Prediction of Energy Consumption in a Smart Home Using Coherent Weighted K-Means Clustering ARIMA Model

J. Jasmine Christina Magdalene[1†] and B.S.E.Zoraida[2††]

[1]Bishop Heber College, Affiliated to Bharathidasan University,Tiruchirappalli, India
[2]Bharathidasan University, Tiruchirappalli, India

Abstract
Technology is progressing with every passing day and the enormous usage of electricity is becoming a necessity. One of the techniques to enjoy the assistances in a smart home is the efficiency to manage the electric energy. When electric energy is managed in an appropriate way, it drastically saves sufficient power even to be spent during hard time as when hit by natural calamities. To accomplish this, prediction of energy consumption plays a very important role. This proposed prediction model Coherent Weighted K-Means Clustering ARIMA (CWKMCA) enhances the weighted k-means clustering technique by adding weights to the cluster points. Forecasting is done using the ARIMA model based on the centroid of the clusters produced. The dataset for this proposed work is taken from the Pecan Project in Texas, USA. The level of accuracy of this model is compared with the traditional ARIMA model and the Weighted K-Means Clustering ARIMA Model. When predicting,errors such as RMSE, MAPE, AIC and AICC are analysed, the results of this suggested work reveal lower values than the ARIMA and Weighted K-Means Clustering ARIMA models. This model also has a greater loglikelihood, demonstrating that this model outperforms the ARIMA model for time series forecasting.
Keywords:
ARIMA, Energy Management, Loglikelihood, RMSE, Weighted K-Means Clustering.

1. Introduction

The twenty-first century has experienced a revolution when it comes to the handling of technology. The advancement of technology, in the form of smart devices and appliances, has turned the lifestyle of people simple yet smart. Not just the smart devices and appliances, the entire home is transforming itself into automation and smart. Along with the X and Y's of smart appliances at home, inclusion of electric grid comes very handy to trace and to forecast the consumption of energy. The smart grid assists the consumers to identify the level of consumption in energy and by doing so, can easily guide the consumers in scheduling their devices to reduce pricing. The proper management of energy thus has extended its importance drastically. It not only benefits the consumers but also aids the utilities to manage energy so as to provide unremitting supply of energy to the consumers even during demanding times. To manage the energy resourcefully, it is essential to have a prospect to predict the energy consumption in a smart home. Forecasting helps the consumers and the utilities to understand further about the necessity of energy by every single home during each of the seasons. Towards calculating the consumption of energy, the previously used data will be significantly handy and therefore a customized forecasting method can to be preferred. The data gathered analyzing the consumption of energy is said to be a time series data, as it is sequential and is recorded in a regular break of time. These time series data comprising four major components are trend, seasonality, clinical and randomness [1].

Researchers in the recent years have embraced time series forecasting module as this prediction model helps numerous organizations and businesses to plan for a better usage in the future. There are a number of statistical and machine learning tools to perform prediction. Regression techniques, decomposition models, ARIMA are a few of the most renowned forecasting models and each model differs in their accuracy level. The accuracy of a model is recognized by analyzing the error rate of the predicted value. The lower the error the higher is the accuracy of the model. Advanced researches in forecasting affirm that one of the popular models for time series forecasting is ARIMA [2]. Randomness is one of the characteristics of time series data, and because of this randomness in data, numerous data mining approaches such as categorization, clustering, and indexing havebecome popular [3]. The clustering approach can be used to find unexpected patterns in time series data. Clustering comparable data points prevents additional data points from interfering with that of a given cluster [4].

In this proposed work, the Weighted K-means Clustering algorithm is improvised to enrich the efficiency of the algorithm and the ARIMA model is used after that to predict the average amount of energy consumption during the various seasons in a smart home. The following is the structure of this research project: Section II is devoted to a literature review. In Section III, the suggested work's methodology is presented, followed by the results and discussions in Section IV. Section V contains the conclusion and recommendations for future work.

Manuscript received October 5, 2022
Manuscript revised October 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.10.23

2. Review of Literature

One of the often sought-after areas in the field of scientific research is the prediction of energy management in a smart metropolis and smart home. Many researchers and authors make noteworthy contributions in this research area by giving variety of unique ideas and novel thoughts. Pasapitch Chujai et.al have successfully forecasted the quantity of energy consumption during different times in a year using ARMA and ARIMA models. In their work they have established that ARIMA is superior for medium- term and long-term forecasting and ARMA proves better for short term forecasting [5]. Cristina Nichiforov et.al have equated ARIMA and Non-linear Auto Regressive neural network (NAR) models and determined that ARIMA is better than NAR for the forecasting of energy consumption [6]. Junwei Miao has applied the ARIMA model to forecast the amount of energy consumption in China and concludes that it shows better precision [7]. Suat Ozturk et.al, have forecasted energy consumption of coal, natural gas and oil in Turkey using ARIMA [8]. Sen et.al have done a case study on an Indian Pig iron organization to forecast energy consumption using ARIMA [9]. Warut Pannakkong et al., have given a ground-breaking hybrid model by combining ARIMA, ANN and K-Means for forecasting time series data and comparison is done with the standard ARIMA and ANN models [10]. Grzegorz Dudek has forecasted the next day electric load curve using K-Means Clustering algorithm [11]. Myong-Hoe-Huh and Yong B.Lim have given significance to clustering variables and to add more meaning to variables to circumvent non-coherence [12]. These works that have been carried out in the area of forecasting of energy consumption shows that ARIMA and Weighted K-Means can be used for time series forecasting. The various principles used in the above said models are also considered for predicting the energy consumption in a smart grid. Muhammed Faizan et.al. have done a comparative study on the applications of clustering techniques in data mining [13]. Jamal Fattah et.al have forecasted the supply chain by using the historical data to predict the future food manufacturing demand [14]. These works that have been carried out in the area of forecasting of energy consumption shows that ARIMA and Weighted K-Means can be used for time series forecasting. The various principles used in the above said models are also considered for predicting the energy consumption in a smart grid.

3. Methodology

When we look at the average amount of energy consumed in a smart home over the course of a year, the appliances that are used play a big role. During summer energy consumption will be high due to the use of air conditioners, while in winter energy consumption will be less and in spring energy consumption will be moderate. The quantity of energy consumed by each device varies, as does the amount of energy consumed from the electric grid. Some homes may have renewable resources as well, resulting in a shift in the amount of energy drawn from the grid. In general, the average amount of energy consumed in a smart home changes with the seasons. Because typical energy use varies with the seasons, energy consumption forecasting is necessary. If utilities can forecast average energy use throughout seasons, it will be easier for them to plan energy supply and assist customers during natural catastrophes and other interruptions. It not only benefits utilities, it also benefits consumers by allowing them to track their usage and schedule the smart appliances in such a way as to avoid higher prices. The past data can be used to make this prediction. Given the importance of historical data, a normal data-driven forecasting model such as Holt Winter's cannot provide the desired accuracy, hence a model-based forecasting method must be explored. The accuracy of the forecasting model will be improved by using a model-based forecasting strategy. In this proposed work, the data point which is the average amount of energy consumed in a month during a particular season is given weights according to the energy consumed. After adding weights to the data points, clusters are formed using K-Means clustering algorithm and then the model-based forecasting technique ARIMA is used to forecast the average amount of energy that will be consumed from the smart grid in the forthcoming years during various seasons. This work is based on data from the Pecan Project in Austin, Texas, for a single residence. The average energy consumption from the smart grid is divided into training and test data for the years 2013 to 2019 (seven years). This dataset is used to forecast the average amount of energy consumed in the following years during distinct seasons. The input variable is the average amount of energy consumption from the smart grid over the course of each month and the function used to represent this input variable is f (EnC). Then, depending on the month and season, weights are applied to each data point. Austin has four distinct seasons, and the weights are calculated so that the month with the lowest value in each season is given greater weight. For a season, the weights are summed in such a way that they add up to one. The month with the lowest season value receives the most weight, while the month with the highest season value receives the least. This is done to avoid the data point with the lowest value becoming insignificant. Here 0.5 is given to the month with the least energy consumption value. 0.2 is given to the month with the highest energy consumption and 0.3 is given to the remaining month. Thus the total weights given to a season is 1. The following equations indicate how weights are added to each month in a season

$$F(x) = n1w1 \quad (1)$$
$$F(y) = n2w2 \quad (2)$$
$$F(z) = n3w3 \quad (3)$$
$$\text{Where } w1+w2+w3=1 \quad (4)$$

Eq.1 F(x) indicates the multiplied value of the weight w1 and the datapoint with the least value in the particular season n1. Eq 2 F(y) indicates the product of the weight w2 and the datapoint with the next lower value n2. Eq 3 F(z) indicates the product of the weight w3 and the datapoint with the highest value. Eq 4 shows that the weights that are added should be equal to 1. The same method is followed for all the seasons in a year.Then Clusters are produced using the K-Means clustering technique after multiplying these weights with each data point. The ARIMA model is used to anticipate the average amount of energy consumption from the smart grid during various seasons in the next years after knowing the centroid. The ARIMA model takes the input p,d,q where p indicates the terms in autoregression, d indicates the difference and q indicates the lagged forecast errors.

The RMSE, MAPE, and loglikelihood of this model Coherent Weighted K-Means Clustering ARIMA (CWKMCA), are compared to the regular ARIMA and Weighted K-Means ARIMA models to determine its correctness. R is the tool that iss utilized to complete this proposed work.

3.1. Input Parameters

The average amount of energy consumed in a smart home is mostly determined by the appliances that are used throughout various seasons and the season also has a significant impact on the amount of energy consumed. The average quantity of energy usage in a month from the smart grid is used as the input parameter for this work.

(i) Each month's energy consumption from a smart grid

1) Average Energy Consumption from a Smart Grid:

Many energy sources are available in a smart house, including the electrical grid, solar panels, and wind turbines, among others. Despite the fact that there are numerous sources, most smart homes still rely on the electrical grid. Each appliance in a smart home consumes different amounts of energy and its usage fluctuates with the seasons. As device energy usage varies, so does the amount of energy drawn from the electrical grid. As a result, the amount of energy consumed varies greatly depending on the season.

3.2. Data Pre-processing

The data must first be pre-processed before the forecasting model can be built. The data for this work comes from a smart home in Pecan Project, Texas. Because the data is collected every minute, it is a time series data. From 2013 to 2019, the average amount of energy utilised per month was calculated during a seven-year period. As part of the pre-processing procedure, some null values in the data gathered every minute are filled with the previous value. The data is then translated to a monthly format. Table 1 shows the average amount of energy usage in kilowatts for various months throughout a seven-year period from 2013 to 2019.

Table I
Average amount of energy consumption from a smart grid in KW

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2013	0.55	0.3	0.32	0.15	0.72	1.47	1.49	2.46	1.81	0.99	0.72	0.47
2014	0.36	0.47	0.29	0.47	1.23	1.52	0.64	0.94	0.96	0.64	0.5	0.8
2015	0.55	0.46	0.57	0.54	0.85	1.08	1.38	1.36	0.84	0.81	0.76	0.77
2016	0.33	0.34	0.44	0.55	0.82	1.00	0.76	1.12	1.16	0.63	0.56	0.77
2017	0.66	0.53	0.36	0.41	0.8	0.85	1.89	0.58	0.97	0.46	0.6	0.71
2018	0.55	0.8	0.31	0.35	0.79	1.09	1.48	1.37	1.24	0.75	0.61	0.56
2019	0.4	0.53	0.29	0.35	0.72	0.83	1.03	1.85	1.8	0.88	0.63	0.61

3.3. Construct and Decompose the Time-series Data

The dataset employed in this study is a time series dataset because it is time-related. By setting the frequency to twelve, this data is turned into a monthly time series. This time series data is deconstructed in order to determine whether the data is cyclical and exhibits trend, seasonality, and unpredictability. The decomposition is carried out in order to select the optimal model for predicting the dataset used in this work. Figure 1 depicts the dataset's trend, cyclicality, seasonality, and randomness.

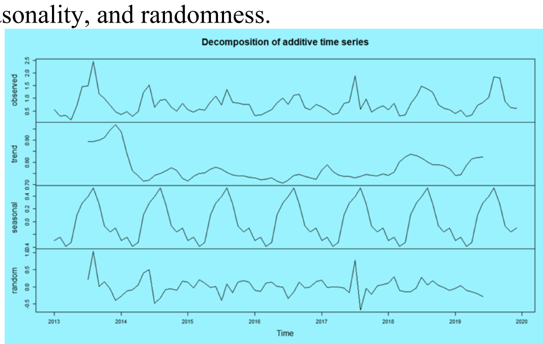

Fig 1: Decomposition of time series data

3.4. Construct the Model

As a first stage, the proposed model is built by assigning weights to the datapoints. Random weights are added to weighted K-Means clustering, but in this model, weights are added depending on the energy consumption in each month during a particular season. Austin has four seasons: Winter (December-February), Spring (March–May), Summer (June–August), and Autumn (September–November). The weights are combined together in such a way that the total equals 1 for a season. The month with the lowest value in a given season receives the maximum weight which is 0.5, while the month with the highest value receives the lowest weight which is 0.2. The major goal of adding weights is to avoid outliers if the data point's value is low and it won't fit into the clusters. The next step is to combine comparable data points using K-Means clustering. The elbow graph, which is represented in Fig2, is used to determine the number of clusters. Because the dataset indicates seasonality and there are four seasons, the number of clusters is set to four. The centroid of the clusters is determined by using the Euclidian approach and datapoints are assigned to the cluster with the shortest distance from the centroid. The ARIMA model is then used to forecast the average energy usage for the next seasons. The ARIMA values for p and q are calculated using the PACF and ACF plots, respectively. Figure 3 depicts the PACF graph, whereas Figure 4 depicts the Auto Correlation Function graph. The value of d is set to 1. d stands for difference. The ARIMA model is used to forecast the average quantity of energy consumption over various seasons in the following years using the values of p,q, and d. Figure 5 shows a graphical representation of the predicted value.

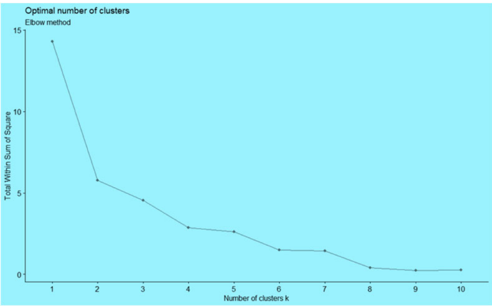

Fig 2. Elbow graph for the dataset

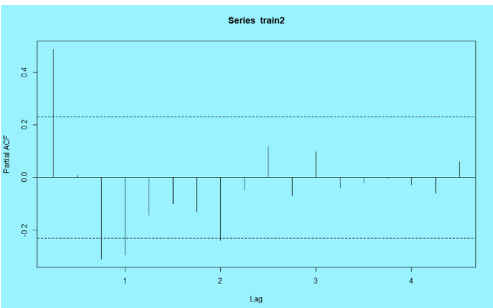

Fig 3. Partial Auto Correlation Function Plot

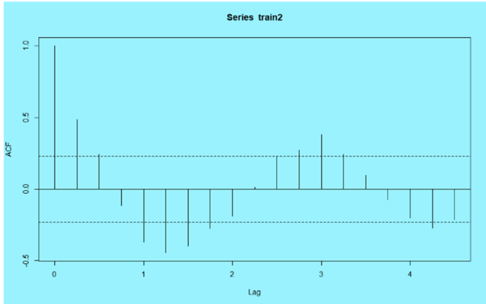

Fig 4. Auto Correlation Function Plot

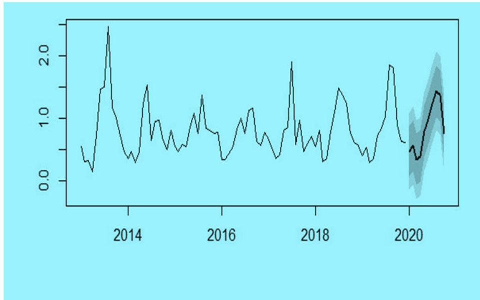

Fig 5. Forecasting using CWKMCA

3.5. Pictorial Representation

In Figure 6, the proposed work is depicted as a conceptual diagram. The suggested model predicts the average quantity of energy that will be consumed in the next seasons.

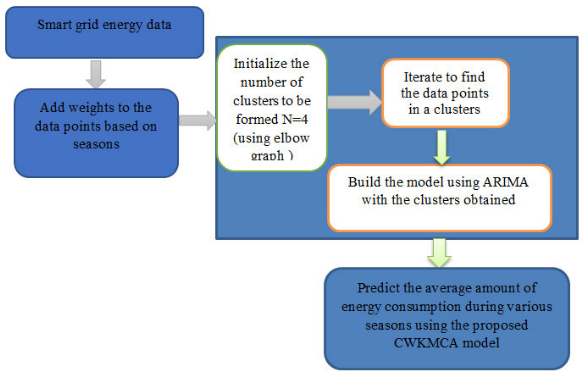

Fig 6. Conceptual Diagram

3.6. Proposed Coherent Weighted K-Means Clustering ARIMA (CWKMCA) algorithm to predict the average amount of energy consumption in a smart grid based on seasonality

Step 1: Read the Variable EnC
Step 2: Calculate the amount of energy consumption for each month in a season
Step 3: Compare the values of energy consumption in each month of a season

Step 4: Give 0.5 weight to the month with the least energy consumption value
Step 5: Give 0.2 weight to the month with the highest energy consumption in a season
Step 6: Give 0.3 to the remaining month in a season
Step 7: Go to Step 3 and perform the same till all the seasons in a year are given weights
Step 8: Multiply the data points with the allotted weights for each month of a season
Step 9: Initialize the number of clusters N = 4 (with the help of Elbow Graph)
Step 10: Create the clusters based on the N value
Step 11: Compute the centroid for each cluster

$$centroid = \frac{d1+d2+\ldots+dn}{n}$$

Step 12: Using K-Means clustering finalize the clusters
Step 13: Build the model with the values obtained in the previous step using ARIMA

ARIMA(p,q,d)

Step 14: Predict the average amount of energy consumption during various seasons using the proposed model
Step 15: Stop

4. Result

ARIMA is a typical model-based forecasting method that takes into account past data and does forecasting based on such data. The K-Means clustering technique, which clusters similar data points, is the most popular and effective clustering technique. The Coherent Weighted K-Means Clustering ARIMA (CWKMCA) method is proposed in this paper to forecast the average amount of energy consumption in a smart house during various seasons. The dataset used in this study was obtained from the Pecan Project in Austin, Texas. The data is collected over seven years, from 2013 to 2019, and is divided into training and test data. This proposed work adheres to the methodology outlined in the previous section. By comparing the suggested model's performance to that of standard ARIMA and Weighted-KMeans Clustering ARIMA, the efficiency of the proposed model is determined. The suggested model's performance is assessed by calculating the values of errors such as RMSE, MAPE, and AIC. Lower values suggest that the model's accuracy is higher. The values of AIC, RMSE, MAPE, and loglikelihood obtained for the dataset using ARIMA, WKMCA, and the suggested CWKMCA models are shown in Table II. When compared to other models, the table clearly illustrates that the proposed model performs better. The MAPE value for ARIMA is 37%, 26% for WKMCA, and 12% for CWKMA, indicating that the suggested model outperforms the others for the dataset. The loglikelihood of ARIMA is -18, it is 25.32 for WKMCA and for CWKMCA it is 72.18. The higher the loglikelihood value, the higher will be the performance of the model. From the result obtained , the loglikelihood value of CWKMCA is higher, indicating that this model performs better. When the RMSE is assessed, the value produced using CWKMCA is lower than ARIMA and WKMCA, implying that the suggested model's accuracy is higher than the other models. The graphical depiction of these values is shown in Figure 7. When comparing the suggested model CWKMCA to ARIMA and WKMCA, the obtained results and representations show that the proposed model CWKMCA performs better for the dataset taken.

Table II
Comparison of the performance of ARIMA and CWKMCA

Parameters	ARIMA	WKMCA	CWKMCA
AIC	49.08	-48.64	-98.18
AICC	50.66	-48.42	-95.64
BIC	61.64	-47.65	-91.05
Loglikelihood	-18.54	25.32	72.18
ME	-0.0093559	-0.0039902	9.23E-04
RMSE	0.3127846	0.0622756	0.0233143
MAE	0.230121	0.0320932	0.0122032
MPE	-16.29267	-6.836557	-21.246465
MAPE	37.18423	25.63582	12.30725
ACF1	0.1625264	0.0164081	0.0105891

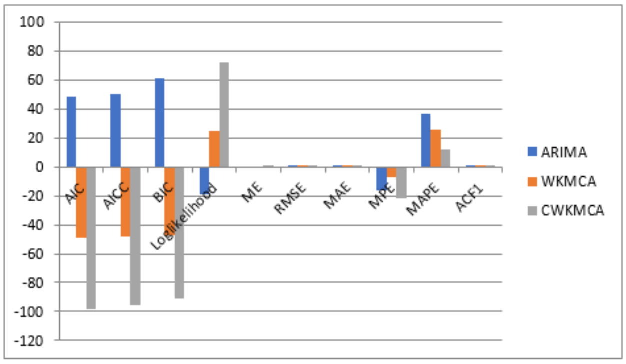

Fig 7. Graphical representation on the comparison of ARIMA, WKMCA and CWKMCA

5. Conclusion

This paper proposes a new technique called Coherent Weighted K-Means Clustering ARIMA (CWKMCA) for predicting the average amount of energy consumed by a smart house from a smart grid in the following seasons. Based on the findings and discussion in the preceding section, it can be stated that the suggested model provides improved accuracy and may be used to forecast time series data with seasonality. According to the results obtained for several parameters such as AIC, AICC, ME, RMSE, and loglikelihood, this model outperforms the ARIMA and Weighted K-Means clustering algorithms. This model predicts how much energy a smart house uses from a smart grid during different seasons. In the future, the amount of energy utilised by each smart device during various seasons can be evaluated, which can aid in the scheduling of smart device usage when prices are not at their highest. Consumers and utilities can both benefit from this forecasting because it lowers prices for consumers and allows utilities to schedule energy supply even during outages.

References

[1] Bruce L. Bowerman, Richard T. O' Connell, & Anne B. Koehler, *"Forecasting, time series, and regression: an applied approach,"* 4th ed. The United States of America: Thomson Brooks, 2005.

[2] Kohiro JM, Otienio RO, Wafula C. *"Seasonal time series forecasting: a comparative study of ARIMA and ANN models"*. Af J Sci Technol. 2004;5(2):41–49.

[3] F. Petitjean , A. Ketterlin , P. Gançarski, *A global averaging method for dynamic time warping, with applications to clustering*, Pattern Recog. 44 (3) (2011) 678–693.

[4] T.W. Liao , *"Clustering of time series data! a survey"*, Pattern Recog. 38 (11) (2005) 1857–1874 .

[5] Pasapitch Chujai, Nittaya Kerdprasop, and Kittisak Kerdprasop, *"Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models"*, Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I, IMECS 2013, Hong Kong, March 13 - 15, 2013.

[6] Cristina Nichiforov, Iulia Stamatescu, Ioana Fagarasan, Grigore Stamatescu, *"Energy Consumption Forecasting Using ARIMA and Neural Network Models"*, 978-1-5386-2059-5/17/$31.00 c 2017 IEEE.

[7] Junwei Miao *"The Energy Consumption Forecasting in China Based on ARIMA Model"*, International Conference on Materials Engineering and Information Technology Applications (MEITA 2015),Published by Atlantis Press,2015.

[8] Suat Ozturk,Feride Ozturk, *"Forecasting Energy Consumption of Turkey by Arima Model"*, DOI: 10.18488/journal.2.2018.82.52.60, 2018.

[9] Sen, Parag & Roy, Mousumi & Pal, Parimal, *"Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization,"* Energy, Elsevier, vol. 116(P1), pages 1031-1038,2016.

[10] Warut Pannakkong, Van-Hai Pham, Van-Nam Huynh, *"A Novel Hybridization of ARIMA, ANNand K-Means for Time Series Forecasting"*, International Journal of Knowledge and Systems Science,Volume 8, Issue 4, October-December 2017.

[11] Grzegorz Dudek**,** *"Next day electric load curve forecasting using k-means clustering"***,**Rynek Energii 92(1):143-149**,** February 2011**.**

[12]Myong-Hoe Huh, Yong B.Lim, *"Weighting variables in K-Means clustering"*,Journal of Applied Statistics, pp. 67– 78, Vol.36, No1, January 2009.

[13] Muhammed Faizan, Megat F. Zuhairi, Shahrinaz Ismail, Sara Sultan, *"Applications of Clustering Techniques in Data Mining: A Comparative Study"*, International Journal of Advanced Computer Science and Applications, Vol.11,No 12, 2020.

[14] Jamal Fattah, Latifa Ezzine, Zineb Aman, Haj El Moussami, Abdeslam Lachhab, "*Forecasting of demand Using ARIMA model*", International Journal of Engineering Business Management, Volume 10:1-9, 2018.

J. Jasmine Christina Magdalene is an aspirant Research Scholar who works as an Assistant Professor in the Department of Computer Applications, Bishop Heber College, Affiliated to Bharathidasan University, Tiruchirappalli, India. She has 14 years of teaching experience and her areas of interests include Data Analytics, Internet of Things, Cloud Computing and Machine Learning.

Dr. B. S. E. Zoraida is an Assistant Professor in the Department of Computer Science, Bharathidasan University, Tiruchirappalli, India. She did her Ph. D at National Institute of Technology, Trichy. She has 18 years of teaching experience and 14 years of profound experience in Research. She has more than 40 research publications to her credit. Her areas of interests include DNA Computing, Bio-inspired Algorithms in Energy Management and Semantic Web.