Clustering Commercial And Residential Electricity Consumption Using K-Means Algorithm

Esnehara Bagundang, Cyrus Rael

Abstract: This study implemented K-means Algorithm to cluster the data set of electricity consumption of clients. The data set was obtained from the Meter Reader Billing Statement System of Sultan Kudarat Electric Cooperative, Inc. (SUKELCO). It aimed to cluster the electricity consumption of commercial and residential clients for the period of four months (January-April 2021). The result of this study shows an interesting fact that majority of both commercial and residential clients belongs to the group with low electricity consumption and there is an increase demand of electricity each month.

Index Terms: Clustering, K-Means Algorithm, Electricity Consumption

1 INTRODUCTION

Sultan Kudarat is a province located in Region 12-SOCCSARGEN, Mindanao, Philippines. It is composed of 12 municipalities namely Bagumbayan, Columbio, Esperanza, Isulan, Kalamansig, Lambayong, Lebak, Lutayan Palimbang, President Quirino, Senator Ninoy Aquino and Tacurong. Sultan Kudarat Electric Cooperative, Inc. (SUKELCO) is a non-profit entity that provide adequate and reliable electric service to its consumers. It covers all the 12 municipalities and divided into areas where area 1 is composed of Isulan and Esperanza. Electricity is one of the human necessities proportional to the increase in human population and economic growth. It is part of the modern life where it supports a lot of activities involved by both residential and commercial clients in a society. Commercial clients pertains to every electricity consumption of commercial buildings while residential clients pertains to every household electricity consumption. Due to the increase demand of electricity, it is important to have knowledge about the electricity consumption pattern of the clients. It can be the basis to determine the amount of electricity production and electricity distribution for specific location and type of client. In order to obtain knowledge on the electricity consumption pattern of clients, clustering technique will be used to come up with electricity consumption profile. It has the ability to measure the similarity between the data sets and group them based on electricity consumption. In this paper, the goal is to cluster the electricity consumption of residential and commercial clients for the period of January-April 2021 of two (2) towns in Sultan Kudarat.

2 REVIEW OF RELATED LITERATURE AND RELATED WORKS

2.1. K-Means Algorithm

K-Means Algorithm is an unsupervised learning algorithm that is used for clustering problems. It allows to group unlabeled dataset where this group are represented by their centers. It aims to minimize the sum of distance between the data point and their corresponding clusters [1]. This algorithm follows a simple and easy way procedure to classify the given data set through a specific number of clusters (k). It takes an unlabeled data set as input and performs two tasks: (1) It determine the best value for K - center points through an iterative process and (2) It creates a cluster for data point which are to a particular k-center [2].

2.2. Elbow Method

Elbow Method is one of the ways in determining the optimal number of cluster in K-Means Algorithm. This method uses the concept of Within Cluster Sum of Squares (WCSS). The WCSS defines the total variations within a cluster. In implementing this method, upon fitting the model with a range values of K and displaying the WCSS in contrast to the number of clusters, the point where it displays a sharp bend like an elbow in the graph means that is the optimal number of cluster [3].

2.3. Outlier Detection

Outliers are data objects which lies in a different distance from the other data objects. It occurs due to variability in the measurement or changes in a system behavior. In data mining, detecting the outlier and removing outlier makes the clustering more reliable [4]. Thus, it is important to remove outliers because including it may produce inaccurate result due to the damaging characteristics of the data processed.

2.4. Data Mining on Electricity Consumption

The use of Advanced Metering Infrastructures (AMI) enables to investigate the characteristics of household energy consumption thus various studies have conducted that employed clustering techniques on the electricity consumption of clients. A study used a data set from a smart meter and implements K-means clustering on energy consumption to group households based on 3 evaluation metrics such as Davies-Bouldin Score, Calinski-Harabaz Score and Silhouette Score [5]. Another study implemented a clustering technique to the data set of electricity consumption of residential buildings. Specifically, it aimed to determine the household appliances that uses the high power consumption [6]. Further, a study implemented K-Means Algorithm to cluster a data set of electricity consumption of 370 clients collected in a year. The data set was aggregated into 4 seasons. It employs 5

- *Esnehara Bagundang is a graduate of Master of Science in Information Technology (MSIT) at Ateneo de Davao University. She is currently a faculty at Sultan Kudarat State University (SKSU). E-mail: esneharabagundang@sksu.edu.ph*
- *Cyrus Rael is a graduate of Master in Information Technology (MIT) at Notre Dame of Marbel University. He is also a faculty at Sultan Kudarat State University(SKSU) E-mail: cyrusrael@sksu.edu.ph*

clusters without eliminating the outliers. Result of this study obtained an interesting pattern where Cluster 0-3 had the biggest average use of electricity in summer and smallest average use in winter. Cluster 4 had the smallest average use of electricity in spring [7].

3 METHODOLOGY

3.1 Data Set

The dataset was obtained from the Meter Reader Billing Statement System of Sultan Kudarat Electric Cooperative, Inc. (SUKELCO). It contains the electricity consumption (kwh) of four type of clients such as Commercial (C), Residential (R), Public Building (P) and Street Light (S) covering January, Februay, March and April 2021. It consists of the following (13) attributes: ConsumerType, BillMonth, Area, TownCode, RouthCode, Account, PresentReading, PreviousReading, KWH Use, TotalBill, Payment Reference, PaymentDate, BillDate.

3.2 Data Preparation

From the original data set it was arranged such that one row represents one client with corresponding electricity consumption (kwh) for the month of January, February, March and April 2021. Data cleaning were employed such as removing rows with null values. Figure 1 shows the example of derived data set that contains the following (7) attributes: ConsumerType, TownCode, Account, Jan, Feb, March, April.

	ConsumerType	TownCode	Account	Jan	Feb	March	April
0	C	35.0	1030.0	56.0	57.0	64.0	75.0
1	C	35.0	1045.0	39.0	43.0	40.0	45.0
2	C	35.0	1050.0	88.0	85.0	77.0	110.0
3	C	35.0	1055.0	212.0	262.0	258.0	236.0
4	C	35.0	1060.0	1121.0	1263.0	1408.0	1814.0
...	...	...	...	...	...	...	...
22641	R	40.0	4700.0	37.0	38.0	42.0	69.0
22642	R	40.0	4800.0	107.0	166.0	176.0	193.0
22643	R	40.0	4900.0	149.0	175.0	163.0	198.0
22644	R	40.0	5000.0	5.0	5.0	5.0	5.0
22645	R	40.0	5001.0	29.0	29.0	24.0	14.0

Columns Description: ConsumerType contains the type of the client such as "C" for Commercial, "R" for Residential, "P" for Public Building and "S" for Street Light. TownCode contains two type of values such as "35" for Isulan and "40" Esperanza. Account contains the account number of each clients. Jan contains the electricity consumption in kwh for January 2021. Feb contains the electricity consumption in kwh for February 2021. March contains the electricity consumption in kwh for March 2021. April contains the electricity consumption in kwh for April 2021.

3.3 Experiment

Though the data set contains four type of client, the focus of this study is on the electricity consumption of commercial and residential clients. The data set can be visualized using matrix D as follows:

$$Dc \mid Dr = \begin{bmatrix} X1 & X2 & X3 & X4 \\ \vdots & \vdots & \vdots & \vdots \\ Xn & Xn & Xn & Xn \end{bmatrix}$$

Dc will be used for clustering the electricity consumption for commercial clients. Dr will be used for clustering the electricity consumption for residential clients. The first column is the amount of electricity consumption in January, the second column is the amount of electricity consumption in February, the third column is the amount of electricity consumption in March and the fourth column is the amount of the electricity consumption in April.

3.4 Tools

In this study python and various libraries such as pandas, matplotlib and scikit-learn were used through Jupyter notebook to perform data visualization, analysis and the clustering task.

4 RESULT AND DISCUSSION

4.1 Data Visualization

4.1.1 Commercial Clients

A total of 1,435 rows of commercial clients were obtained in the data set. Figure 2, 3, 4 and 5 shows the electricity consumption of this type of client in January, February, March and April 2021 respectively. The x-axis is the client and the y-axis is the electricity consumption.

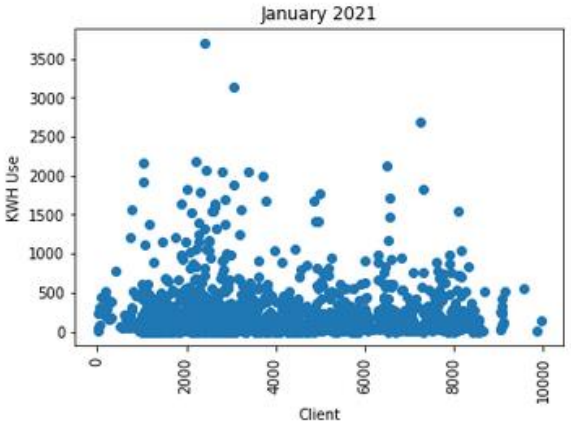

Figure 2. Electricity Consumption Profile of Commercial Clients in January 2021

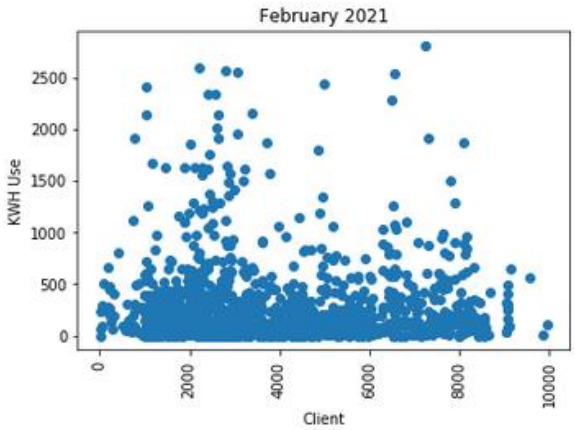

Figure 3. Electricity Consumption Profile of Commercial Clients in February 2021

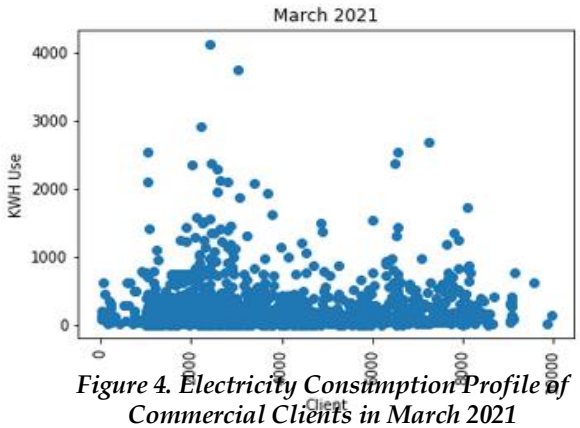

Figure 4. Electricity Consumption Profile of Commercial Clients in March 2021

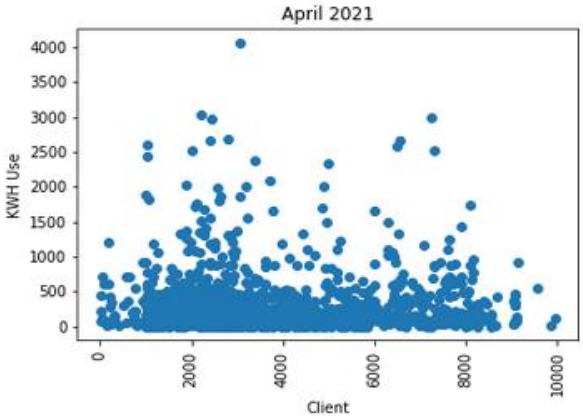

Figure 5. Electricity Consumption Profile of Commercial Clients in April 2021

4.1.2 Residential Clients

A total of 20,001 rows of residential clients were obtained in the data set. Figure 6, 7, 8 and 9 shows the electricity consumption of this type of client in January, February, March and April 2021 respectively. The x-axis is the client and the y-axis is the electricity consumption.

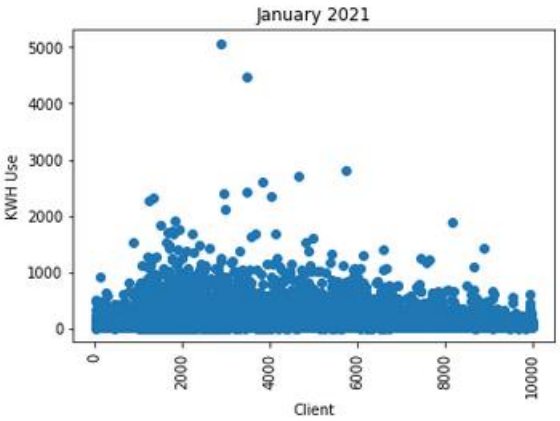

Figure 6. Electricity Consumption Profile of Residential Clients in January 2021

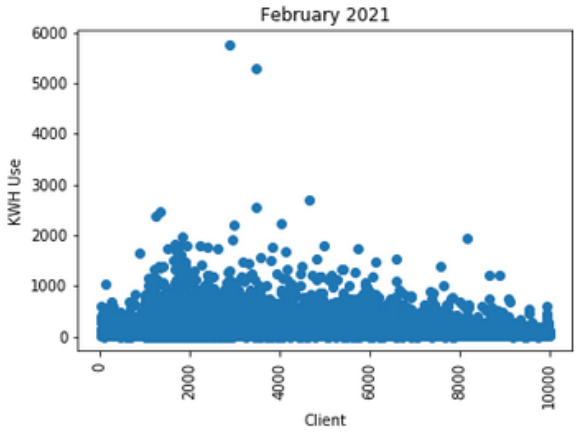

Figure 7. Electricity Consumption Profile of Residential Clients in February 2021

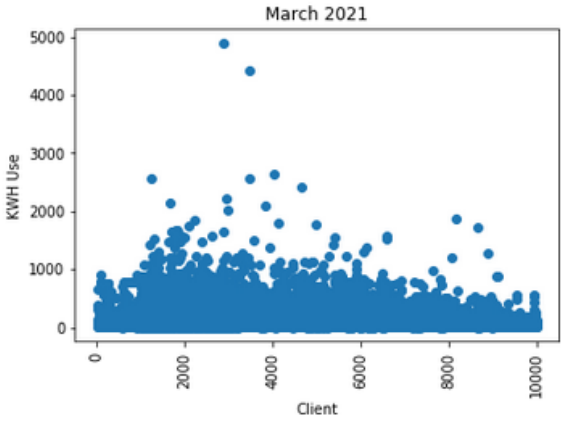

Figure 8. Electricity Consumption Profile of Residential Clients in March 2021

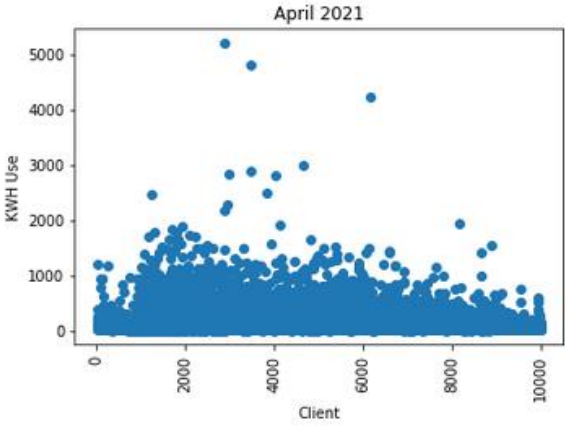

Figure 9. Electricity Consumption Profile of Residential Clients in April 2021

4.2 K-Means Clustering

4.2.1 Commercial Clients

In clustering the electricity consumption of commercial clients we use the data Dc as input and outliers were removed. Using the Elbow Method it shows the optimal number of cluster as 3. Table 1 shows the cluster result of electricity consumption of

commercial clients. It can be seen that Cluster 0 has the highest average electricity consumption followed by Cluster 1 then Cluster 2. Cluster 2 has the highest number of cluster member followed by Cluster 1 then Cluster 0.

Table 1. Commercial Clients Clustering Result

Cluster Number	Number of cluster member	V1	V2	V3	V4
0	12,407	49.85	48.45	47.98	54.55
1	5,789	163.43	164.12	162.28	181.59
2	1,805	354.73	370.79	362.76	406.96
TOTAL	20,001				

4.2.2 Residential Clients

In clustering the electricity consumption of commercial clients we use the data Dr as input and the same process was done where the outliers were removed and using the Elbow Method it shows the optimal number of cluster as 3. Table 2 shows the cluster result of electricity consumption of residential clients. It can be seen that Cluster 2 has the highest average electricity consumption followed by Cluster 1 then Cluster 0. Cluster 0 has the highest number of cluster member followed by Cluster 1 then Cluster 0.

Table 2. Residential Clients Clustering Result

Cluster Number	Number of cluster member	V1	V2	V3	V4
0	98	748.44	818.29	781.50	897.27
1	366	313.80	339.42	342.15	371.07
2	971	69.52	70.40	71.26	77.34
TOTAL	1,435				

Table 1 and Table 2 Description: V1 is the average electricity consumption (kwh) in January 2021, V2 is the average electricity consumption (kwh) in February 2021, V3 is the average electricity consumption (kwh) in March 2021 and V4 the average electricity consumption (kwh) in April 2021.

4.3 DISCUSSION

4.3.1 Commercial Clients

There are three clusters of commercial clients which was described below:

1. Group 1 (Cluster 0) has the high average electricity consumption with 98 members. The highest average electricity consumption of this group is 1268kwh and the lowest average electricity consumption is 574kwh.
2. Group 2 (Cluster 1) has the mid average electricity consumption with 366 members. The highest average electricity consumption of this group is 575kwh and the lowest average electricity consumption is 203kwh.
3. Group 3(Cluster 2) has the low average electricity consumption with 971 members. The highest average electricity consumption of this group is 207kwh.

4.3.2 Residential Clients

There are three clusters of residential clients which was described below:

1. Group 1 (Cluster 0) has the low average electricity consumption with 12,407 members. The highest average electricity consumption of this group is 89.42kwh and the lowest average electricity consumption is 0.60kwh.
2. Group 2 (Cluster 1) has the mid average electricity consumption with 5,789 members. The highest average electricity consumption of this group is 223.70kwh and the lowest average electricity consumption is 84.13kwh.
3. Group 3 (Cluster 2) has the high average electricity consumption with 1,805 members. The highest average electricity consumption of this group is 515kwh and the lowest average electricity consumption is 214.50kwh.

5 CONCLUSION

Based on the result, this paper obtained an interesting fact that majority of both commercial and residential clients belongs to the group with low electricity consumption and there is an increase demand of electricity in each month. With this, the electricity provider can plan and determine the amount of electricity production in the preceding months.

6 REFERENCES

[1] E. Alpaydin, Introduction to Machine Learning, 2010.

[2] A. Kanungo, David M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu., "An Efficient k-means Clustering Algorithm: Analysis and Implementation," 2002.

[3] M. Syakur, Khotimah, B, Rohman, Eka and Dwi Satoto, Budi, "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster," 2018.

[4] G. Guojun and Michael, Kwok-PoNgb, "k-means clustering with outlier removal," Pattern Recognition Letters, vol. 90, pp. 8-14, 2017.

[5] Z. Zhang and T. Zimet, "K-means Based Clustering Analysis of Household Energy Consumption," 2018.

[6] M. A. R. Fitri, YakubEmail, N. N. I. Sulaiman and M. Z. A. Rashid, "Energy Consumption Clustering Analysis in Residential Building," 2019.

[7] Y. Amri, A. Lailatul Fadhilah, Fatmawati, N. Setiani and Septia, "Analysis Clustering of Electricity Usage Profile Using," IOP Conf. Series: Materials Science and Engineering, 2020.