JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1770-1777
https://publishoa.com
ISSN: 1309-3452

Predicting the Usage of Energy in a Smart Home Using Improved Weighted K-Means Clustering ARIMA Model

J. Jasmine Christina Magdalene
Asst. Professor, Computer Applications, Bishop Heber College, (Affiliated to Bharathidasan University), Tiruchirappalli, India
jjasminebhc@gmail.com

B. S. E. Zoraida
Asst. Professor, Computer Science and Engineering, Bharathidasan University, Tiruchirappalli, India
b.s.e.zoraida@gmail.com

Abstract- Modern development in making the whole world to act smartly has not left the home. The homes in modern era are becoming smart. The household appliances irrespective of their sizes are designed to behave smartly according to the needs of the human beings. By doing so, it helps the human beings in multiple ways. One of the ways to enjoy the benefits in a smart home is the efficiency to manage the electric energy. When electric energy is managed in a proper way, it helps the people in saving the amount that is spent on using electrical appliances and it also helps them to save the energy so that during the time of natural calamities there is no shortage of energy. To achieve this prediction of energy consumption plays a very important role. Based on the prediction done, the smart devices can be scheduled to operate so that pricing at peak hours can be reduced. Energy storage and production from renewable resources can be handled in a better way based on this forecasting of energy consumption during various seasons in a year. In this proposed new model Improved Weighted K-Means Clustering ARIMA (IWKMCA), prediction on energy consumption in a smart home is done based on the requirements during the various seasons in a year. This proposed model enhances the weighted k-means clustering algorithm to form clusters with better cluster points. Weights are added to the data points that belong to a particular cluster and based on the centroid of the clusters that are formed forecasting is done using the ARIMA model. This is done to increase the accuracy in forecasting and to reduce the various forecasting errors. The average amount of energy consumed from the smart grid by a smart home in the Pecan Project in Texas, USA is taken as the dataset for this proposed work and the average amount of energy consumed during various seasons by smart devices in the smart home is forecasted. This model shows better accuracy level when compared with the traditional ARIMA model and the Weighted K-Means Clustering ARIMA Model. The result of this proposed work shows lesser values than the ARIMA and Weighted K-Means Clustering ARIMA model when the forecasting errors like RMSE, MAPE, AIC, AICC are considered. This proposed work also shows a higher loglikelihood value than the ARIMA model proving that this model excels the standard ARIMA model in all aspects for time series forecasting.

Keywords- Smart Home, Energy Management, Weighted K-Means Clustering, ARIMA, RMSE, AIC, AICC.

INTRODUCTION

This era is a technological era and smart is the buzz word all over the world. Everybody is getting attracted to the smart devices and appliances and the home is also becoming smart. A smart home is a home where most of the appliances are smart and electric grid is not an exception. The smart grid helps the consumers to know how much of energy is consumed by which device so that the consumers can schedule their devices to reduce pricing. Hence managing the energy has gained importance as it not only helps the consumers but it also helps the utilities to manage energy so that they can provide uninterrupted supply of energy to the consumers even during the time of outages. To manage the energy efficiently prediction of energy consumption in a smart home is crucial. Prediction helps the consumers and the utilities to know more about the requirement of energy by each home during each season. To predict the consumption of energy the historical data will be of use and hence a model-based forecasting method has to be used. The energy consumption data is said to be a time series data as it is sequential and recorded in regular interval of time. There are four major components in a time series data. They are trend, seasonality, cyclical and randomness [1].

In the recent years research in the area of time series forecasting has gained more importance and interest as it helps various organizations and businesses to plan for future. There are a number of statistical and machine learning tools to perform prediction. Regression techniques, decomposition models, ARIMA are some of the most famous forecasting models and each model differs in their accuracy level. The accuracy of a model is identified by knowing the error rate of the predicted value. The lower the error the higher is the accuracy of the model. Recent research shows that one of the popular models for time series forecasting is ARIMA [2]. One of the components of time series data is randomness and because of this randomness in data many data mining techniques like classification, clustering and indexing have gained importance [3]. The unforeseen patterns in a time series data can be identified by using the clustering technique. Clustering of similar data points helps in the non-interference of other data point which doesn't have the similarity to that of a particular cluster [4].

In this proposed work the Weighted K-means Clustering algorithm is modified to enhance the efficiency of the algorithm and the ARIMA model is used after that to predict the average amount of energy consumption during various seasons in a smart home. The organization of this research work is as follows. The review of literature is dealt in Section II. The methodology of the proposed work is given is Section III followed by the result and discussion in Section IV. The conclusion and future work are given in Section V.

REVIEW OF LITERATURE

One of the sought-after areas in the field of scientific research is the prediction of energy management in a smart city and smart home. Many researchers and authors are making significant contributions in this research area by giving various different ideas and thoughts. Jamal Fattah et.al have modelled and forecasted the demand of finished food product to help in managing the supply chain. In their work they have concluded that ARIMA model can be used for time series data. [5]. Cristina Nichiforov et.al have compared ARIMA and Non-linear Auto Regressive neural network (NAR) models and concluded that ARIMA is better than NAR for the forecasting of energy consumption [6]. Farjana Mahia et.al have used ARIMA model t to forecast the electricity consumption for various industries in China. They have proved that the predictions were stable and of high precision [7]. Suat Ozturk et.al, have forecasted energy consumption of coal, natural gas and oil in Turkey using ARIMA [8]. Cristina Nichiforov et al have compared the accuracy of forecasting of energy consumption using ARIMA and Neural Network Models. They have concluded that the accuracy of ARIMA is better than the Neural Network model when it comes for forecasting of energy consumption [9]. Warut Pannakkong et al., have given an innovative 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]. Haihong Bian et.al., have proposed a study on the forecasting of short-term power load using k-means and non-clustering based algorithms. They have concluded that cluster-based k-means algorithm gives better accuracy than the non-cluster based algorithms [11]. Wei Yang et.al., have done a research on adding weights to the traditional k-means clustering and proved that there is enhancement in the performance of the traditional algorithm after adding weights. [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.

METHODOLOGY

When we consider the average amount of energy consumption in a smart home during various seasons, it greatly relies on the appliances that are used. The amount of energy consumption during summer will be high as air conditioners are used and when winter is considered heaters are used and there will be a moderate usage of energy during spring. As the amount of energy consumption by each device varies, the amount of energy used from the electric grid also varies. Some houses may have renewable resources too and hence the energy drawn from the grid changes. In general, the average amount of energy consumption in a smart home varies according to different seasons. Since there are changes in the average amount of energy consumption during various seasons prediction of energy consumption becomes a necessity. If the average energy consumption during various seasons is predicted, it will help the utilities to plan the supply of energy and to help the consumers at the time of natural disasters and during other outages. It not only helps the utilities; it also helps the consumers to know about their usage and schedule it in such a way so that they can avoid higher pricing. This prediction can be done using the historical data. As historical data is considered standard data-driven forecasting model like Holt Winter's cannot give the expected accuracy, hence model-based forecasting method has to be considered. Model-based forecasting method will enhance the accuracy of the forecasting model. In this proposed work the data point is given weights according to the season in the Weighted K-Means clustering algorithm and the weighted data are clustered and then ARIMA which is a model-based forecasting technique is used to predict the average amount of energy consumed during various seasons in a smart home.

This work is implemented using the data that is taken from the Pecan Project, Austin USA for a single home. The average energy consumption from the smart grid is taken for the years 2013 – 2019 (seven years) and it is split into

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1770-1777
https://publishoa.com
ISSN: 1309-3452

training and test data. This dataset is used to forecast the average amount of energy consumption during various seasons in the forth coming years. The average amount of energy consumption from the smart grid during each month is taken as the input and f(EC) is the function used to represent this input variable. With this input clusters are formed using K-Means clustering algorithm based on the centroid that is calculated. Once the cluster is formed, weights are added to each data point in the cluster to which they belong. Random weights are given to each data point in a particular cluster so as the total weights add up to 1 for a cluster. The random weights that are generated are multiplied with each data point. Then the ARIMA model is used to predict the average amount of energy consumption from the smart grid during various seasons of the forthcoming years. The accuracy of this model Improved Weighted K-Means Clustering ARIMA (IWKMCA) is assessed by comparing the RMSE, MAPE and loglikelihood with the standard ARIMA and Weighted K-Means ARIMA models. R tool is the software that is used to implement this work.

A. Input Parameter:

The average amount of energy consumption in a smart home depends mostly on the appliances that are used in different seasons and season also plays a major role in the amount of energy usage. To carry out this work, the average amount of energy consumption in a month from the smart grid is taken as the input parameter. (i) Energy consumption from a smart grid during each month

1) Average Energy Consumption from a Smart Grid:

A smart home has many energy sources like the electrical grid, solar, windmill etc., though there are many sources most of the smart homes still rely on the electrical grid. The energy consumption by each appliance in a smart home varies and its usage also varies based on the various seasons. As the energy consumption by devices varies, it has an impact on the energy drawn from the grid also. Hence there are always fluctuations in the amount of energy consumed during various seasons.

B. Data Preprocessing

The first step to build the forecasting model is to preprocess the data. The data that is taken for this work is from a smart home in Pecan Project, Texas, USA. This data is a time series data as the data is collected every minute. The average amount of energy consumed in each month is taken for a period of seven years from 2013 to 2019. This data that is collected for every minute has some null values and they are filled with the previous value as a process of preprocessing. Then the data is converted into a monthly data. The average amount of energy consumption in kw during various months for the period of seven years from 2013 to 2019 is given in Table 1.

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

C. Construct and Decompose the Timeseries Data

The dataset that is used in this work is a time series data as it is relative to time. This data is converted into a monthly time series data by given the frequency as twelve. This time series data is decomposed so as to check whether the data is cyclical and shows trend, seasonality and also to find the randomness of the data. The decomposition is done to choose the best model for predicting the dataset taken for this work. The trend, cyclical, seasonality and randomness of the dataset is shown in Fig1.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1770-1777
https://publishoa.com
ISSN: 1309-3452

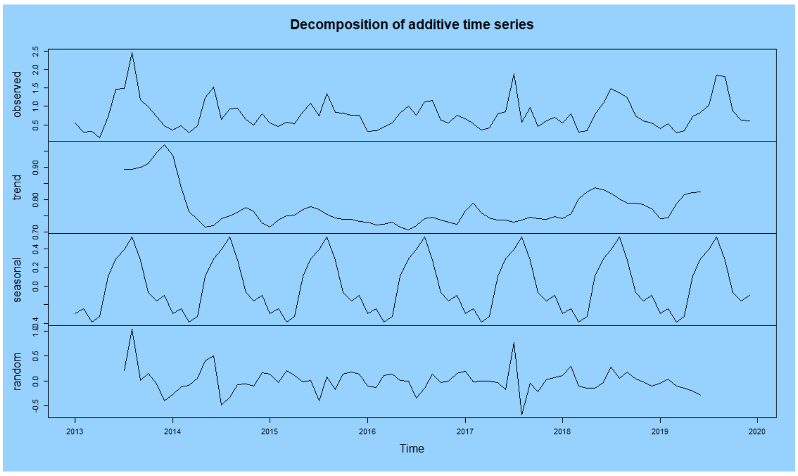

Fig 1: Decomposition of time series data

D. Construct the Model

The proposed model is constructed by forming clusters using k-means clustering to combine similar data points as an initial step. The elbow graph is used to find the number of clusters formed and it is depicted in the graph shown in Fig2. Here the number of clusters is taken as 4 as the dataset shows seasonality and the number of seasons are 4. The Euclidian method is used to find the centroid of the clusters and the data points are moved to the respective cluster based on the shortest distance from the centroid. Then weights are added to each data point in the cluster . In weighted K-Means clustering, random weights are added but, in this model, weights are added in such a way that it all adds up to 1 for a cluster. The main idea for adding weights is to avoid outliers if the value of the data point is less and it may not be included in the clusters. Then the ARIMA model is used to predict the average amount of energy consumption in the upcoming seasons. The values for p and q in the ARIMA are obtained from the PACF and ACF plots respectively. The graph for PACF is shown in Fig 3 and the Auto Correlation Function graph is shown in Fig 4. The value for d is taken as 1. d indicates the difference. By making use of the values of p,q and d the ARIMA model is used to forecast the average amount of energy consumption during various seasons in the upcoming years. The graphical representation of the predicted value is shown in Fig 5.

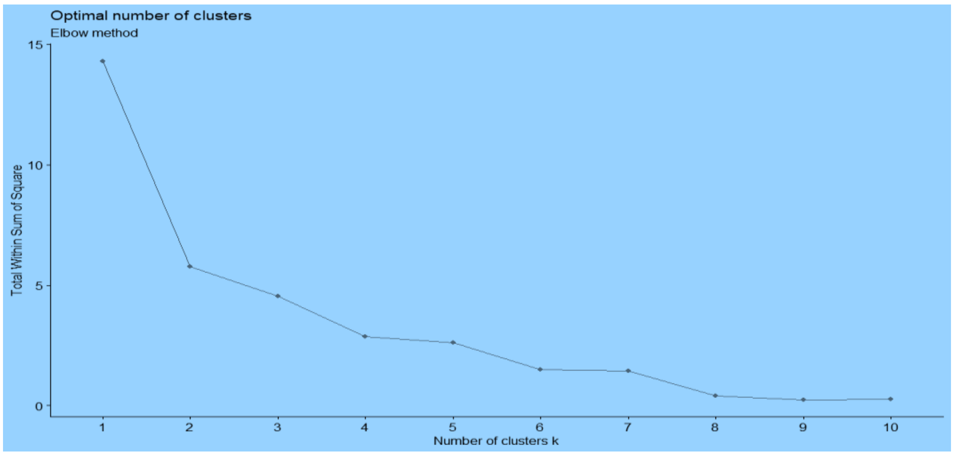

Fig 2. Elbow graph for the dataset

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1770-1777
https://publishoa.com
ISSN: 1309-3452

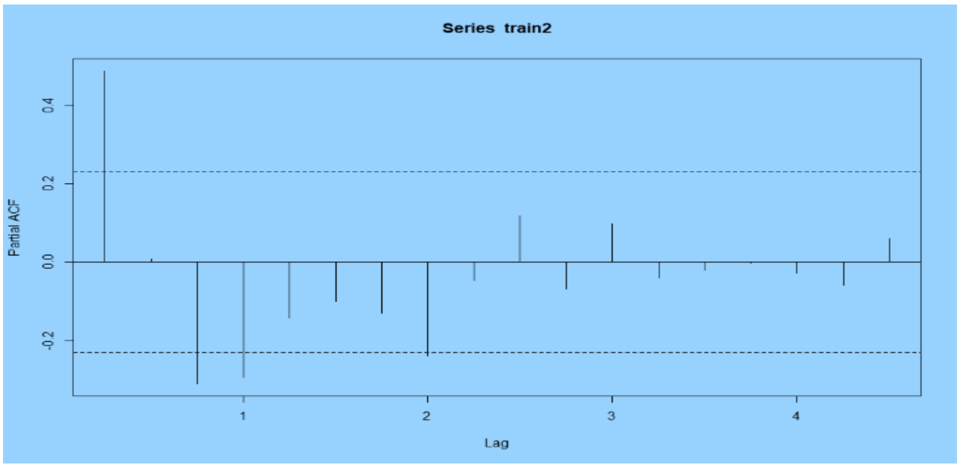

Fig 3. Partial Auto Correlation Function Plot

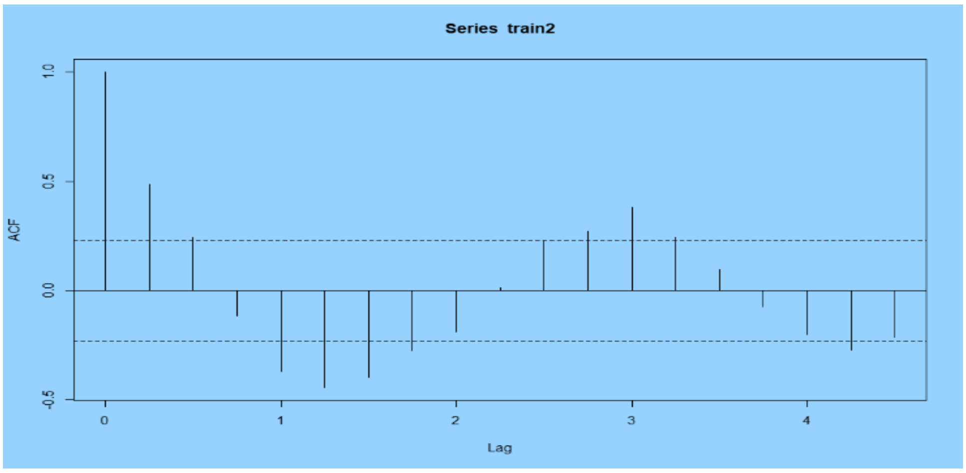

Fig 4. Auto Correlation Function Plot

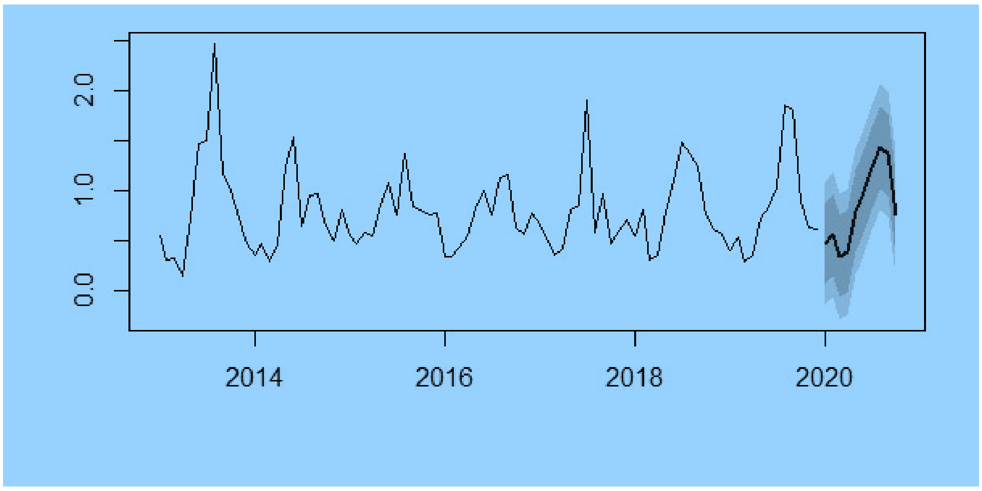

Fig 5. Forecasting using IWKMCA

E. Pictorial Representation

The proposed work is pictorially represented as a conceptual diagram in Fig 6. The average amount of energy that will be consumed in the forthcoming seasons is predicted using the proposed model.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1770-1777
https://publishoa.com
ISSN: 1309-3452

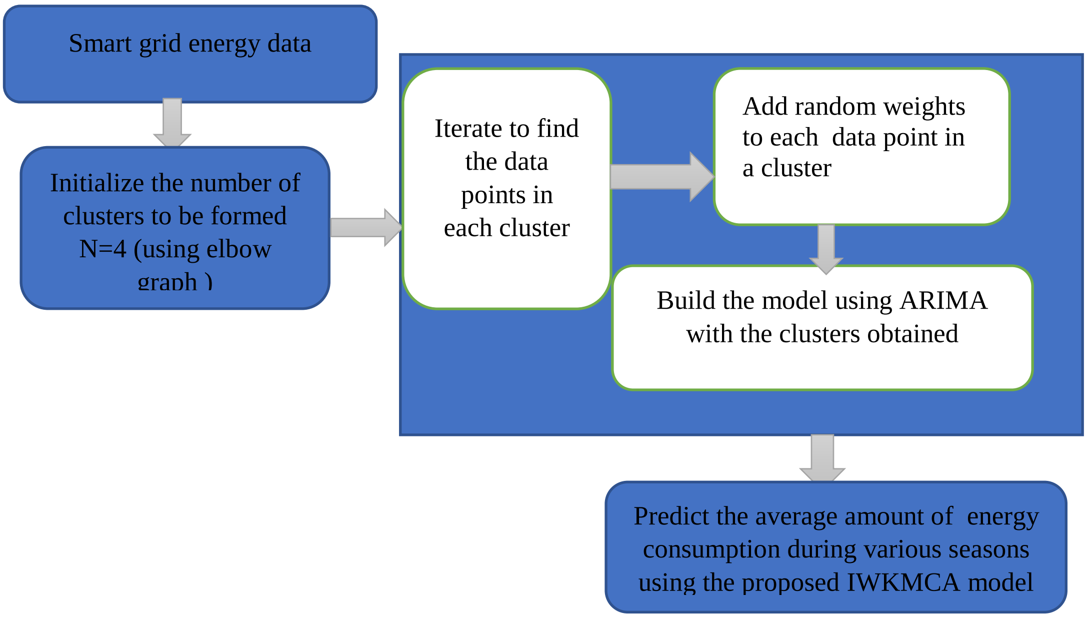

Fig 6. Conceptual Diagram

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

---

Step 1: Read the Variable EC
Step 2: Initialize the number of clusters N = 4 (with the help of Elbow Graph)
Step 3: Create the clusters based on the N value
Step 4: Compute the centroid for each cluster

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

Step 5: Using K-Means clustering finalize the clusters
Step 6: Add random weights to the data points of each cluster
Step 7: Check whether the weights add up to 1 else goto Step 6
Step 8: Multiply the data points with the allotted weights for each data point in a cluster
Step 9: Build the model with the values obtained in the previous step using ARIMA
ARIMA(p,q,d)
Step 10: Predict the average amount of energy consumption during various seasons using the proposed model
Step 11: Stop

RESUTS AND DISCUSSION

One of the standard model-based forecasting methods is ARIMA which considers historical data and based on the historical data forecasting is done. The most popular and effective clustering technique is the K-Means clustering which clusters similar data points. This work Improved Weighted K-Means Clustering ARIMA(IWKMCA) is proposed to predict the average amount of energy consumption during various seasons in a smart home. The dataset that is used to carry out this work is taken from Pecan Project, Austin, Texas. The data for seven years i.e., from 2013 to 2019 is taken and it is split into training data and test data. This work follows the methodology that is discussed in the previous section. The efficiency of the proposed model is identified by comparing it with the performance of the standard

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 1770-1777
https://publishoa.com
ISSN: 1309-3452

ARIMA and Weighted-KMeans Clustering ARIMA. The performance of the proposed model is measured by taking the values of the errors like RMSE, MAPE and AIC. Lower values indicate that the accuracy of the model is higher. The following Table II. shows the values of AIC, RMSE, MAPE, loglikelihood which are obtained for the dataset using ARIMA, WKMCA and the proposed IWKMCA models. The table clearly shows that the performance of the proposed model is better when compared with other models. The MAPE value for ARIMA is 37%, it is 26% for WKMCA and 17% for IWKMA proving that the proposed model performs better for the dataset taken. The log likelihood value of IWKMCA is higher when compared to the other two models showing this model shows better performance. When the RMSE is considered, the value that is obtained using IWKMCA is less when compared to ARIMA and WKMCA stating that the accuracy is higher in the proposed model than the other models. Fig 7. depicts the graphical representation of these values. From the result that is obtained and by observing the representations, the proposed model IWKMCA shows better performance for the dataset taken when compared to ARIMA and WKMCA.

TABLE II
COMPARISON OF THE PERFORMANCE OF ARIMA, WKMCA AND IWKMCA

Parameters	ARIMA	WKMCA	IWKMCA
AIC	49.08	-48.64	-94.18
AICC	50.66	-48.42	-93.75
BIC	61.64	-47.65	-87.95
Loglikelihood	-18.54	25.32	50.09
ME	-0.0093559	-0.0039902	7.84E-04
RMSE	0.3127846	0.0622756	0.0334053
MAE	0.230121	0.0320932	0.0233021
MPE	-16.29267	-6.836557	-18.357576
MAPE	37.18423	25.63582	17.41836
ACF1	0.1625264	0.0164081	0.01169

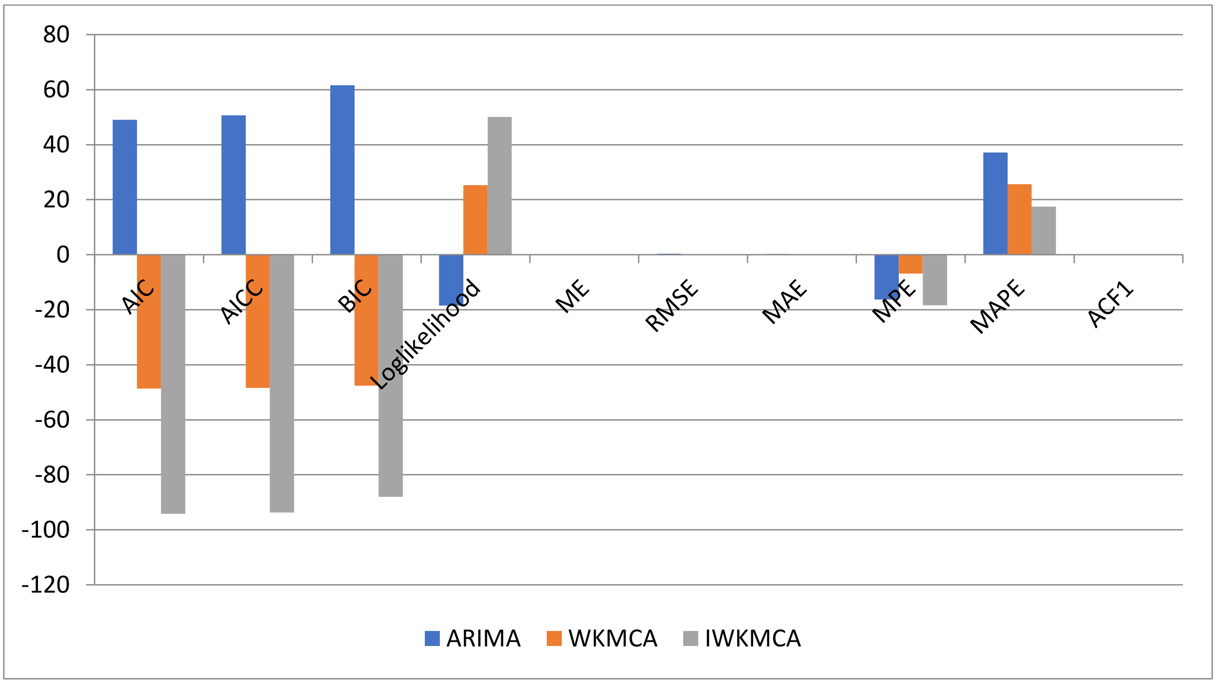

Fig 7. Graphical representation on the comparison of ARIMA, WKMCA and IWKMCA

CONCLUSION

A new algorithm Improved Weighted K-Means Clustering ARIMA (EWKMCA) is proposed in this work to predict the average amount of energy that will be consumed by a smart home from a smart grid in the upcoming seasons. Based on the result that is obtained and discussed in the previous section, it can be concluded that the proposed model gives better accuracy and can be implemented for forecasting time series data which shows seasonality. The performance of this model is better when compared with the ARIMA and Weighted K-Means clustering algorithm as per the values obtained for the various parameters like AIC, AICC, ME, RMSE and loglikelihood. This model forecasts the energy consumed from a smart grid in a smart home during various seasons. In future, energy consumed by each smart device during various seasons can be considered which can help in scheduling the usage of the smart devices when the price is not in its peak. The consumers and the utilities can make use of this forecasting so that the price is reduced for the consumers and the utilities can schedule the supply of energy even at the time of outages.

REFERENCES

1. Yathish H R,, S.R.Swamy, “Recent Trends in Time Series Forecasting- A Survey”,International Research Journal of Engineering and Technology(IRJET), vol 7, issue 4,April 2020.
2. Hussan Al-Chalabi, Yamur K. Al-Douri and Jan Lundberg.” Time Series Forecasting Using ARIMA Model- A Case Study of Mining Face Drilling Rig”, ADVCOMP 2018:The Twelfth International Conference on Advanced Engineering Computing and Applications in Sciences, IARIA 2018.
3. Xu Lu and Tianzhong Zhao,”Research on Time Series Data Prediction based on clustering algorithm- A Case Study of Yuebao”, AIP Conference Proceedings, August 2017.
4. 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, Volume 11,No12,2020 .
5. Jamal Fattah, Latifa Ezzine, Zineb Aman, Haj El Moussami and Abdeslam Lachhab,” Forecasting of Demand Using ARIMa model”, International Journal of Engineering Business Management,Volume 10:1-9.
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. Frajana Mahia, Arpita Rani Dey, Md Abdul Masud, Mohammed Sultan Mahmud, “Forecasting Electricity Consumption using ARIMA Model”, 2019 International Conference on Sustainable Technology for Industry 4.0(STI),Published by IEEE 2020.
8. Suat Ozturk,Feride Ozturk, “Forecasting Energy Consumption of Turkey by Arima Model”, DOI: 10.18488/journal.2.2018.82.52.60, 2018.
9. Cristina Nichiforov, Iulia Stamatescu, Ioana Fagarasan, GrigoreSta, "Energy Consumption Forecasting Using ARIMA and Neural Network Models, 2017 5th International Symposium on Electrical and Electronics Engineering, Publisher:IEEE, December 2017.
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. Haihong Bian,Yigun Zhong, Jianshuo Sun**,** Fangchu Shi,**”**Study on Power Consumption Load Forecast based on K-Meand Clustering and FCM-BP Model**”,** Energy Reports, Volume 6, Supplement 9,2020, Pages 693-700.
12. Wei Yang , Hua Long, Lihua Ma, Huifang Sun, “Research on Clustering Method Based on Weighted Distance Density and K-Means”, Procedia Computer Science, Volume 166, 2020, Pages 507-511.