [14]. V. Michalakopoulos, I. Papias, E. Sarantinopoulos, E. Sarvas, V. Marinakis, D. Askounis, A hyperparameter-space clustering methodology of residential electricity loads, *Applied Soft Computing* 181 (2025) 113497.
- [15]. Y. Li, X. Chu, D. Tian, J. Feng, W. Mu, Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm, *Applied Soft Computing* 113 (2021) 107924.
- [16]. O. Adeleke, O.J. Adebawale, I. Ayoade, I. Olawuyi, T.-C. Jen, Residential energy consumption dynamics: A SHAP-based interpretation, k-means clustering, and predictive modelling, *Green Energy and Resources* 4 (1) (2026) 100169.
- [17]. R. Delage, T. Nakata, Multivariate empirical mode decomposition and recurrence quantification for the multiscale, spatiotemporal analysis of electricity demand—A case study of Japan, *Energies* 15 (17) (2022) 6292.
- [18]. F.N. Melzi, A. Same, M.H. Zayani, L. Oukhellou, A dedicated mixture model for clustering smart meter data: Identification and analysis of electricity consumption behaviors, *Energies* 10 (10) (2017) 1446.
- [19]. S. Palaniappan, S. Karuppannan, D. Velusamy, Categorization of Indian residential consumers electrical energy consumption pattern using clustering and classification techniques, *Energy* 289 (2024) 129992.

- [20].Y. Wang, T. Cao, H. Wang, Q. Fu, J. Chen, Y. Lu, An interactive clustering method via metric learning for analyzing energy consumption patterns, *Journal of Building Engineering* 98 (2024) 111143.
- [21].E. Sarmas, A. Fragkiadaki, V. Marinakis, Explainable AI-Based Ensemble Clustering for Load Profiling and Demand Response, *Energies* 17 (22) (2024) 5559.
- [22].J. Wang, Z. Chen, W. Zhang, X. Du, Clustering-based identification of typical urban block morphologies and energy consumption archetypes: a case study of Beijing, China, *Energy and Buildings* 353 (2026) 116923.
- [23].U. Misra, G.R. Sahoo, S. Prakash, M. Srivastava, Unraveling Power Consumption Pattern Using Multidimensional Scaling and Clustering Analysis, *IEEE Access* 13 (2025) 13573–13583.
- [24].V.R. Chifu, T. Cioara, I. Danci, C.B. Pop, I. Anghel, Whale optimization-based clustering for buildings' energy consumption profiles, *Energy Reports* 14 (2025) 1207–1221.
- [25].E.T. Habtemariam, M. Martínez-Ballesteros, A. Troncoso, F. Martínez-Álvarez, A novel approach based on clustering and optimized ensemble deep learning for energy consumption forecasting in Ethiopia, *Neurocomputing* 637 (2025) 130027.
- [26].E.L. Ofetotse, E.A. Essah, R. Yao, Evaluating the determinants of household electricity consumption using cluster analysis, *Journal of Building Engineering* 43 (2021) 102487.
- [27].J. Hu, Y. Wang, X. Ai, Y. Qu, J. Wang, Y. Liu, Application of BIM-DesignBuilder and K-means clustering-random forest for prediction of building energy consumption, *Energy* 360 (2026) 141613.
- [28].A.B. Owolabi, A. Yahaya, M. Amir, A.O. Yakub, M. Kavagic, D. Suh, Predicting energy consumption of building clusters at the design stage using machine learning models, *Ain Shams Engineering Journal* 16 (8) (2025) 103481.
- [29].R. Estrada, P. Farinango, K. Santana, V. Asanza, Energy Consumption Prediction System based on clustering techniques, *Procedia Computer Science* 251 (2024) 170–177.
- [30].D.T. Kusuma, N. Ahmad, S.S. Syed Ahmad, I.B.M. Sangadji, Y. Arvio, An efficient clustering approach in electrical energy consumption patterns, *Bulletin of Electrical Engineering and Informatics* 14 (2) (2025) 1168–1177.
- [31].L. Zhu, J. Liu, C. Hu, Y. Zhi, Y. Liu, Analysis of Electricity Consumption Pattern Clustering and Electricity Consumption Behavior, *Energy Engineering: Journal of the Association of Energy Engineering* 121 (9) (2024) 2639–2653.