

# ELECTRICITY CONSUMPTION PATTERNS USING SOM-BASED TWO-LEVEL CLUSTERING OF RESIDENTIAL HOUSEHOLDS

Rinku Chavda

PhD Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India  
rinku.chavda.12@gmail.com

Dr. Sohil D. Pandya

Department of Computer Applications, CMPICA, CHARUSAT, Charotar University of Science and Technology (CHARUSAT), CHARUSAT campus, Changa 388421, India  
sohilpandya@gmail.com

Dr. Chetan D. Kotwal

Department of Electrical Engineering, Sardar Vallabhbhai Patel Institute of Technology, Vasad-388306, Gujarat, India  
chetan.kotwal@gmail.com

## Abstract

Electricity consumption in residential households' accounts for a huge share of energy use, which encourages researchers, policy makers towards energy conservation of residential household. The goal of Clustering using Unsupervised Machine Learning algorithms is to find likenesses in the data point of D-EC and cluster similar data points of D-EC together. This paper emphasis on application of machine learning techniques for pattern identification with the implementation of dimension reduction technique Self-Organizing Map algorithm trailed by unsupervised clustering algorithms such as K-means, Gaussian Mixture clustering, MiniBatchKMeans, Agglomerative Clustering, Spectral Clustering on D-EC data of electrical usage. Consumers classified into 4 clusters and labelled as per the usage pattern, also identified as a consistent and non-consistent consumer based on their electricity usage. A proposed methodology concluded using SOM followed by Kmeans algorithm which helpful for awareness and alerts to consumers based on consumption patterns participates in demand-supply decision making with timely recommendation.

**Keywords:** Electricity Consumption Patterns; Load Profiles; K-means Clustering; Self-Organized Map; Dimension Reduction.

## 1. Introduction

Electricity used by electric appliances is known as electricity usage of household. Electricity consumption is the genuine electricity demand made on current electricity supply. Energy consumption in residential households' accounts for a huge share of global energy use, which motivates many researchers and policy makers towards energy conservation of residential household and buildings. Residential electricity consumption is the total electricity used by households for different end-uses like cooking, refrigerating food, cooling rooms, lights, washing-machine, dish-washer and many other day-to-day activities. It is useful to understand how these end-uses of electricity contribute to the total REC. The electricity consumption patterns of different households demonstrate high discrepancy, due to the fact that their energy usage decision makings are usually affected by various factors such as day time, temperature, no of persons in family, no of rooms [Zhou and Yang, 2016]. A time-series data of electricity consumption of the end-use can help to understand the patterns of change in electricity consumption. This is useful to predict the future demand for electricity from households. A better understanding of the electricity consumption patterns can provide useful insights for designing policy and technological responses for managing and meeting the increasing demand. Electricity consumption patterns of a consumer may differ significantly within consumer clusters due to numerous influences like weather condition, temperature, lifestyle, etc. and difficult to discover the behavioral attributes of consumers without advanced data analysis and mining techniques and based on analysis of electricity consumption patterns, user electricity consumption profiles can be generated [Rinku and Sohil, 2018]. Exploring the changing aspects of residential electricity usage at finely-resolved timescales is gradually practical with the rising accessibility of high-resolution

data and diagnostic methods to characterize those [Satre-Meloy et al., 2020]. Machine learning and Data Mining technique includes a series of algorithms, in which a defined methodology lookup for similarities in data points, based on similarities generate clusters to make classification system, data analytics, and constructing expressive outcomes. This paper will emphasis on the application of machine learning techniques for pattern identification with the implementation of dimension reduction technique i.e., Self-Organizing Map [SOM] algorithm followed by various unsupervised clustering algorithms such as k-means clustering algorithm, Gaussian Mixture clustering, MiniBatchKMeans, Agglomerative Clustering, Spectral Clustering on a dataset of electrical energy usage. The scientific objective of this work is the detection of consumption pat-terns in time series data of electricity consumption to generate consumer profiles. The remainder of the paper is structured as follows. Section 2 presents previous research work proposed by researchers in area of identification of consumption profiles. Section 3 explains SOM based Two-level Clustering Methodology. Section 4 discusses implementation of methodology on D-EC dataset and outcomes. The final section provides concluding remarks.

## 1. Literature Review

Segmentation of consumers based on time series data i.e., Daily Electricity Consumption [D-EC] useful to generate consumption profile which shows the usage behaviors of consumers. While producing consumption profile, consumer electricity usage and pattern depends on different factors such as income, house structure, no. of appliances, no. of rooms, no. of persons in house, age group, day type, weather condition, outside and inside temperature, etc. play effective role while doing analysis. Electricity Consumption Profiles [ECPs] are used in several analytical applications such as consumption control and estimation, future demand of electricity, providing tariff plan, detecting unusual electricity consumption and developing power market strategies [Satre-Meloy et al., 2020].

The purpose of Clustering using Unsupervised Machine Learning algorithms is to treasure likenesses in the data point of D-EC and cluster similar data points of D-EC together. Grouping data points using clustering gives us understanding of underlying patterns of different clusters or group. Numerous Unsupervised Clustering algorithms developed to implement clustering concept in data points of D-EC. Several researchers have used various algorithms to generate ECPs based on electricity usage, no. of appliances, temporal data, etc. Choice making process with initiation of use of Artificial Intelligence accomplished via Scientific growths and new machine learning techniques [Undavia et al., 2016].

Using k-means clustering, [Amin et al., 2019] have designed an automated method for grouping the apartments with comparable usage profiles and result “peaks” and “lows” in daily data patterns to divide the data into a set of parts, which can be demonstrated with linear regression and using the strong connection of weather temperature and the energy usage, also designed an ensemble of linear regression models which are built for different seasons [month] and time of the day. [Perez-Chacon et al., 2016] worked on the Kmeans clustering algorithm in the Apache Spark framework and anticipated in order to discover patterns from a time series for two buildings. Partitioning-based clustering method K-Means and K-Medoids discussed in aspect for understanding electricity load profile on two different types of datasets also used decision tree to categorize the data into seven dissimilar branch-es which benefits for forecast or classify the data [Ali et al., 2016]. [Nepal et al., 2019] presented a scheme to rise the performance of the K-means clustering algorithm by choosing improved initial centroids percentile method based on experimental cumulative distribution and numeral of clusters. [Cerquitelli et al., 2018] proposed an advanced scalable procedure named as Consumption Duration Curve Time Series to determine electricity consumption shapes over time of consumers with a comparable behaviour articulated by the duration arcs of their hourly consumption using K-Means clustering algorithm. [Razaque et al., 2017] measured the similarity between groups and based on similarity consumers are categorized into different load-profiles i.e., clusters using k-means clustering algorithm. [Abeykoon et al., 2016] concluded that unsupervised K-Means clustering algorithm shows a higher accurate results over other classifiers i.e., supervised learning methods neural networks, support vector machines, unsupervised learning method Mean-Shift to detect electricity consumption patterns of domestic power usage and used Silhouette classifier to select a clustering algorithm with improved performance in real time to detect usage pattern of devices. Clustering method K-means algorithm proposed by [Al-Wakeel and Wu, 2016] reveals that the benefit to have small numbers of clusters when the high correlation between the clustered daily profiles and for enormously dynamic residential load profiles, it is chosen to have large numbers of clusters. [Wen et al., 2018] proposed an improved K-means algorithm analysis to improve the proficiency and efficiency of electricity consumption data in which principal component analysis used to condense the dimensions of smart meter time series data and the preliminary cluster centers optimized. [Li Kangping, et al., 2016] stated as a large-scale dataset including about 216 houses’ consumption records in American, shown that the improved K-means has a better performance than traditional K-means algorithm and be able to extract typical consumption patterns. [Choi et al., 2019] analysed the electricity usage of home through Davis Boulden Index and Silhouette Index

to find the detailed optimal number of clusters in the K-means algorithm. A substantial presence of autocorrelation coefficients extracted and united them into successive clustering using K-Means clustering algorithm of smart meter electricity data [Tureczek et al., 2018]. A hierarchical clustering algorithm through Ward linkage method was applied to cluster the normalized load profiles and to identify the optimal number of partitions between 2 and 10 clusters, Davies-Bouldin [DB] Index and the Silhouette Index were used as competence measures [Capozzoli et al., 2017]. [Viloria et al., 2020] applied the K-means clustering algorithm on the electricity consumption data of different customers and as a result of which allowed to identify up to 10 groups of customers with similar behaviour. For the identification of groups of households with different consumption patterns of energy carriers, k-means algorithm was applied by the [Piekut 2020] and to assess the association between the risk of energy deficiency and specific socio-economic and demographic features of the legislatures of these households. K-Means Clustering algorithm was used by [Amri et al., 2016] to apply segmentation on the dataset of electricity consumption from 370 consumers and authors obtained a stimulating pattern that an immense group of clients consume the lowest electricity in spring season where another group consumed lowest electricity consumption during winter season. A proposed study tested hierarchical clustering, k-means clustering, FCM clustering, follow-the-leader clustering, and fuzzy relation clustering algorithms that are being used in generating typical load profile [TLP] and the k-means clustering technique judged to be the most efficient option if minimizing the MAE value between TLP and RLP is the over-riding priority [Kim et al., 2011]. Investigation employed a data-driven approach to recognize the shape of seasonal residential demand profiles, ideal number of normalized representative residential electricity usage profiles within each season via a Kmeans clustering algorithm, and recognize associations to the different profiles based on collected dataset of 103 households in Austin, TX [Rhodes et al., 2014]. To excerpt feature parameters to characterize the electricity consumption level and patterns a clustering-based probability distribution methodology was anticipated using Kmeans clustering algorithm to recognize the differences of electricity consumption levels and patterns [Xu et al., 2021]. K-means clustering algorithm applied to diagnose the shape of the daily profiles, also applied clustering to daily profiles taking into account de-tailed features such as daytime, night time, morning, evening and mean of the standard deviation for each period [Yilmaz et al., 2019]. Using K-means clustering algorithm on normalized loads of different household shown notable variance in topmost load times, which is valuable for peak consumption reduction planning [Zhang and Zimet, 2018]. [Souza et al., 2009] projected a technique of clustering which composed of two stages by combining SOM and Ant k-means to reduce computational time by the hierarchical clustering method or partitive for the large and complex sets of data.

A clustering-based electricity power consumption level prediction technique based on a deep autoencoder and SOM presented for consumer profiling using smart sensor data to obtain clustered data to create different stages of the electricity consumption level of each consumer [Ullah et al., 2020]. A two-levels approach was implemented on electricity consumption data of 24 flats in a residential estate using a SOM with the k-means algorithm and results in five daily energy profiles, that can be ascribed to different types of family and/or habits [Ferrando et al., 2019]. [Toussaint and Moodley, 2020] illustrated a method for prompting and representing expert province knowledge and application requirements to formalize clustering purposes using variations of kmeans, self-organizing maps, combination of the Kmeans & SOM algorithms and guided the evaluation and selection of clustering algorithms using Evaluation Metrics for Clustering with Internal and External Metrics. Unsupervised machine learning techniques such as Hierarchical cluster analysis, Grade Data Analysis, Sequential Association Rules, C-Means Clustering, and Multidimensional Scaling have been presented and employed by [Gajowniczek and Ząbkowski, 2015] to exhibit specific consumption usage patterns observed at an individual household based on on-off states of appliances like washing machine, dish washer, tumble dryer, kettle and microwave oven.

[Satre-Meloy et al., 2020] presented a unified investigation to cluster a sample of 269 peak-period electricity consumption profiles from households to capture temporal differences in consumption shapes using K-means clustering algorithm and several variants of Hierarchical Agglomerative clustering. [Alonso et al., 2020] presented three diverse hierarchical-based clustering approaches based on dissimilarity measures figured based on quantile auto-covariances, and simple and partial autocorrelations on 30-minute time series smart-meter dataset and assessed the performance of each clustering model whose results show how the gained clusters recognize appropriate consumption behaviours of households. A two-stage methodology developed for the clustering and classification of electricity consumers to find load profile based on pattern identification procedures, such as Kmeans clustering algorithms, Kohonen adaptive vector quantization, Fuzzy K-means clustering, and Hierarchical clustering and performance of each approaches compared using different adequacy measures such as Mean square error or error function, Clustering Dispersion Indicator, Mean Index Adequacy, Similarity Matrix Score, Davies-Bouldin Score, Ratio of within cluster sum of squares to between cluster variation [Tsekouras et al., 2007].

A bottom-up prediction framework proposed by [Gao et al., 2018] to determine household consumption profiles using fuzzy c-means clustering, in which energy consumption behaviour characteristics of each household appliance was viewed as the main building block and in order to improve the performance of the foretelling process, a similar day extraction process was proposed to select the day most similar to the projected day from numerous past days. [Li et al., 2018] presented a clustering-based policy to identify Typical Daily Electricity Usage [TDEU] profiles of multiple buildings using an Agglomerative Hierarchical clustering and a Gaussian mixture model-based clustering used to generate profile of each individual building. [Albert and Rajagopal, 2013] have established a dynamic model based on Hidden Markov Model which builds dynamic consumption profiles based on temperature for individual and aggregated consumers and also addressed how this method may be used in exercise to understand consumption and its components mainly driven by response based on outside temperature. The different states of power consumption the Hidden Markov Model modelling method of pattern identification from single-user single-appliance, single-user multi-appliances, multi-users multi-appliances progressively in-depth analysis and the main algorithms of learning and decoding introduced, and the concept of association possibility matrix presented [Lin et al., 2018]. Typical Electricity Consumption Pattern of each consumer identified using Adaptive DBSCAN followed by K-means clustering algorithm used to group consumers showing similar Typical Electricity Consumption Pattern together [Wang et al., 2018]. [Zhou et al., 2017] explored the daily energy usage patterns of low-voltage residential consumers in China and proposed FCM based process model to obtain typical electricity consumption profiles when the number of clusters was set to three, eight, and ten and a fuzzy cluster validity index was used to determine an appropriate number of clusters. An improved fuzzy clustering model developed for the monthly electricity consumption pattern mining of daily electricity consumption data of 1200 households to sustainance the development of custom-made and targeted publicizing policies and the enhancement of energy competence [Zhou et al., 2017]. A data-driven segmentation approach was presented by [Afzalan and Jazizadeh, 2019] using real-world data for more than 300 households for demand response which has successfully distinguished users based on demand reduction potential with a cross-appliance comparison of flexibility assessment for different segments of users and findings have shown high variation of demand reduction potential based on consumption patterns. [Ardakanian et al., 2014] applied an auto-regression model i.e., Periodic Auto Regression with eXogenous variables on 1000 residential consumers to generate simple, easy to understand and effective profiling framework based on consumption at different day timings and external temperatures. [Liu et al., 2015] proposed Periodic Auto Regression [PAR] algorithm to excerpt consumption profiles of consumers and the time series similarity search algorithms to discover analogous consumers to form cluster. A flexible DR scheme in smart grid with K-means clustering algorithm of residential customers and broadly considering the factors such as Building type, House construction year, Total room number, Total square footage, Total appliance number were proposed by authors to extract new features from past data to describe customers' characteristics and clustering methods applied to explore their electricity consumption behaviours [Dai et al., 2019]. [Kleiminger et al., 2013] offered and assessed an approach that influences electricity smart meters as occupancy sensors using stateless and stateful classification algorithms where stateless classifiers were Support Vector Machines [SVM], K-Nearest Neighbour [KNN] and Thresholding [THR] and stateful classifier was Hidden Markov model [HMM] using a data set collected during an 8-month long experiment run in 5 households. A method proposed by [Liang et al., 2019] to detect power constantly consumed by some appliances that are never turned off [baseload] segments from daily consumption profiles using Sliding Window Linear Regression and shown the high accuracy performance of the proposed method and its adaptiveness to the heterogeneity in energy consumptions across customers also discussed how the projected method can be employed to classify customers with high baseload energy saving capacities and low effectual refrigerator-freezer. A combination of Symbolic aggregate approximation technique as a data size reduction process and hierarchical clustering algorithm employed to find daily load curve of customers and segregated into certain clusters based on similarity [Rajabi et al., 2019]. [Rinku and Sohil, 2018] have presented a literature survey for algorithms and technologies used in smart electricity meter data analytics which focuses on electricity consumption usage, pattern, profile and load forecasting. Authors have presented how R can be used to see daily, monthly and quarterly consumption pattern and profile based on consumers' usage. With the help of Apache Hadoop and R, analysis of time series smart meter data by applying ARIMA and ARMA models to forecast the demand for electricity have been introduced by [P. Mathiyalagan et al., 2017]. Energy consumption behavioural shapes of different households show high variance, due to the fact that their decision makings about electricity usage affected by various intra-personal, inter-personal and outside factors was presented by [Zhou and Yang, 2016] to enhance the understanding of social issues in energy consumption from the information science [or more specifically data science] perspective. [Fikru and Gautier, 2015] deliberated the influence of weather variation on energy consumption by using 5-min interval weather and energy data of two residential households and outcomes specified that the relationship between energy consumption and weather is sensitive to what time of the day/night it is as measured by the statistical implication of the during days in most of the regression outputs. A monthly electricity consumption data of 70 households assembled between 2012 and 2014 to appraise the load profile and based

on analysis it was concluded that electricity load profile differs suggestively with the home size, income and number of persons in home [Lhendup et al., 2015]. Energy consumption pattern of 24 sample households which were clustered into three categories based on the income level was compared for various functional utilities reinforced usage of appliances and showed that thermal comfort appliances subsidize more to the consumption for medium and low-income group compared to high-income group because of unfamiliarity regarding energy-effectual appliances [Sonawane and Gumaste, 2020].

## 2. Methodology

The literature review on Electricity Consumption Patterns and generating Consumer Profiles, several unsupervised clustering algorithms have been used by various authors such as Kmeans Clustering, Hierarchical Clustering, Fuzzy C-Means Clustering, Self-organizing map for reduction of data dimensions. The formation and justification of the SOM based Two-level Clustering Methodology consist of three phases: Interpretation of Data, Model Establishment and Model Evaluation.

- (1). Interpretation of D-EC data phase consists of clustering tendency and data pre-processing of D-EC data points.
- (2). Model establishment phase consists of two approaches. First approach applied clustering algorithms on D-EC data points without applying data reduction technique and Second approach applied data reduction technique to reduce dimensionality of D-EC data points. Both approaches have used Internal Validity Indices (IVI) to obtain optimum Number of Clusters using Distortion (sum of squared distances to centers) obtaining optimum number of clusters for D-EC data and applied five clustering algorithms i.e., Kmeans Clustering, Hierarchical Agglomerative Clustering, Gaussian Mixture Clustering Model, Spectral Clustering, MiniBatchKMeans with and without using data reduction algorithm Self-Organizing Maps as a first step of methodology.
- (3). Model Evaluation phase consists of identification of consumption range for each cluster and applied Internal Evaluation Measures on Unsupervised Clustering Algorithms to select algorithm whose performance is remarkable.

Phases defined to execute SOM based Two-level Clustering Methodology have followed steps described in below algorithm.

- step 1. Cluster Tendency (CT) of D-EC data points using Hopkins Statistic (H)
- step 2. Data Pre-processing – Data cleaning with preparing features ( $f_i$ ) to generate Electricity Consumption Patterns (ECPs).
- step 3. Execute Unsupervised clustering algorithms on D-EC to generate ECPs
  - (1). First Approach – Apply Internal Validity Indices (IVI) to find optimum Number of Clusters ( $k$ ) using Distortion ( $D$ ) followed by execution of Unsupervised Clustering algorithms
    - (i). Observation on consumption range formed by each clustering algorithms and outcomes identified.
  - (2). Second Approach – Self-Organizing Map based Two-level Clustering Methodology to generate ECPs
    - (i). Apply Self-Organizing Map on D-EC for dimension reduction
    - (ii). Internal Validity Indices (IVI) to find optimum Number of Clusters D-EC data with SOM-Index followed by Unsupervised Clustering Algorithms
    - (iii). Observe performance of clustering algorithms using score calculated using Evaluation Metrics i.e.,
      1. Davies-Bouldin Score (DB),
      2. Calinski-Harabasz Score (CH),
      3. Silhouette Coefficient (S)
    - (iv). Select Clustering algorithm whose performance remarkable
- step 4. Observations and Outcomes based on Consumption of D-EC data

This process is followed by a discussion of the methodology. The overall technical approach is illustrated in Fig. 1.

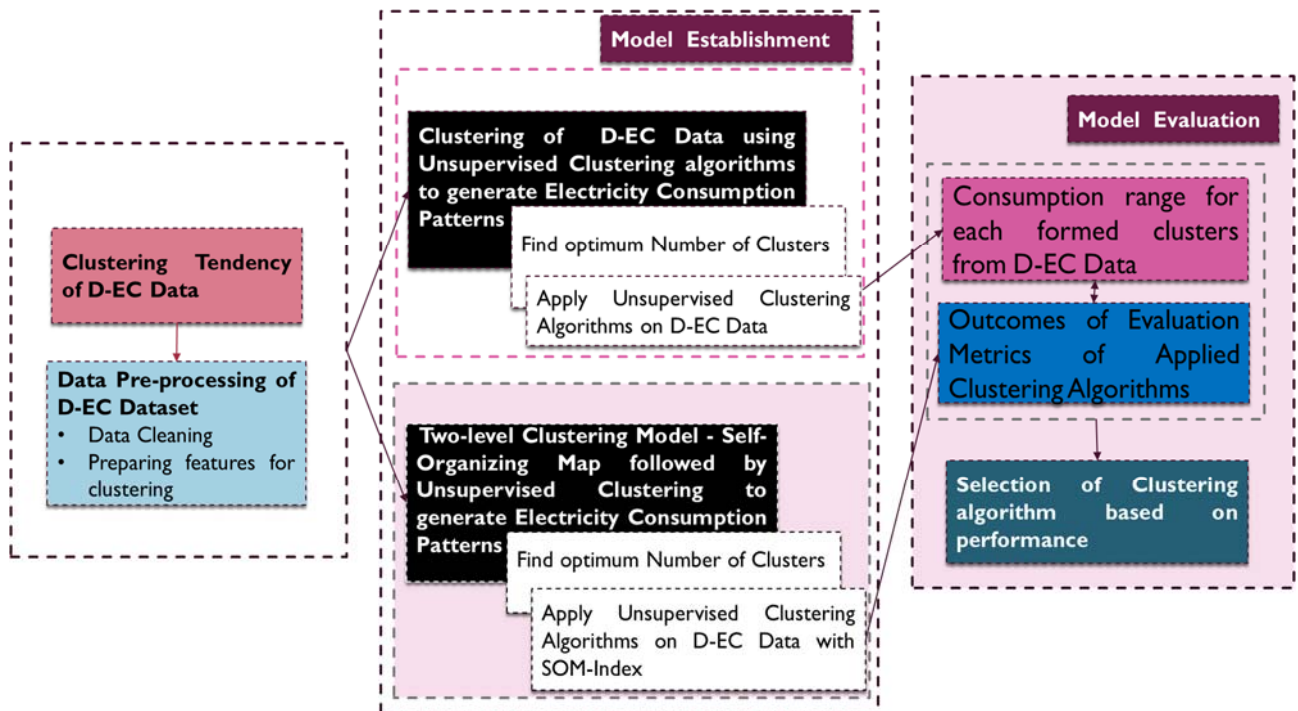


Fig. 1 Self-Organizing Map based Two-level Clustering Methodology to generate Electricity Consumption Patterns (ECPs)

### 3.1. Interpretation of D-EC Data

#### 2.1.1. Clustering tendency

A clustering approach, the analyst has to check whether the time series data comprises meaningful groups or not. If groups exist, then how many clusters are there? This method is defined as the measuring of clustering tendency or the possibility of the clustering analysis. A Hopkins statistical test which allows to guess if the data follow a uniform distribution.

#### 2.1.2. Data pre-processing

Data pre-processing aims to facilitate the training/testing process by appropriately transforming and scaling the entire dataset. Measurements of electric power consumption in 6445 household with a daily sampling rate over a period of one and half years with 536 measurements gathered between July 2009 and December 2010 (18 months) as Comma Separated Values.

#### 2.1.3. Data cleaning and preparing features of clustering

Data cleaning is the method of sensing and modifying (or removing) imprecise or missing records from dataset and refers to identifying imperfect, unfitting, or irrelevant parts of the data and then replacing, reshaping, modifying, or deleting unused data.

### 3.2. Model Establishment

The model establishment comprises of two approaches based on dimension reduction technique using Self-Organizing Maps followed by unsupervised clustering algorithms.

- (1). *First approach.* Internal Validity Indices using Distortion method followed by Unsupervised Clustering algorithms have been applied on D-EC data and compared results of algorithms based on the consumption range.
- (2). *Second approach.* Two-level Clustering Methodology used for dimension reduction using Self-Organizing Map applied on D-EC data followed by unsupervised clustering algorithms to discover clusters of D-EC data to generate consumption patterns and profiles based on similarities of data points.

### 3.3. Model Evaluation

Clustering algorithms evaluated based on some similarity or dissimilarity measures such as distance between cluster points, ratio of within-cluster distances to between-cluster distances. If the algorithm can unite similar data points and separate the dissimilar data points well, then it has performed well. For unsupervised clustering algorithms, the clustering result being evaluated based on the data points that was grouped by itself, is known as internal evaluation. Internal Evaluation Measures for Unsupervised Clustering Algorithms used to find performance of clustering algorithms. Internal evaluation metrics named used to check the performance of clustering algorithms. Based on the result obtained using internal evaluation metrics, clustering algorithm has been selected whose performance is noteworthy.

## 3. Implementation of Two-level Clustering Methodology on D-EC Dataset

### 3.1. Clustering Tendency and Data Pre-processing

#### 3.1.1. Clustering tendency

Before applying any clustering method on data points, it vital to assess whether the data points contain expressive clusters or not. If clusters are present in data, then how many clusters are there? This procedure is demarcated as the assessing of clustering tendency or the feasibility of the clustering analysis.

*Hopkins statistical test.* A Hopkins statistical test allows to speculate if the data points follow a uniform distribution or not. Based on the equation (1), if the Hopkins Statistics score inclines to 0, it means the data points are not consistently distributed. Hence bunching of similar data points can be beneficial to categorize the observations. The null hypothesis means no meaningful cluster happens when the Hopkins Statistics score is around 0.5.

$$H = \frac{\sum_{i=1}^m u_i^d}{\sum_{i=1}^m u_i^d + \sum_{i=1}^m w_i^d} \quad \text{Eq. (1)}$$

Hopkins Statistic test on D-EC data points applied and H score is 0.038 which is very near to 0 that means data point of D-EC is significantly clusterable dataset.

#### 3.1.2. Data pre-processing

##### (1) Dataset

The dataset used for implementing SOM based Two-level Clustering Methodology to generate Electricity Consumption Patterns (ECPs) consists of electricity readings of 6445 household with a daily sampling rate over a period of one and half years with 536 measurements collected between July 2009 and December 2010 as Comma Separated Values shown in Table 1.

Date	ID1000	ID1001	ID1002	ID1003	ID1004	...
14-07-2009	11.203	6.744	6.355	24.183	50.057	...
15-07-2009	8.403	6.949	8.972	26.659	48.813	...
16-07-2009	7.225	7.255	8.794	32.017	32.555	...
17-07-2009	11.338	7.19	8.306	33.032	46.727	...

Table 1. Daily Electricity Consumption (D-EC) of Households

##### (2) Data Cleaning

The D-EC data contains some lost values in the measurements nearly 5.3 % due to many reasons such as may be new consumers in the middle of observation period, some Meter IDs between 1000-7444 not observed or, some meter IDs that stop to be member in the middle of observation period. All the missing values are set with 0 value of the dataset.

##### (3) Preparing features of D-EC dataset for clustering.

Before applying clustering algorithms, transformation of data points needed to extract features from D-EC data points. Identified 538 variables to detect the consumption behaviour of every costumer which were listed in Table 2. From 538 feature variables, 536 features were Daily electricity usage in the observation period (in KW) i.e., 14th July 2009 to 31st December 2010. Remaining two features out of 538 features were meterID for each consumer and Total consumption in the observation period (total KW) for each consumer.

meterID	14-07-2009	15-07-2009	16-07-2009	...	31-12-2010	total_kw
ID1000	11.203	8.403	7.225	...	27.907	5515.675
ID1001	6.744	6.949	7.255	...	52.841	5090.375
ID1002	6.355	8.972	8.794	...	10.103	5352.83
ID1003	24.183	26.659	32.017	...	32.449	16305.58

Table 2. Extracted features from D-EC data for Clustering

### 3.2. Model Establishment

#### 3.2.1. First approach

Before applying clustering algorithms, Internal Validity Indices used to determine optimum Number of Clusters using Distortion method which calculates sum of squared distances from each point of D-EC to its assigned center. Applying the Elbow method on D-EC data points, sum of squared distances to centers were calculated known as Distortion score between the data points and their respective assigned clusters centroid using Kmeans clustering. By plotting, sum of squared distances to centers with its cluster number, identified that at cluster number 4 the graph flattens out and from cluster number 4 onwards the difference of SSE between the cluster numbers were very minimum shown in Fig. 2. Hence, 4 clusters considered as optimum number of clusters for D-EC data for next step of applying unsupervised clustering algorithms.

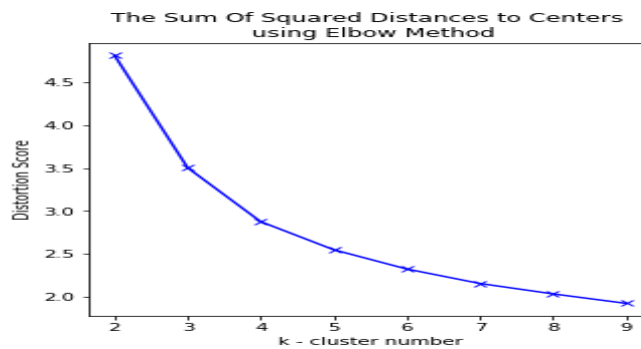


Fig. 2 The Sum of Squared Distances to Centers - Elbow Method

As mentioned in methodology section, five clustering algorithms i.e., Kmeans Clustering, Hierarchical Agglomerative Clustering, Gaussian Mixture Clustering Model, Spectral Clustering, MiniBatchKMeans applied on D-EC data points without applying dimension reduction technique i.e., Self-Organized Map and compared results based on the consumption range formed by each cluster shown in Fig.3. According to consumption range mentioned in Table 3, there was an issue of overlapping of D-EC data-points over the identified clusters means clear segmentation of data points was not identified.

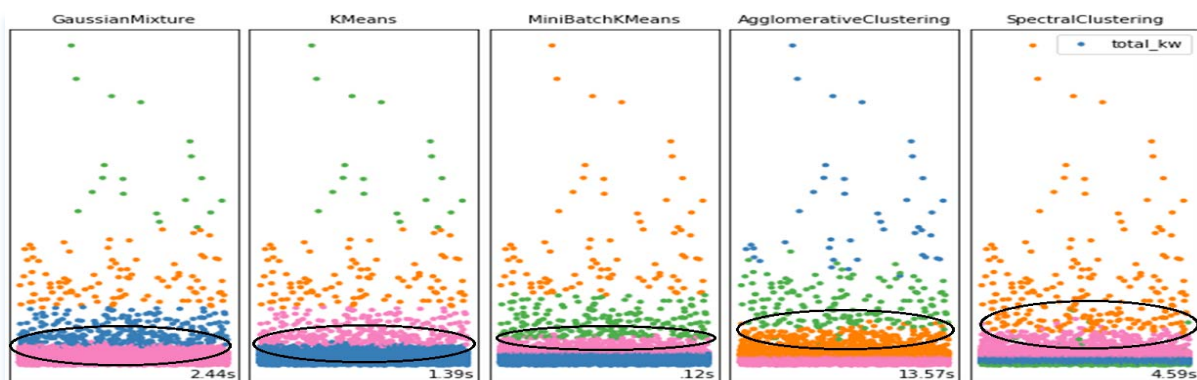


Fig. 3 Overlapping of D-EC data points in more than one cluster -Result of Unsupervised Clustering Algorithms



Gaussian Mixture Clustering				Kmeans Clustering			MiniBatchKMeans Clustering			Hierarchical Agglomerative Clustering			Spectral Clustering		
Cluster Number	Total no. of Consumers	min (kWh)	max (kWh)	Total no. of Consumers	min (kWh)	max (kWh)	Total no. of Consumers	min (kWh)	max (kWh)	Total no. of Consumers	min (kWh)	max (kWh)	Total no. of Consumers	min (kWh)	max (kWh)
0	6014	0	70.23	6014	0	70.23	5186	0	70.23	3787	0	39.12	288	0	128.9
2	315	54.43	173.4	315	54.43	173.4	1007	54.43	173.4	2451	17.26	130.5	3161	2.966	29.62
1	97	172.6	405.2	97	172.6	405.2	180	172.6	405.2	166	51.05	339.4	2778	15.95	163.5
3	19	412.09	953.76	19	412.09	953.76	72	412.09	953.76	41	267.1	953.8	218	97.31	953.8

Table 3. Consumption Range generated after applying Unsupervised Clustering Algorithms

### 3.2.2. Second approach (Two-level Clustering on D-EC data).

Before applying unsupervised clustering algorithms mentioned in methodology section, Two-level Clustering Methodology used for dimension reduction using SOM applied on D-EC data followed by unsupervised clustering algorithms to discover clusters of D-EC data points which is useful to generate consumption patterns and profiles constructed by means of similarities of data points. Self-Organizing Map used to reduce the dimension of D-EC data points to generate SOM-Index of D-EC data. The number of K clusters is an important parameter which influence the performance of the algorithm. Internal Validity Indices Sum of squared distance (Distortion), Davies-Bouldin Score and Silhouette Score with KMeans Clustering algorithm applied on SOM-index with average daily load to determine optimal value of k. As shown in Fig. 4, Experimental results of D-EC data with SOM-Index with average daily load, at cluster no. 4 the SSE starts flatten out as well as Silhouette Score and Davies-Bouldin Score were high and low respectively at cluster number 4. Unsupervised clustering algorithms Kmeans Clustering, Hierarchical Agglomerative Clustering, Gaussian Mixture Clustering Model, Spectral Clustering, MiniBatchKMeans applied on average daily consumption with SOM-Index to generate clusters based on similarities of electricity consumption usage.

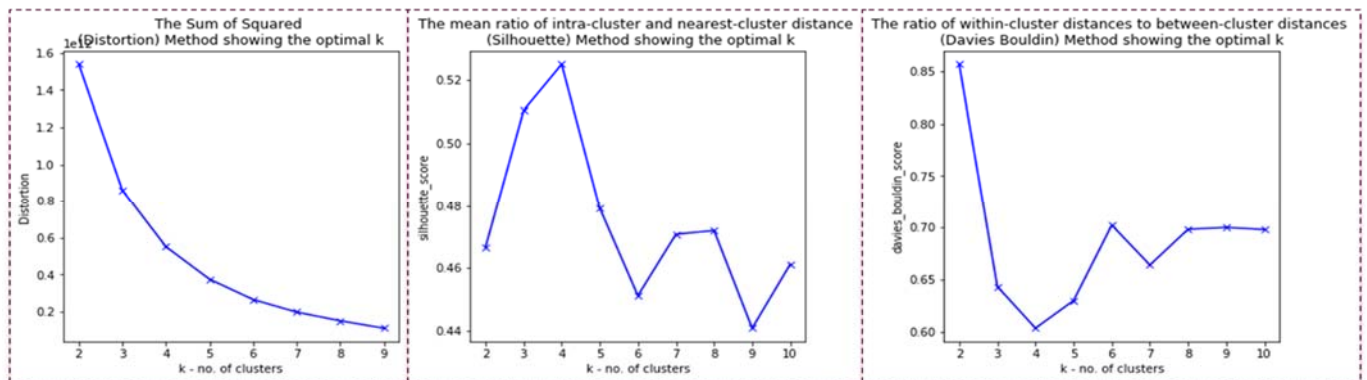


Fig. 4 Internal Validity Indices - Optimum Number of Clusters using SOM-Index

### 3.3. Model Evaluation using Internal Evaluation Measures

Clustering is evaluated based on similarity or dissimilarity measures such as distance between cluster points. If the algorithm can unite similar data points and separate the dissimilar data points well, then it has performed well. Internal Evaluation Measures for Unsupervised Clustering Algorithms used when a clustering outcome assessed based on the D-EC datapoints that clustered itself, it is termed as internal evaluation. Davies Bouldin Score, Silhouette Score, and Calinski Harabasz Score used to check the performance of clustering algorithms. Experimental results of described Evaluation metrics for all unsupervised clustering algorithms concludes that performance of K-Means clustering algorithm was superior as compare to other applied clustering algorithms. KMeans outperforms based on higher score of Silhouette and Calinski-Harabasz as well as lower score for Davies Bouldin shown in Fig. 5 and Table 4. As per Evaluation Metrics Outcomes of each algorithm, SOM based Two-level Clustering followed by Kmeans Clustering Algorithm is considered as highly significant clustering algorithm for D-EC data.

Evaluation Metrics	Gaussian Mixture Clustering	Kmeans Clustering	MiniBatchKMeans Clustering	Agglomerative Clustering	Spectral Clustering
Silhouette Score ( $\uparrow$ )	0.66	0.79 ( $\uparrow$ )	0.57	0.53	0.47
Calinski Harabasz Score ( $\uparrow$ )	9419.12	13953.19 ( $\uparrow$ )	9958.30	11455.01	2013.51
Davies-Bouldin Score ( $\downarrow$ )	0.65	0.49 ( $\downarrow$ )	0.62	0.64	0.75

Table 4. Internal Evaluation Results of Clustering algorithms

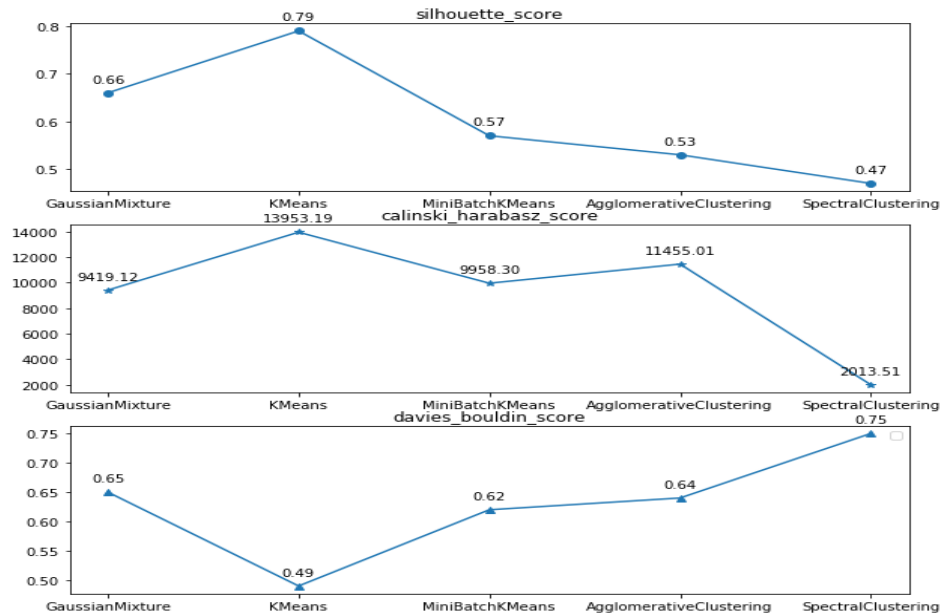


Fig. 5 Result of Internal Evaluation Metrics of Unsupervised Clustering Algorithms

### 3.4. Observations and Outcomes

Based on SOM-KMeans Cluster Index produced using SOM based Two-level Clustering followed by Kmeans Clustering Algorithm, consumers of D-EC data points divided into four groups, following Table 5 shows cluster wise total number of consumers, its daily and monthly average consumption boundaries with its usage label. As per the usage, 93.5% consumers are labelled as Low Usage Consumers who belongs to the cluster 0 and their daily consumption range is 0 – 62.5 kWh. Out of total consumers, 4.8% consumers are labelled as Moderate Usage Consumers who belongs to cluster 2 and their daily consumption usage range is 63 – 182 kWh. In cluster 1, 1.4% consumers identified as Elevated Usage Consumers where their daily consumption range is relatively high than consumption range of cluster 0 and 2 i.e., 183 – 428 kWh. 17 consumers of cluster 3 i.e., 0.3% of D-EC dataset have been labelled as Extravagant Usage Consumers based on their consumption i.e., 453 – 954 kWh of daily usage.

Sr. No.	Cluster No.	Total no. of Consumers	Percentage (%)	Averaged Consumption (in kWh) range formed by each cluster		Usage Label for Cluster
				Daily	Monthly	
1	0	6028	93.5	0 - 62.5	0 - 1862.1	Low Usage
2	2	307	4.8	62.9 – 182	1873.7 – 5419.8	Moderate Usage
3	1	93	1.4	182.3 – 428.1	5429.5 – 12747.1	Elevated Usage
4	3	17	0.3	452.8 – 953.8	13484.4 – 28401	Extravagant Usage

Table 5. Total Consumers of each cluster with its consumption range

#### 4.4.1 Sub-clustering of consumers based on average daily electricity (ADE) usage.

Sub-clustering applied on four clusters found using SOM-Kmeans Index with respect to average daily electricity usage. Based on ADE observations, many of the consumers of D-EC dataset have consumed low electricity or may be house-holds were not in use for more than almost 5 to 6 months for example MeterID 1335 of cluster 0 operated as irregular consumption of electricity who have consumed between 0-10 kWh for many days or months

during the year shown in Fig. 6 same as MeterID 3800 and 6368 of cluster 1 have not consumed electricity for almost 6 to 7 months.

For the identification of such non-consistent group of consumers, Kmeans Clustering algorithm applied on the outcomes instigated from SOM based Two-level Clustering followed by Kmeans Clustering Algorithm together with average daily load to identify the consistent and non-consistent consumers in each cluster and results were shown in Fig.7. Table 6 represented cluster-wise sub-clustering results concluding as 92.75%, 91.21, 94.62 and 100% consumers of cluster 0, 2 and 1 respectively called as Consistent Users and remaining i.e., 7.25%, 8.79 and 5.38 consumers of cluster 0, 2 and 1 respectively called as Non-Consistent Users based on their ADE. Based on ADE, months were classified into three groups.

1. In the month of January, March, May, Jul and August, 2358 consumers of cluster 0, 45 consumers of cluster 1, 127 consumers of cluster 2 used more electricity than ADE for approx. 14 days of each month whereas 3670, 50, 183 consumers of cluster 0, 1 and 2 respectively consumes less than ADE for ap-prox. 17 days of the month January, March, May, Jul and August. Hence for January, March, May, Jul and August months, the electricity consumption was more than other months i.e., February, April, June, September, October, November, December.
2. In the month of October and December, 4160, 63, 225 consumers of cluster 0, 1, 2 respectively consumed electricity less than ADE for approx. 18 days of both the months and for 12 days consumers have used electricity more than ADE.
3. In the month of February, April, June, Sep and November, consumers of each clusters i.e., 0,1, and 2 have used electricity approximately ADE for all the days of the months.
4. 8 consumers out of 17 consumers of cluster 3, in the month of January, March, May, Jul, October, November and December, have used electricity less than ADE approx. 14 days of each month and greater than ADE for approx. 17 days. Henceforth these months would be called as above ADE group and remaining months i.e., Feb, April, June, August, September are called as below ADE months in which 9 consumers have used electricity less than ADE for approx. 17 days and for remaining days they have used electricity greater than ADE for approx. 14 days.
5. As per the usage of consumers and ADE, three group were identified i.e., Above ADE, below ADE and equal to ADE group. January, March, May, July and August belong to above ADE group, October and December belongs to be-low ADE group and February, April, June, Sep and November belong to the ADE group.

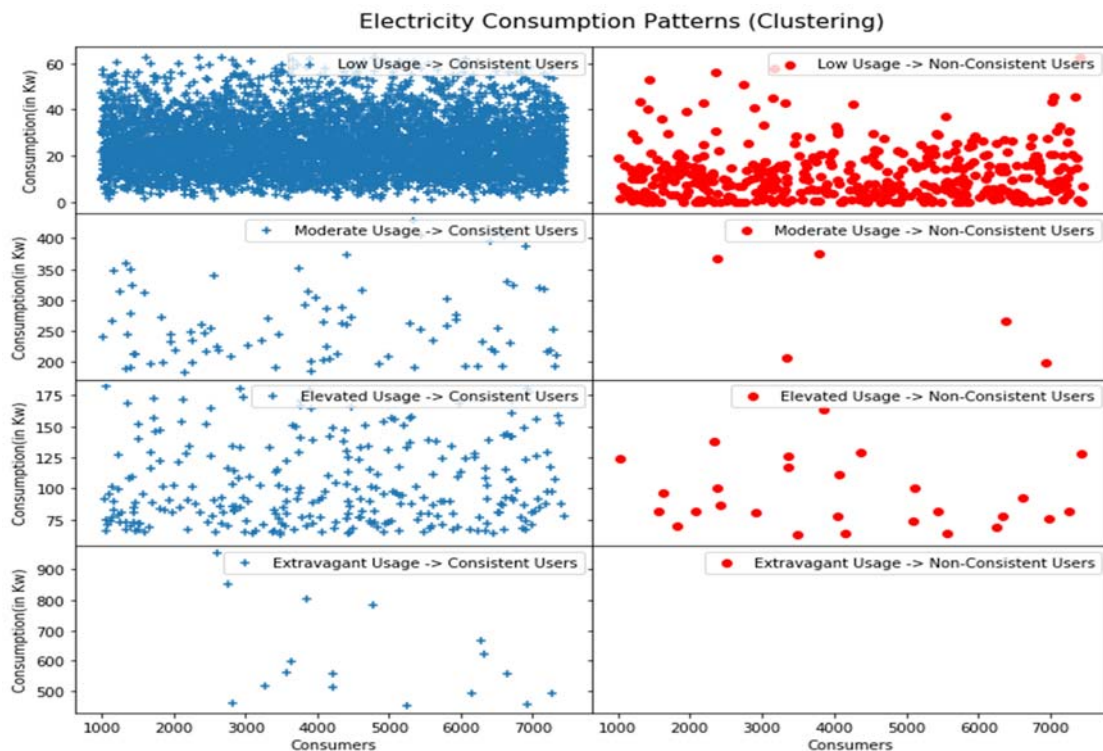


Fig. 6 Sub-clustering-based Consistent and Non-Consistent Consumers of each cluster

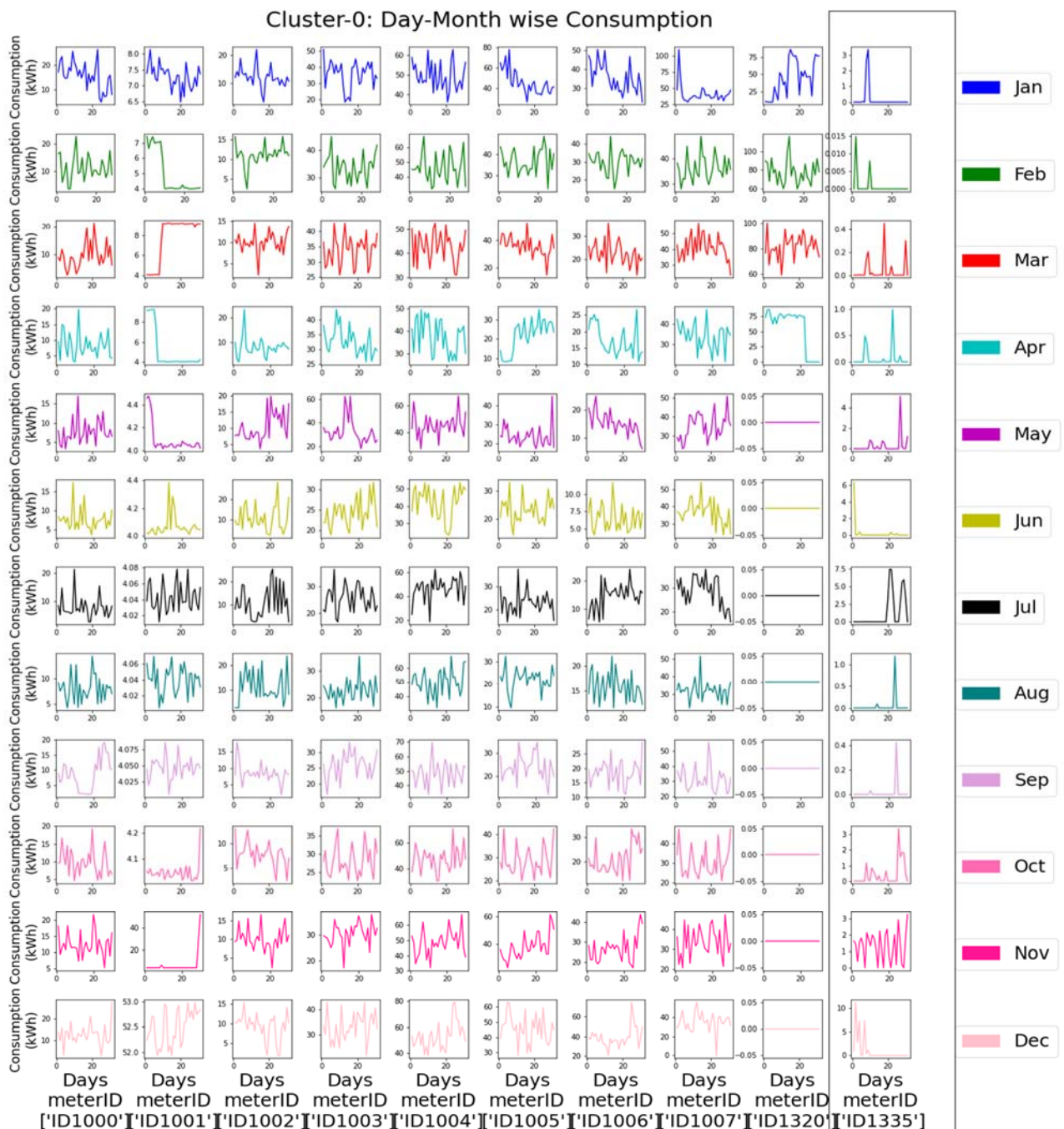


Fig. 6 Day-Month wise consumption of 10 Households from Cluster-0

Cluster No.	Sub-Cluster No.	Total no. of Consumers	Percentage (%)	Daily Averaged Consumption (in kWh) range formed by each cluster	1st Level Label for Consumers	2nd Level Label for Consumers
0	0	5591	92.75	0.89 - 62.53	Low Usage	Consistent User
	1	437	7.25	0 - 62.38		Non-Consistent User
2	0	280	91.21	62.93 - 182.01	Moderate Usage	Consistent User
	1	27	8.79	62.92 - 163.47		Non-Consistent User
1	0	88	94.62	182.33 - 428.07	Elevated Usage	Consistent User
	1	5	5.38	198.24 - 375.4		Non-Consistent User
3	0	17	100	452.84 - 953.76	Extravagant Usage	Consistent User

Table 6. Sub-clustering-based consumer labels of each cluster originated based on ADE



## 5. Conclusion

Self-Organizing Map based Two-level Clustering Methodology to generate Electricity Consumption Patterns (ECPs) was proposed and implemented to identify electricity profile of residential consumers based on time series data of Daily Electricity Consumption. The analysis results in four daily electricity profiles as based on their daily usage habits are labelled as Low Usage Consumers, Moderate Usage Consumers, Elevated Usage Consumers and Extravagant Usage Consumers. In each generated profile, consumers were identified as consistent and non-consistent consumers based on their average daily electricity consumption usage. Based on ADE observations, many of the consumers of D-EC dataset have consumed low electricity or may be households were not in use for more than almost 5 to 6 months. The proposed methodology opens an opportunity to look upon the unexplored domain of electricity consumption patterns and profile through data analytics and data mining techniques which will help to develop socio-economic applications to help society for awareness for managing their usage of energy, usage of their household in such way so that it affects their budget to decrease the cost, and from the provider's point of view, based on the selection of area or city, provider gets insight of total consumption of energy for particular day, time duration. Also, useful to have a model for implementation of dynamic tariff structure by visualization of unified consumption patterns, profiling and future need of electricity.

## References

- [1] Abeykoon, V., Nishadi, K., Senevirathna, R., Ranaweera, P., and Udawapola, R. (2016) 'Electricity Consumption Pattern Detection', Annual Research Symposium (ARS)
- [2] Afzalan, M., and Jazizadeh, F. (2019) 'Residential loads flexibility potential for de-mand response using energy consumption patterns and user segments', *Applied Energy*, Elsevier, vol. 254(C), doi:10.1016/j.apenergy.2019.113693
- [3] Albert, A., and Rajagopal, R. (2013) 'Building dynamic thermal profiles of energy consumption for individuals and neighborhoods', IEEE International Conference on Big Data, doi:10.1109/bigdata.2013.6691644
- [4] Ali, U., Buccella, C., and Cecati, C. (2016) 'Households electricity consumption analysis with data mining techniques', IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society, pp. 3966-3971, doi:10.1109/iecon.2016.7793118
- [5] Alonso, A. M., Nogales, F. J., and Ruiz, C. (2020) 'Hierarchical Clustering for Smart Meter Electricity Loads Based on Quantile Autocovariances', *IEEE Transactions on Smart Grid*, Vol. 11 No. 5, pp. 4522-4530, doi:10.1109/tsg.2020.2991316
- [6] Al-Wakeel, A., and Wu, J. (2016) 'K-means Based Cluster Analysis of Residential Smart Meter Measurements', *Energy Procedia*, Vol. 88, pp. 754-760, doi:10.1016/j.egypro.2016.06.066
- [7] Amin, P., Cherkasova, L., Aitken, R., and Kache, V. (2019) 'Analysis and Demand Forecasting of Residential Energy Consumption at Multiple Time Scales', IFIP/IEEE Symposium on Integrated Network and Service Management (IM), pp. 494-499.
- [8] Amri, Y., Fadhilah, A. L., Fatmawati, Setiani, N., and Rani, S. (2016) 'Analysis Clustering of Electricity Usage Profile Using K-Means Algorithm', *IOP Conference Series: Materials Science and Engineering*, Vol. 105, pp. 012-020. doi:10.1088/1757-899x/105/1/012020
- [9] Ardakanian, O., Koochakzadeh, N., Singh R.P., Golab, L. and Keshav, S. (2014) 'Computing electricity consumption profiles from household smart meter data', *CEUR Workshop Proceedings*, Vol. 1133, pp. 140-147
- [10] Bellala, G., Marwah, M., Arlitt, M., Lyon, G., Bash, C., and Shah, A. (2017) 'Data Analytics for Managing Power in Commercial Buildings', *ACM Transactions on Cyber-Physical Systems*, Vol. 1 No. 4, pp. 1-25. doi:10.1145/3110219
- [11] Capozzoli, A., Piscitelli, M. S., and Brandi, S. (2017) 'Mining typical load profiles in buildings to support energy management in the smart city context', *Energy Procedia*, Vol. 134, pp. 865-874. doi:10.1016/j.egypro.2017.09.545
- [12] Cerquitelli, T. et al, (2018) 'Discovering electricity consumption over time for residential consumers through cluster analysis', 2018 International Conference on Development and Application Systems (DAS), pp. 164-169 doi:10.1109/daas.2018.8396090
- [13] Chae, Y., Horesh, R., Hwang, Y., and Lee, Y. (2016) 'Artificial Neural Network Model For Forecasting Sub-hourly Electricity Usage In Commercial Buildings' *Energy And Buildings*, Vol. 111, pp. 184 - 194.
- [14] Choi, H. W., Qureshi, N. M. F., and Shin, D. R. (2019) 'Analysis of Electricity Consumption at Home Using K-means Clustering Algorithm', 2019 21st International Conference on Advanced Communication Technology (ICACT), pp. 639-643, doi:10.23919/icact.2019.8701981
- [15] D. Alberg, and Last, M. (2018) 'Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms', *Vietnam Journal of Computer Science*, pp. 241-249, doi:10.1007/s40595-018-0119-7
- [16] Dai, B., Wang, R., Zhu, K., Hao, J., and Wang, P. (2019) A Demand Response Scheme in Smart Grid with Clustering of Residential Customers. 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) doi:10.1109/smartgridcomm.2019.8909776
- [17] Deb, C., Eang, L. S., Yang, J., and Santamouris, M. (2016) 'Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks', *Energy and Buildings*, Vol. 121, pp. 284-297, doi:10.1016/j.enbuild.2015.12.050
- [18] Ferrando, M., Marchenko, A., Erba, S., Causone, F., and Carlucci, S. (2019) 'Pattern Recognition and Classification For Electrical Energy Use In Residential Buildings', *Proceedings of the 16th IBPSA Conference Rome, Italy*, pp. 2246-2253, 10.26868/25222708.2019.210750.
- [19] Fikru, M. G., and Gautier, L. (2015) 'The impact of weather variation on energy consumption in residential houses', *Applied Energy*, Vol. 144, pp. 19-30. doi:10.1016/j.apenergy.2015.01.040
- [20] Gajowniczek, K., and Zabkowski, T. (2015) 'Data Mining Techniques for Detecting Household Characteristics Based on Smart Meter Data', *Energies*, Vol. 8 No. 7, pp. 7407-7427. doi:10.3390/en8077407
- [21] Gao, B., Liu, X., and Zhu, Z. (2018) 'A Bottom-Up Model for Household Load Profile Based on the Consumption Behavior of Residents', *Energies*, Vol. 11 No. 8, 2112. doi:10.3390/en11082112
- [22] Ghofrani, M., Hassanzadeh, M., Etezadi-Amoli, M., and Fadali, M. S. (2011) 'Smart meter based short-term load forecasting for residential customers', 2011 North American Power Symposium, doi:10.1109/naps.2011.6025124
- [23] Jetcheva, J. G., Majidpour, M., and Chen, W.-P. (2014) 'Neural network model ensembles for building-level electricity load forecasts', *Energy and Buildings*, Vol. 84, pp. 214-223, doi:10.1016/j.enbuild.2014.08.004
- [24] Kim, Y.-I., Ko, J.-M., and Choi, S.-H. (2011) 'Methods for generating TLPs (typical load profiles) for smart grid-based energy programs', 2011 IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG), doi:10.1109/ciasg.2011.5953331

- [25] Kleiminger, W., Beckel, C., Staake, T., and Santini, S. (2013) 'Occupancy Detection from Electricity Consumption Data', Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys'13, doi:10.1145/2528282.2528295
- [26] Gajowniczek, K., and Zabkowski, T. (2014) 'Short Term Electricity Forecasting Using Individual Smart Meter Data', Procedia Computer Science, Vol. 35, pp. 589–597, doi:10.1016/j.procs.2014.08.140
- [27] Lhendup, T., Drukpa, T., Chodon, P., and Subba, A. (2015) 'Residential Electricity Use Load Profile Using Monthly Electricity Consumption'
- [28] Li Kangping, et al. (2016) 'Analysis on residential electricity consumption behavior using improved k-means based on simulated annealing algorithm', 2016 IEEE Power and Energy Conference at Illinois (PECI), doi:10.1109/peci.2016.7459209
- [29] Liang, H., Ma, J., Sun, R., and Du, Y. (2019) 'A Data-driven Approach for Targeting Residential Customers for Energy Efficiency Programs', IEEE Transactions on Smart Grid, Vol. 11 No. 2, doi:10.1109/tsg.2019.2933704.
- [30] Li, K., Ma, Z., Robinson, D., and Ma, J. (2018) 'Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering', Applied Energy, Vol. 231, pp. 331–342. doi:10.1016/j.apenergy.2018.09.050
- [31] Lin, G., Pan, F., Yang, Y., Yang, L., He, G., and Fan, S. (2018) 'The Pattern Recognition of Residential Power Consumption Based on HMM', 2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), pp. 413–418, doi:10.1109/isgt-asia.2018.8467905
- [32] Ling, J., Doris, L., Alex, S., Sam, B., Kesheng, W., C, Anna, S., Annika Todd. (2017) 'Comparison of Clustering Techniques for Residential Energy Behavior using Smart Meter Data', AAAI Workshops - Artificial Intelligence for Smart Grids and Buildings
- [33] Liu, X., Golab, L., Golab, W., and Ilyas, I. F. (2015) 'Benchmarking Smart Meter Data Analytics', In Proceedings of the 18th International Conference on Extending Database Technology (EDBT), pp. 385–396, <https://doi.org/10.5441/002/edbt.2015.34>.
- [34] Lu, N., Du, P., Guo, X., and Greitzer, F. L. (2012) 'Smart meter data analysis', PES T&D 2012, doi:10.1109/tde.2012.6281612.
- [35] Lusi, P., Khalilpour, K. R., Andrew, L., and Liebman, A. (2017) 'Short-term residential load forecasting: Impact of calendar effects and forecast granularity', Applied Energy, Vol. 205, pp. 654–669, doi:10.1016/j.apenergy.2017.07.114
- [36] Nepal, Yamaha, Sahashi, and Yokoe. (2019) 'Analysis of Building Electricity Use Pattern Using K-Means Clustering Algorithm by Determination of Better Initial Centroids and Number of Clusters', Energies, Vol. 12 No. 12, 2451. doi:10.3390/en12122451
- [37] Perez-Chacon, R., Talavera-Llames, R. L., Martinez-Alvarez, F., and Troncoso, A. (2016) 'Finding Electric Energy Consumption Patterns in Big Time Series Data, Advances in Intelligent Systems and Computing, pp. 231–238, doi:10.1007/978-3-319-40162-1\_25
- [38] Piekut, Marlena. (2020) 'Patterns of Energy Consumption in Polish One-Person Households', Energies, Vol. 13 No. 21, 5699, <https://doi.org/10.3390/en13215699>
- [39] P. Mathiyalagan, Shanmugapriya, A., and Geethu, A. V. (2017) 'Smart meter data analytics using R and Hadoop', 2017 IEEE International Conference on Electro In-formation Technology (EIT), doi:10.1109/eit.2017.8053438
- [40] Rajabi, A., Eskandari, M., Jabbari, M., Ghavidel, S., Li, L., Zhang, J., and Siano, P. (2019) 'A pattern recognition methodology for analyzing residential customers load data and targeting demand response applications', Energy and Buildings, doi:10.1016/j.enbuild.2019.109455.
- [41] Razaque, F., Soomro, N., Samo, J., Dharejo, H., and Shaikh, S. (2017) 'Analysis of Home Energy Consumption by K-Mean', Annals of Emerging Technologies in Computing (AETiC), Vol. 1 No. 1, pp. 1–6, doi:10.33166/AETiC.2017.01.001
- [42] Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., and Webber, M. E. (2014) 'Clustering analysis of residential electricity demand profiles', Applied Energy, Vol. 135, pp. 461–471, doi:10.1016/j.apenergy.2014.08.111.
- [43] Rinku, C., Sohil, P. (2018) 'Analyzing Electricity Consumptions Pat-tern for Profiling and Forecasting: A Review', International Journal of Innovative Research & Studies, pp. 236–242
- [44] Satre-Meloy, A., Diakonova, M., and Grünewald, P. (2020) 'Cluster analysis and prediction of residential peak demand profiles using occupant activity data', Applied Energy, Vol. 260, doi:10.1016/j.apenergy.2019.114246
- [45] Sonawane, GA., and Gumaste, KS. (2020) 'Global Challenges in Energy and Environment', Lecture Notes on Multidisciplinary Industrial Engineering, pp. 177–200, doi:10.1007/978-981-13-9213-9\_18
- [46] Souza, J. R., Ludermit, T. B., and Almeida, L. M. (2009) 'A Two Stage Clustering Method Combining Self-Organizing Maps and Ant K-Means', Lecture Notes in Computer Science, pp. 485–494. doi:10.1007/978-3-642-04274-4\_51
- [47] Toussaint, W. and Moodley, D. (2020) 'Clustering Residential Electricity Consumption Data to Create Archetype that Capture Household Behaviour in South Africa', South African Computer Journal, Vol. 32 No. 2, pp. 1–34, <https://doi.org/10.18489/sacj.v32i2.845>
- [48] Tsekouras, G. J., Hatzigargyriou, N. D., and Dialynas, E. N. (2007) 'Two-Stage Pat-tern Recognition of Load Curves for Classification of Electricity Customers', IEEE Transactions on Power Systems, Vol. 22 No. 3, pp. 1120–1128, doi:10.1109/tpwrs.2007.901287
- [49] Turczek, A., Nielsen, P., and Madsen, H. (2018) 'Electricity Consumption Clustering Using Smart Meter Data', Energies, Vol. 11 No. 4, doi:10.3390/en11040859.
- [50] Ullah, H., Ul H., Muhammad, R., Lee, and Baik. (2020) 'Deep Learning Assisted Buildings Energy Consumption Profiling Using Smart Meter Data', Sensors, Vol. 20 No. 3, doi:10.3390/s20030873.
- [51] Undavia, J. N., Dolia, P., and Patel, A. (2016) 'Customized Prediction Model to Predict Post-Graduation Course for Graduating Students Using Decision Tree Classifier', Indian Journal of Science and Technology, Vol. 9 No. 12, doi:10.17485/ijst/2016/v9i12/83335
- [52] Vaidya N.M. and Patel K.K. (2020) 'Learner Performance and Preference Meter for Better Career Guidance and Holistic Growth', In: Fong S., Dey N., Joshi A. (eds) ICT Analysis and Applications, Lecture Notes in Networks and Systems, Springer, Singapore, Vol. 93, [https://doi.org/10.1007/978-981-15-0630-7\\_5](https://doi.org/10.1007/978-981-15-0630-7_5)
- [53] Viloria, A., Naveda, A., Palma, H., Nunez, W. and Nunez, L. (2020) 'Electrical Consumption Patterns through Machine Learning', Journal of Physics: Conference Series, doi:10.1088/1742-6596/1432/1/012093.
- [54] Wang, F., Li, K., Duic, N., Mi, Z., Hodge, B.-M., Shafiekhah, M., and Catalao, J. P. S. (2018) 'Association rule mining based quantitative analysis approach of house-hold characteristics impacts on residential electricity consumption patterns', Energy Conversion and Management, Vol. 171, pp. 839–854. doi:10.1016/j.enconman.2018.06.017
- [55] Weicong, K., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., and Zhang, Y. (2017) 'Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network', IEEE Transactions on Smart Grid, Vol. 10 No. 1. doi:10.1109/tsg.2017.2753802
- [56] Wen, L., Zhou, K., and Yang, S. (2018) 'A shape-based clustering method for pat-tern recognition of residential electricity consumption', Journal of Cleaner Production, Vol. 212, pp. 475–488, doi:10.1016/j.jclepro.2018.12.067.
- [57] Xu, J., Kang, X., Chen, Z. et al. (2021) 'Clustering-based probability distribution model for monthly residential building electricity consumption analysis', Building Simulation, Vol. 14, pp. 149–164. <https://doi.org/10.1007/s12273-020-0710-6>
- [58] Yilmaz, S., Chambers, J., and Patel, M. K. (2019) 'Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management', Energy, Vol. 180, pp. 665–677, doi:10.1016/j.energy.2019.05.124.
- [59] Zhang, Z., and Zimet, T. (2018) 'K-means Based Clustering Analysis of Household Energy Consumption'
- [60] Zhou, K., Yang, C., and Shen, J. (2017) 'Discovering residential electricity consumption patterns through smart-meter data mining: A case study from China', Utilities Policy, Vol. 44, pp. 73–84, doi:10.1016/j.jup.2017.01.004

- [61] Zhou, K., and Yang, S. (2016) 'Understanding household energy consumption behavior: The contribution of energy big data analytics' *Renewable and Sustainable Energy Reviews*, Vol. 56, pp. 810–819, doi:10.1016/j.rser.2015.12.001
- [62] Zhou, K., Yang, S., and Shao, Z. (2017) 'Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study', *Journal of Cleaner Production*, Vol. 141, pp. 900–908, doi:10.1016/j.jclepro.2016.09.165

### Authors Profile



**Rinku Chavda**, Rinku Chavda is working as an Asst. Professor at Dr Subhash Technical Campus, Junagadh, Gujarat, India. She has completed her M.C.A. from SVIT, Vasad under Gujarat University, Ahmedabad, India. She has 10+ years' experience of teaching in MCA. She has expertise in Data Mining, Data Analytics, Android App development, Internet of Things.



**Dr. Sohil Pandya**, He is working as an Assistant Professor, Department of Computer Applications, CMPICA, CHARUSAT, Charotar University of Science and Technology (CHARUSAT), CHARUSAT campus, India. He is having 16+ years of experience in teaching and research. He has completed his Ph. D. from Kadi Sarva Vishwavidyalaya, Gujarat, India. His area of expertise is Data Mining, Information Retrieval, Internet of Things.



**Dr. Chetan Kotwal**, He received Bachelor in Electrical Engineering degree and Masters in Electrical Engineering from M.S. University of Baroda, India and PhD from Indian Institute of Technology, Roorkee. He is working as a Professor and Head of Electrical Engineering Department, SVIT, VASAD, Gujarat, India. He is having more than 23+ years of experience in teaching and research. His area of expertise is SMART GRID, Cooperative Energy Utilization, CYBER PHYSICAL SECURITY, Optimum sharing of Energy AI and POWER GRID.