

# Domestic electricity consumption analysis using data mining techniques

Prof.S.S.Darbastwar

*Assistant professor, Department of computer science and engineering, Dkte society's textile and engineering institute, Ichalkaranji, Maharashtra, India*

## ABSTRACT

*Smart Grid improves the electricity grid infrastructure by introducing new powerful communication system between consumer and supplier. Implementation of smart meter increases the availability of detail level of consumer electricity load profile data. To improve an efficient planning and development of these new power system, a primary challenge is to analyze the electricity consumption data. To analyze the energy consumption or achieve our objective it choose the best analytic process is data mining techniques including exploratory data analysis and pre-processing, frequent patterns mining and association, classification/characterization, clustering . In this paper, it use these techniques and apply on dataset. Explain and evaluate which techniques are useful for the better understanding of electricity load profile consumption data*

**Keyword :** - Data Mining ,Classification, Clustering, Electricity consumption analysis.

## 1. Introduction

In past few years, energy analytic is emerging and lots of researches on electricity consumption analysis like consumer segmentation, characterization, prediction and knowledge extraction from smart meter had been done. The data mining technique mostly used classification, clustering of electricity demand patterns and cluster analysis of smart meter in data. However, mostly Research on the smart meter analysis but in this paper it target the consumer (domestic) and utilities (micro grid) level analysis.

In this paper the data set consisting of load profile consumption data. It apply the different data mining techniques such as data preprocessing, statistical analysis, frequent patterns and association, clustering data classification, cluster prediction.

### 1.1 Problem statement

Proposed work focuses to electricity consumption analysis using Data mining techniques. It will illustrate the purpose and complete declaration for the development of the system. It will also explain the system constraints and interactions with the different data set.

### 1.2 Scope

Electricity analyzer is useful for analysis of electricity consumption. The paper will classify the electricity consumption data of consumers and thus create a profile according to attributes which approximate their load shapes. This automated system will help to avoid accidents and reduce congestion

- To build models which approximate their load shapes for certain subsets of customers and self-reliance estimates for those load shapes for different consumer weather conditions, times of year, and days of the week.
- Producing models and visualizations which helps us to promote understanding underlying structures which may be in load profile data and to identify the relative importance and interactions between various useful variables.
- Understanding of electrical usage patterns within a domestic.

## 2. Related work

### 2.1 Fintan McLoughlin, Aidan Duffy, Michael Conlon, “A Clustering Approach to Domestic Load Profile Characterization Using Smart Metering Data”, Conference Paper 2015.

This paper presented a clustering methodology for creating a series of representative electricity load for domestic sector. The availability of increasing amount of data to electricity utilities through implementation of traditional way of analysis using meter reading information such as descriptive statistics which was very difficult. When processes are applied on data there were possibilities of losing data therefore other methods of analyzing data need to be used so that this information is not lost.

### 2.2 Thair Nu Phyu. Survey of classification techniques in data mining. In Proceedings of the International Multi Conference of Engineers and Computer Scientists, volume 1, 2009.

This paper presents several major kinds of classification method including decision tree induction, Bayesian networks, k-nearest neighbor classifier, case-based reasoning, genetic algorithm and fuzzy logic techniques. The goal of this survey was to provide a comprehensive review of different classification techniques in data mining. Decision trees and Bayesian Network (BN) generally have different operational profiles, when one is very accurate the other is not and vice versa.

## 3. Overview

It is widely understood that electricity plays a vital role in both the production and consumption of goods and services within an economy. To analysis of electricity consumption it use the best computational technique that is data mining techniques. Data mining techniques include data pre-processing, data analysis, frequent pattern mining, clustering, classification and prediction. It use Google forms to collect the information of each home that includes all the information about monthly consumption, home appliances, number of members, number of persons etc. This data set is in raw format therefore pre-processing of data set is require. Data discretization is a part of data pre-processing with particular importance and in the form of ranges or binding especially for numerical data. Selection of attributes such as units, rooms, person, equipment, water source, type of residence. Clustering is a collection of same group similar or data object or in other groups. The main objective is to find similarities between data objects with the help of specific characteristics found in the data and grouping these similar data objects into clusters. For clustering K-mean algorithm is used in the paper.

## 4. Data mining techniques

In this paper, it use different data mining techniques for analysis of electricity load profile. Figure 1 shows the process how it apply different data mining techniques on data and detail discussion on these in the following section.

- A. Data Preprocessing Data in the real world is raw format or incomplete that lacking new attribute values and certain interesting attributes, or sometimes data having only aggregate data like in electricity load profile and sometimes data having noisy containing errors or outliers. Also inconsistent containing discrepancies in the data in the form of codes or names. Major preprocess techniques are follows: 1) Data cleaning: The data is filled with any missing values, smooth noisiness in the data, identify or try to remove outliers, and resolve inconsistencies in the data. 2) Data integration: In this process the integration of multiple databases, data cubes, or files into a single or useable format. 3) Data transformation: In this phase data further process for normalization and aggregation of data. 4) Data reduction: In this process reduced representation in volume but produces the same or similar analytical results. 5) Data discretization: This process also part of reduction but with particular importance and in the form of ranges, or binning especially for numerical data.

## B. Clustering

Clustering is a collection of same group similar or data objects or in other groups. Its also finding the dissimilar to the data objects in other groups. In cluster analysis, the main objective is to find similarities between data objects with the help of specific characteristics found in the data and grouping these similar data objects into clusters. 1) Clustering Algorithms: In this paper, it use partitioning base clustering method. Partitioning a database D of n objects into a set of k clusters, such that the sum of squared distances is minimized (where  $c_i$  is the centroid or medoid of cluster  $C_i$ ).It use two algorithms K-mean and K-medoids. K-means (MacQueen's, Lloyd's): Each cluster is represented by the center of the cluster.K-means is an unsupervised partition classification algorithm, which requires the exact information of the number of clusters in order to operate

## C. Data Classification

Classification is also known as pattern recognition, discrimination, or supervised learning. There is a lot of Classification approaches, including the use of decision trees and rule induction; density estimation; and artificial neural networks. A decision tree is a tree-like graph structure or model. It also looks like an inverted tree because it grows downwards and it has root at the top. The representation of the data provides the user more advantage compared with other methods that are being meaningful and easy to understand or interpret. The goal is to create a classification model that use for prediction

## 5. Functional requirements

Functional requirements define the fundamental actions that take place in the software in accepting and processing the data-sets and in processing and generating the graphs and predict.

1. Accept the data-sets then clean csv files and assign a class: The system shall accept the data-sets of electricity consumption. Apply Data pre-processing techniques on given data to clean the csv files. Assign the classes on the basis of attributes.
2. Perform clustering algorithm: For clustering it use K-mean's Clustering Algorithm in which each cluster is represented by the center of the cluster. Clustering algorithms make heavy use of distance between two vectors to compare data points and clusters. It use Euclidean Distance formula to calculate distance between clusters.
3. Create rules and predict the consumption for next month: Create the rules for clusters having similar attributes so amount of electricity to be generated can be predicted.
4. Show graphical representation: The graphical representation of data provides the user more advantage compared with other methods that are the meaningful and easy to understand.

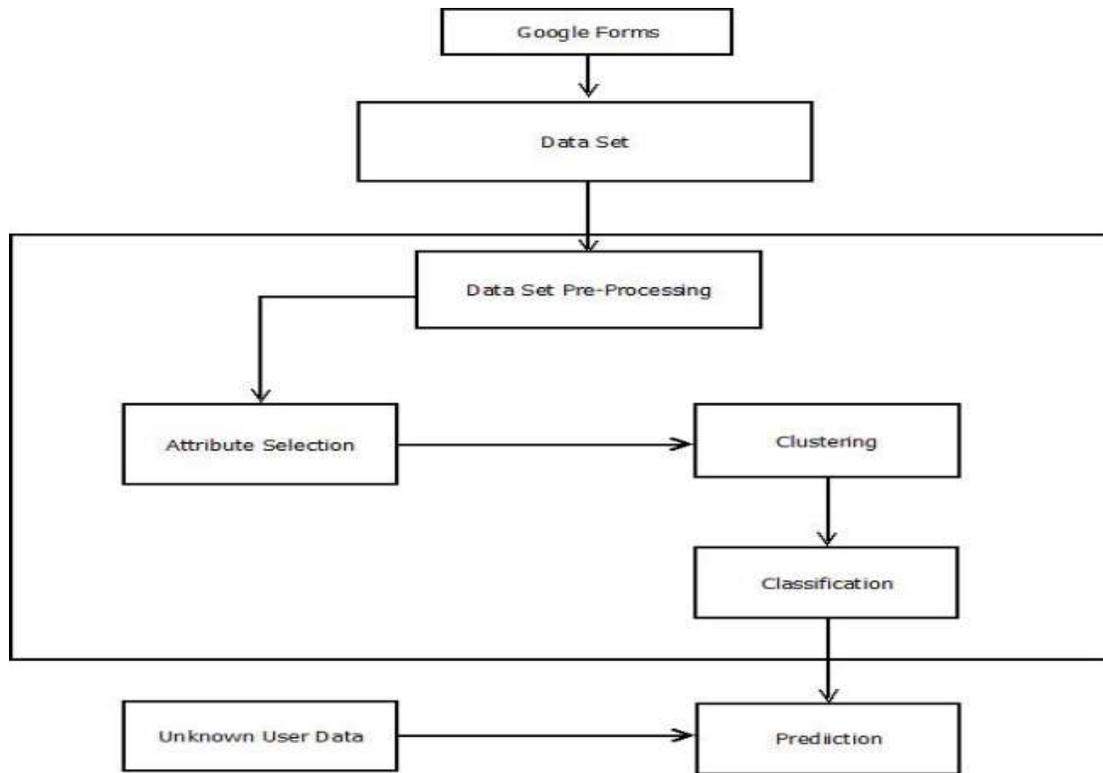


Figure 1 : Module Diagram of domestic electricity consumption analysis using data mining techniques

As shown in Module diagram through google form data will be accepted and will be kept in data set. Further dataset preprocessing is done for appropriate mining algorithm K means algorithm is applied on preprocessed data set after attribute selection. Then result analysis is done. That will be useful in cluster predictions.

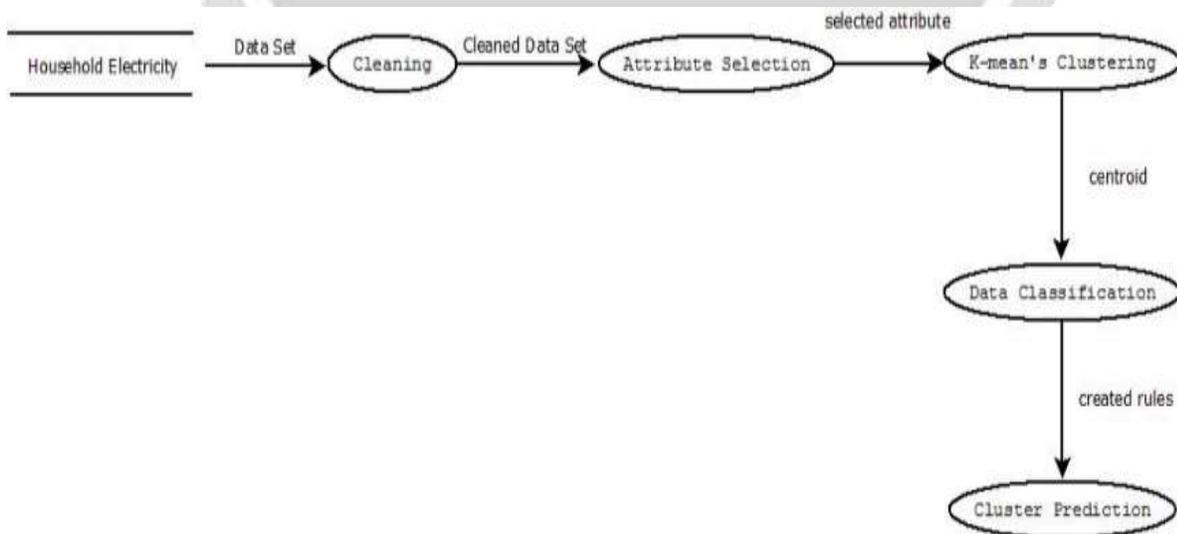


Figure 2: Data flow Diagram of domestic electricity consumption analysis using data mining techniques

## 6. CONCLUSIONS

In this paper, it discuss in detail major data mining techniques for understanding electricity load profile. After applying these techniques on datasets it can easily understand the significance of data. Also analyze how it can improve the new power system by understanding large scale load profile data. All these techniques useful in the peak load reduction. The detail domestic level datasets understanding which domestic level devices take more load consumption as compared to other. That will also helpful in demand side management. In future, it can improve the results by applying these results in real time or more detail level datasets like seasonal and user socio-demographic information.

## 7. REFERENCES

- [1] Usman Ali, Carlo Cecati, "Household Electricity Analysis Consumption Analysis with Data Mining Techniques", Conference Paper October 2016.
- [2] Eintan McLoughlin, Aidan Duffy, Michael Conlon, "A Clustering Approach to Domestic Load Profile Characterization Using Smart Metering Data", Conference Paper 2015.
- [3] Tri Kurniawan Wijaya, Dipyaman Banerjee, "A Cyber Physical Simulator for Demand Respond Systems", 2014.
- [4] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht, "Smart: An open data set and tools for enabling research in sustainable homes", August 2012.
- [5] Rakesh Agrawal, Tomasz Imielinski, and Arun Swami, "Mining association rules between sets of items in large databases", ACM, 1993.