CLASSIFICATION OF CUSTOMER BEHAVIOUR BASED ON SMART METER DATA

SHEIKH AMIR ASYRAAF BIN SHEIKH YUSOFF

BACHELOR OF ELECTRICAL ENGINEERING WITH HONOURS
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

CLASSIFICATION OF CUSTOMER BEHAVIOUR BASED ON SMART METER DATA

SHEIKH AMIR ASYRAAF BIN SHEIKH YUSOFF

A report submitted
in partial fulfillment of the requirements for the degree of
Bachelor of Electrical Engineering with Honours

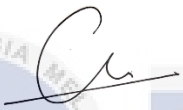
Faculty of Electrical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2021

DECLARATION

I announce that this study entitled 'CLASSIFICATION OF CUSTOMER BEHAVIOUR BASED ON SMART METER DATA is the result of my own research, except as indicated in the references. The thesis has not been approved for any degree and is not submitted simultaneously to any other degree applicant.

Signature : 

Name : SHEIKH AMIR ASYRAAF BIN SHEIKH YUSOFF

Date : 3 July 2021

APPROVAL

I hereby declare that I have reviewed this study entitled "CLASSIFICATION OF CUSTOMER BEHAVIOUR BASED ON SMART METER DATA" and that, in my view, this thesis complies with the partial fulfillment of the award of the Bachelor of Electrical Engineering with Honours.

Signature : .....

Supervisor Name : DR. INTAN AZMIRA BINTI WAN ABDUL RAZAK

Date : 5.7.2021

DEDICATIONS

I want to devote my Final Year Project Project report to my beloved parents, who are always brave to my report and offer moral support. In addition, it is my sincere gratitude and warmest respect that I devote this project report to my supervisor to assist and help me in making this project and report successful.

ACKNOWLEDGEMENTS

In this project, I have made efforts. Without the kind assistance and aid from several individuals and organisations, it would not have been possible. I want to give all of them my heartfelt gratitude. I am deeply grateful to my previous supervisor Dr. Khairul Anwar bin Ibrahim and my current supervisor Dr. Intan Azmira binti Wan Abdul Razak for their advice, continuous oversight, required project knowledge and encouragement in completing this project. I also want to appreciate Dr. Nurdiana binti Nordin for her guidance on the coding. Next, I want to thank Dr. Harriezan Ahmad who is my supervisor from TNB for the valuable guidance and advice to succesfully complete this project proposal. Last but not least, I want to express my appreciation for the kind cooperation and support of my parents and colleagues, who helped me to complete this project.

ABSTRACT

Smart grids contribute to balancing output, delivery and usage by gathering information on the network. There were massive amount of data collected by the smart meter. Thus, the information data must be classify first into a few cluster to make sure the process of balancing between supply and demand more accurate and efficient. The main objective for this project are to simulate and perform the classification of customer behaviour based on smart meter data and to analyse the result of the classification on the customer behaviour. There are a few techniques that can be used to achieve the objective such as k-means, hierarchical and natural jenks natural break. After brief research and comparison, the method of K-means clustering techniques was used to classify the customer behaviour in this project. As a result, the customer was classified into domestic cluster, commercial cluster and industrial cluster by doing 4 case study which is clustering daily profiles, clustering daily mean working days profiles, daily mean weekend profiles and weekly mean profiles.

ABSTRAK

Grid pintar menyumbang untuk menyeimbangkan output, penghantaran dan penggunaan dengan mengumpulkan maklumat di rangkaian elektrik. Terdapat sejumlah besar data yang dikumpulkan oleh meter pintar. Untuk menggunakan data sepenuhnya, data mesti dianalisis dengan mendalam dan proses ini memerlukan banyak masa dan wang. Oleh itu, data maklumat mesti diklasifikasikan terlebih dahulu menjadi beberapa kelompok untuk memastikan proses pengimbangan antara penawaran dan permintaan lebih tepat dan efisien. Objektif utama projek ini adalah untuk mensimulasikan dan melakukan klasifikasi tingkah laku pelanggan berdasarkan data meter pintar dan menganalisis hasil klasifikasi terhadap tingkah laku pelanggan tersebut. Terdapat beberapa teknik yang boleh digunakan untuk mencapai objektif seperti k-means, hierarki dan "jenk natural break". Setelah membuat kajian dan perbandingan ringkas, kaedah teknik pengelompokan "K-means" digunakan untuk mengklasifikasikan tingkah laku pelanggan dalam projek ini. Sebagai hasil projek, pengguna dapat diklasifikasi kepada kluster domestic, kluster komersial dan kluster industri dengan melakukan 4 kajian kes iaitu, pengklusteran harian, pengklusteran purata harian hari bekerja, pengklusteran purata harian hujung minggu dan pengklusteran purata mingguan.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AMI	-	Advanced Metering Infrastructure
TNB	-	Tenaga Nasional Berhad
DSM	-	Demand Side Management
CVMM	-	C-Vine Mixture Copula Models
GVF	-	Goodness of Variance Fit

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

The introduction of the study was clarified in this chapter, following the background of the project and the problem statement. For the background of the project, a detail brief on the development of the customer classification from smart meter data has been explained. Moreover, this chapter also discuss about why should I choose this topic which is known as problem statement and motivation. Next, objectives and scope for this project also explained through this chapter.

1.2 Project Background

Smart grids are implementations of smart energy systems that alter the electricity flow between producers and consumers through communication technology and digital data. Smart grids contribute to balancing output, delivery and usage by gathering information on the state of the network. Predictive, communicative and controllable would become the effect of the electrical system [1]. The goal is to create the power grids more versatile, able to incorporate more specifically the activities of all linked users, including manufacturers, customers and users, to make sure lower cost coordination between supply and demand, while at the same time encouraging effective integration of green technologies such as renewable energy.

A two-way flow of power and knowledge has revolutionized energy production and use through the smart grid. Advanced Metering Infrastructure (AMI) has gained growing prominence around the world as an effective demand-side data source. The electrical industry will have a better understanding of electrical consumption behavior by making full use of the information gathered by AMI. Researchers have investigated several data mining strategies for load profiling from the viewpoint of various technological methods, such as direct clustering, indirect clustering, clustering assessment requirements, and customer segmentation, to incorporate this strategy [2].

1.3 Motivation

Smart meter data will offer a great deal of benefits to suppliers and customers. Smart meter data must be processed, grouped into a few clusters and configured to ensure that smart meter data is useful in our everyday lives. How to derive valuable data from the enormous information volume recorded at distribution network terminals such as households and office buildings is part of the main issue of smart metering. For instance, a smart meter has time interval of 30 minutes and total of one million meter in one distribution network, so it will gain more than 15 billion data recorded for one year [1]. Machine learning methods are a useful collection of algorithms that can make evaluation of the data become easier, interpret and discover hidden information in smart meter data to solve this question.

Data clustering is to identify related groups with the same trend of consumption. Data clustering extensively researched to group customers and expose their behavior in energy usage, leading to personalized energy management systems for individual users. In addition, customer classification data can be used to forecast the consumption of electricity so that energy supply companies such as TNB to estimate the amount of energy expected to be supplied to the user and less power dissipation can occur.

It can also enable us to take care of the atmosphere and can significantly cooperate with the fragile economy during the Covid-19 pandemic by balancing the supply and demand of electricity so there was no energy waste. All these advantages have been enormous factor and incentive for me to successfully complete this project.

1.4 Problem Statement

In this urbanisation era, most of electrical consumer in Malaysia equipped with smart meter especially in Melaka, Kuala Lumpur, Selangor and Putrajaya. Smart meter was used as an replacement to traditional meter to collect the data on consumption of electricity for consumer. In the course of communication and metering, technical advancements enable utilities to analyze and store vast quantities of data relevant to the services. The deployment of 2-way power meters enables user consumption to be recorded in a high-resolution manner.

However, the customer behaviour was still unknown because the data always random and different between every users. It will cost a lot of money and time to analyze the data of the smart meter one by one for each day. The high demand for sampling rates in smart metering and the massive number of users mean that vast volumes of data need to be saved and analyzed more complexly. To solve this problem, classification of the customer into some cluster based on their consumption on electrical usage data collected by smart meter needed to be done. This information can help them plan the grid, target consumers in response to demand and energy efficiency programmes, getting more accurate energy forecasts and others.

The typical consumption curves of consumption can provide valuable insights into the different consumer groups such as for the design systems, consumer engagement policy, communications, alternative tariff set up approaches and demand forecasting devices. Computational intelligence methods are beginning to be used widely in the course of machine learning to derive information from data from the power source [3].

1.5 Objectives

1- To study and analyses the load profile data from smart meter.
2- To model the cluster of customers based on K-means clustering technique by using Python software.

1.6 Scope

1- The hourly load data were provided by the industry which is Tenaga Nasional Berhad (TNB).
2- The load data were taken from 737 customers randomly in Peninsular Malaysia from 1st August 2017 until 1st September 2017.
3- Modelling of customer classification was designed using K-means techniques and simulated using Python software.

1.7 Report Outline

Five chapters are included in this project report:

The background, statement of problem, goals, and scope for this project is clarified in Chapter 1.

Chapter 2 discusses the literature analyses of the consumer classification and the smart meter.

Chapter 3 includes information on the manner in which the aims of this project are accomplished.

The findings will be discussed in Chapter 4 based on the proposed method in Chapter 3.

The conclusion of this project's final report and some potential work that needs to be carried out includes in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This section is about the review on article, journal, website, book and other sources on the topic. A simple and quick understanding of the data classification, load profile, customer classification and smart meter. All the details have been discussed in this chapter.

2.2 Data Classification

Data classification is defined as the method of grouping knowledge through appropriate categories so that it can be used more effectively. The classification process basically makes it easier to find and retrieve data. In terms of risk management, data management, enforcement, and data security, the classification of data is especially important. Classification of data requires marking data in order to quickly scan and trace. It also eliminates various duplications of records, which can reduce storage and backup costs and improve the search process.

The technology is currently used for a variety of purposes, primarily to support data security. Data are often listed to be available conveniently, to ensure compliance with regulations and to meet some other business or personal goals. In certain cases, the classification of data is a legal requirement under certain timeframes [4].

2.2.1 Process of Data Classification

There were 3 steps for processing the data classification such as below:

2.2.1.1 Understand the Current Setup

Taking a proper look at existing data and all corporate rules is perhaps the best starting point for effective classification data. Before doing classification, the details must be define. Data comprehension is an important feature of classification.

2.2.1.2 Create a Data Classification Policy

Without appropriate policy it is virtually difficult to be in line with data security standards in an enterprise. Our highest goal should be to build a strategy.

2.2.1.3 Prioritize and Organize Data

Now that our existing data is protected by a policy and an image, now is the time to identify the data perfectly. Choose the right way to arrange our data on the basis of your collected information.

2.2.2 Benefits of Data Classification

The data classification has many advantages, such as making it easy to search data. In order for modern organizations to get an understanding of the vast number of data available at any moment, it is important to identify data. Data classification gives a straightforward view of all data under an organization's jurisdiction and an understanding of where the data is saved, how to effectively access the data and how to ensure it is better secured from future security hazards. When the data classification is applied, it offers a structured structure that makes for more appropriate details.

2.3 Load Profile

The customer load profile is the identification of each customer's energy consumption trend. There is a need to consider past consumer load profile data properly. All customers' added load profile data is the electricity needed by customers. Load profile data are also the deciding appliances for the maximum generation of electricity. Optimal energy output must concentrate and balance demand and power surpluses. Other characteristics that impact energy use are the load profile for consumers. Demographic variables like occupancy levels, equipment number, environmental factors such as temperature and others. For the optimum generation of load profiles, the electricity load profile data along with the most control population value are considered.

The generation of load profiles is a blueprint for various user trends [5]. In addition, the load profile also ensures the equilibrium between manufacturing and use.

For companies concerned with electricity transmission and electricity delivery or in the electricity sector, this knowledge about demand for loads is very necessary to have adequate capacity. Load profile can be extended to many areas, such as dynamic price control, load management, demand management, vendor optimization, load forecasting and automation. Figure 2.1 is an example of load profile for one of the smart data recorded in the database. For every one hour interval, the eletrical power usage was recorded and plot into graph as below.

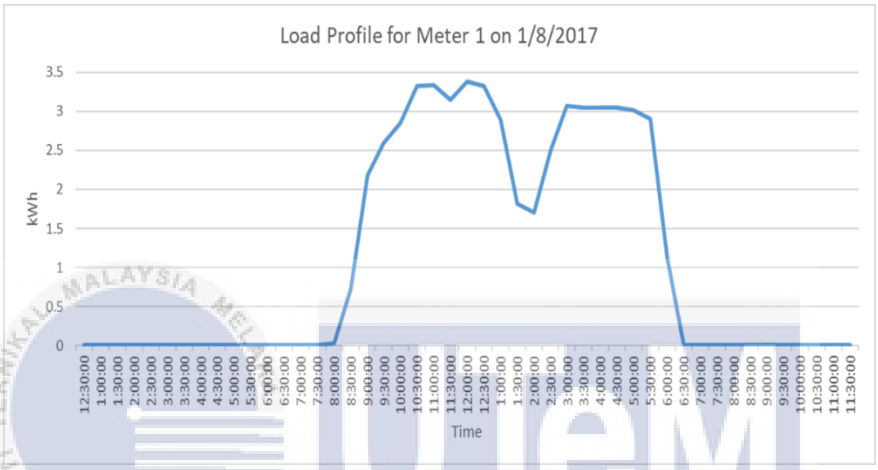


Figure 2.1: Example of Load Profile

2.3.1 Type of Load Profile

Generally, there are 3 type of load profile in Malaysia which are domestic, commercial and industrial. Each of the type had different behaviour and pattern of power consumption.

2.3.1.1 Domestic Load Profile

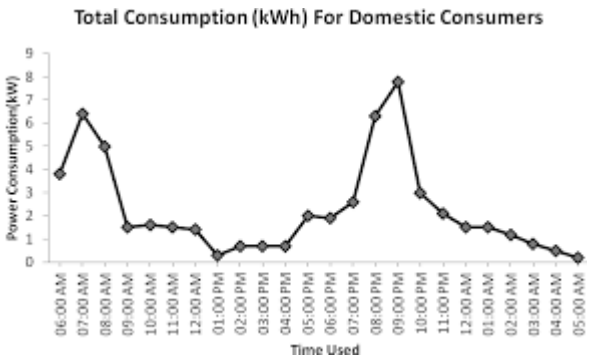


Figure 2.2: Domestic Load Profile

Figure 2.2 shows the average total consumption for domestic customer in Malaysia [6]. It shows that the peak hour electricity consumption is between 7.00 am to 9.00 am and from 7.00 pm to 10.00 pm.

2.3.1.2 Commercial Load Profile

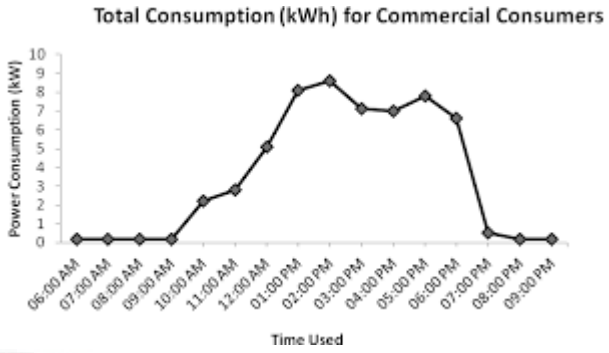


Figure 2.3: Commercial Load Profile

Figure 2.3 shows power consumption profile for electrical appliances and equipment for commercial sector. It shows the peak hour for using this building operated is between 10.00 am to 6.00 pm [6].

2.3.1.3 Industrial Load Profile

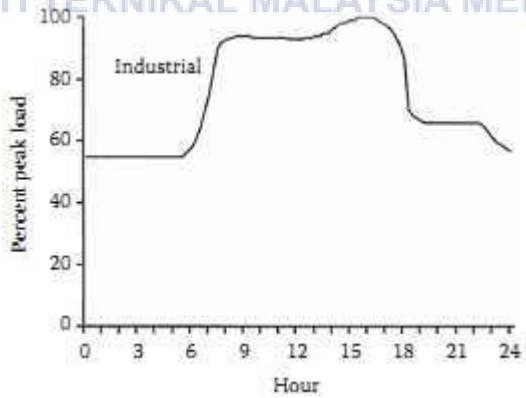


Figure 2.4: Industrial Load Profile

Figure 2.4 shows power consumption profile for electrical appliances and equipment for industrial sector. It shows the peak hour for using this building operated between 7.00 am to 6.00 pm [6].

2.4 Customer Classification

Customer classification is the act of finding out and identifying common characteristics in a group of customers. The most regular electrical load type is in accordance with the billing types used by utilities. This includes residential, corporation, industrial and other classifications. Home customers are domestic consumers while, clearly, business and manufacturing users are business customers. Such consumer classifications cover counties, federal and state governments, power cooperatives and educational establishments [7].

The clustering of electricity usage knowledge has been the subject of significant studies in recent years to assess the actions of electricity customers. The normal applications range from DSM modeling and simulation, load forecasting, tariff setting, marketing and incorrect data sensing. The clustering strategies that have been established are mainly K-means [3]. Data planning in these applications is of considerable significance, which specifies the information to be derived from the clustering and the capability of the approaches used to produce successful performance.

2.4.1 Purpose of Customer Classification

Nearly any power source worldwide uses consumer load profiles for grid optimization, market forecasting and demand side control. This is also a critical technique for improving organizational efficacy and stability of the power grid [7].

2.4.2 Advantages of Customer Classification

A better understanding of the load profile would allow utilities to provide an efficient marketing policy by supplying their consumers with a better tariff structure, as compared with typical Inclining Block Tariff, for example time-of-use. It also allows the utility to upgrade and modernize its operating processes to make service and energy generation more effective [7].