Identification and Performance Evaluation for Effective Utilization of Electrical Energy Resource using K-Means Clustering Algorithm

M.Anita Priscilla Mary [1], Dr.M.S.Josephine [2], Dr.V.Jeyabalaraja [3], Dr.S.Nirmala Sugirtha Rajini [4]
[1] *Research Scholar,* [2, 4] *Professor, Department of Computer Applications,*
[3] *Professor, Department of Computer Science & Engineering,*
[1, 2, 4] *Dr.M.G.R Educational and Research Institute, Maduravoyal, Chennai-600095,*
[3] *Vellammal Engineering College, Surapet, Chennai-600066.*

Abstract

Data Mining is used to distinguish the interesting relations between different variables in the databases. Data Mining is used to unpack unrevealed patterns in the data. Clustering is a method used to identify group of objects that are like each other, which helps to find the differences and similarities between the data. Clustering is applied to various sectors of energy resources by using K-Means algorithm. Efficiency of algorithm is found by modifying existing K-Means algorithm applying Python.

Keywords: *Efficiency, Utilization, Identification.*

1. Introduction

Electrical Energy conservation is an efficient method of using energy consumption. Energy cost may be reduced by adapting the effective method in energy usage[1]. Fourteen years of statistical data is taken for analysis of energy consumption to find which sector needs effective utilization of energy. K-Means clustering process is applied to find which sector needs effective utilization. Advantages and disadvantages of K-Means methodology are identified in order to enhance the existing K-Means algorithm[2]. Accuracy is identified which is compared with actual K-Means algorithm and Proposed Altered K-Means algorithm. Effective utilization of power demand is analyzed by identifying the various sectors of energy consumption[3]. The various energy sectors are Domestic, Commercial, Industries, Agriculture and others (Public lighting, water works, cottage industries, traction (railways), Bulk supplies to licenses and miscellaneous sales). The areas of energy consumption are identified and methods devoir to be ensues in order to have effective utilization. Implementation of data resource is done using Python.

2. Related work

Tianke sun, Tieyen Zhang, Yun Teng(2019) has described about monthly electricity consumption based on X12 and STL decomposition model for integrated energy system using seasonal and trend decomposition model [4].

Vaishali.S, Mrs. Rajni N. Pamnani(2017) in their paper they have proposed a model for effective prediction on electricity consumption using efficient analysis of house hold characteristics. It is implemented using analysis of load profile and load estimation using support vector machine [5].

Navjit Kaur, Amrit Kaur(2016) has recommended a predictive model for forecasting electricity utilization using value forecasting is implemented using Artificial Neural Network [6].

3. Mode and Mechanism

ISSN: 2005-4238 IJAST

3.1 Dataset

The data needed for clustering means need to store continuously accessed data need to be close together so that accessing can be done in few operations. Fourteen years of statistical data is collected of various Energy Consumption sectors such as Domestic, Commercial, Industries, Agriculture and others [7]. The converted into tabular data where each column of the table represent a different value [8]. It is categorized into columns such as Division and Data Values in order to apply clustering and find which sector needs effective utilization for energy consumption.

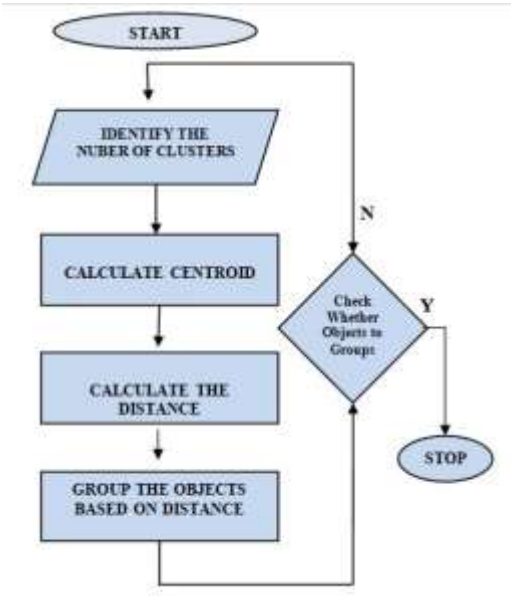
3.2 Mode

Python is a persuasive programming language which has high aligned data architecture and has effectual accession to object oriented programming. Python has built in support for task automation. Pair programming facility and more contribution to open source makes Python as best programming language [9]. In Python Jupyter notebook is used since it has expressive code feature, it is used as a server to create web applications. It can handle huge data and enables to perform complex mathematics. It can work on different platform and uses simple English language. Python supports the usage of apportionment and assortment which help the coding to be delineating in a modular style and code can be reused across various projects.

3.3 Mechanism

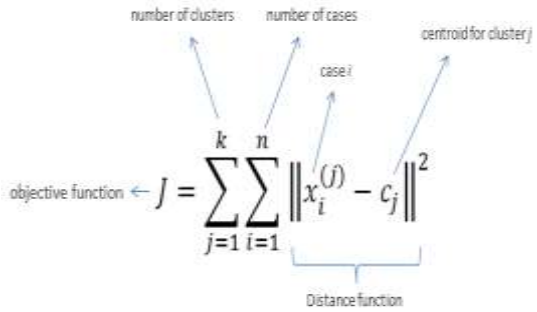
K-means clustering technique is sole of the exceedingly used clustering algorithms because of its homogeneity. K-means algorithm is a consolidated maneuver which partitions the dataset into k distinct subgroups called clusters, where each data point belongs to only one cluster [10]. It keeps data points of similar into one faction and dissimilar data points into another faction.

3.3.1. Flowchart of K-means



3.3.2. Working of Actual K-means Algorithm

Step 1 : First initialize the k points, called means inconstantly.
Step 2 : Typecast each item to its neighboring mean and update the mean's coordinate which are the intermediate of the items categorized in that mean so far.
Step 3 : Repeat the procedure for given cipher of monotony and at the end clusters are formed



$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

3.3.3. Advantages of K-means

- Relatively simple to implement.
- It can scale to large data sets.
- Guarantees convergence.

3.3.4. Disadvantages of K-means

- If a centroid is initialized to be far away point it may not be initialized to any cluster.
- It may have more clusters linked for a single centroid.

3.3.5. Proposed Altered K-means Clustering Algorithm

Step 1 : Digitize the centroid randomly from the given predetermined data points.

Step 2 : Compute the distance for each new data points from the cluster centers and find the minimum of those distances.

Step 3 : Choosing a data point as next centroid from the given data points is correlated with the distance from the adjacent and antecedent selected centroid.
a) Check whether minimum distance is beneath than they are clustered. Our model is relevant than the quondam model as it will cluster to the exact centroid, Inception value. If the predicament is true then assign the new data point to the proportional cluster. If the predicament is not satisfied assign it as new cluster with new data point.

Step 4 : We can repeat all the steps until all data points.

4. Experimental Result

4.1. Comparison of Actual K-means Algorithm and Proposed Altered K-means algorithm

Table 1: Performance comparison

Measures	Existing K-Means	Proposed Altered K-Means
Accuracy Score	0.833	0.98
Iterations	3	4
Inertia	143.3345	89.6384
Elapsed Time	-0.00010	-0.000333

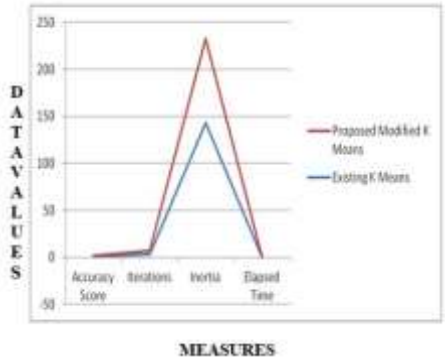


Figure 1: Performance Evaluation

When we see the performance comparison of accuracy score in Actual K-means algorithm and Proposed K-means it is found that accuracy score is high in Proposed Altered K-means algorithm [11]. This proves that the Proposed Altered K-means has solved the disadvantage of Existing K-means algorithm of clustering more clusters to the same centroid [12]. The iteration is higher in Proposed Altered K-Means algorithm which proves it takes more one iteration to match each data set to the correct centroid [13, 14].

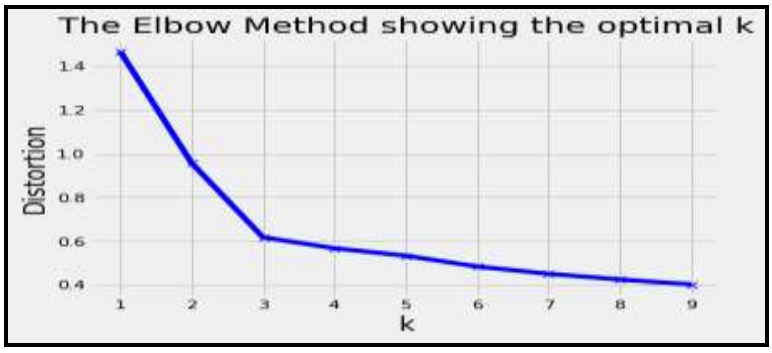
4.2 Identifying Clusters using Cluster Distance

Table 2: Cluster Distance for Five Objects

Object	C1-Distance	C2-Distance	C3-Distance	Cluster
Object 1	3.482	1.501	3.883	C2
Object 2	2.008	1.308	1.539	C2
Object 3	3.049	1.245	3.661	C2
Object 4	2.225	1.906	1.128	C3
Object 5	3.245	5.095	3.990	C1

For the given data sets it has created three clusters by applying proposed Altered K-means algorithm based on cluster distance the data points are clustered to C1, C2 and C3.

Figure 2: Visualization of Optimal K using Elbow Method using Python



The Elbow Method is a definite, highest typical method to determine this optimal value of k and for each value, we are calculating the mass of squared distances from each point to its assigned center

4.3 Cluster Representation

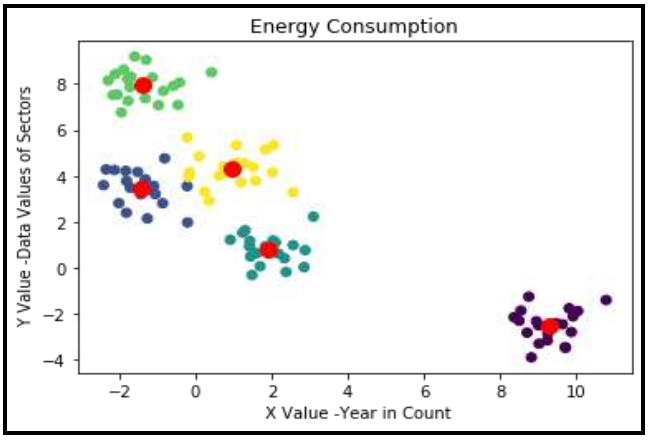


Figure 3: Visualization of Clusters for Energy Consumption of Various Sectors using Python

In energy consumption five different sectors of Energy Consumption data were given to form clusters. Five clusters are formed with centre points. Red point represents the cluster centres.

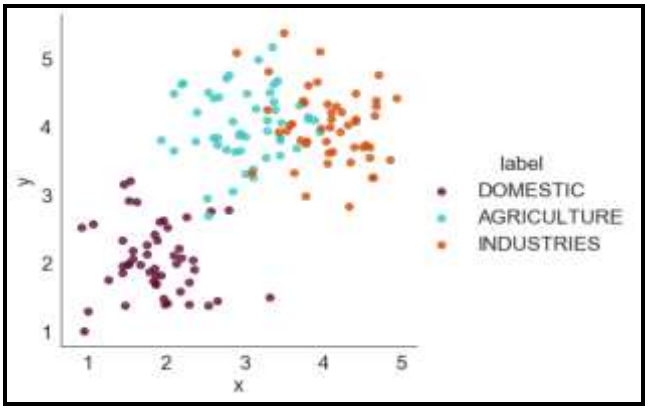


Figure 4: Identifying Clusters using Labels using Python

After applying Proposed Altered K-Means clustering it is identified that Domestic, Agriculture and Industries are in high level of Energy Consumption in different categories of Energy sectors.

5. Results and Discussion

After implementing K-means clustering various sectors of electricity usage it is found that effective utilization is mainly needed in the areas namely Domestic, Agriculture and Industry. Effective utilization in Domestic can be done by following the methods like replacing light bulbs by LEDs which reduces 25-80% of electricity. Purchasing of Energy Efficient appliances can pauperize energy consumption in Domestic sectors. In Industry effective utilization can be done shifting of timing of their operations away from peak demand hours to off peak time. Cogeneration of powers should be initiated in all industries. In Agriculture effective utilization can be done by using drip irrigation for specific crops which will consume up to 80% water and reduce pumping energy requirement, irrigation of crops can be done during night time.

ISSN: 2005-4238 IJAST

6. Conclusion

Existing K-Means algorithm is improvised by Proposed Altered K-Means algorithm, after applying the mutated algorithm to the data accuracy level was high compared to Actual K-Means method. This algorithm was mainly applied to find out where the effective utilization is mainly required in various sectors of Energy Consumption. After applying the algorithm the sectors are identified namely Domestic, Agriculture and Industries. Hence several methods have to be followed to reduce the Energy Consumption in the following sectors to sequence the tourney of future demand.

References

[1] M.Anita Priscilla Mary**,** M.S.Josephine, "Analysis and Forecasting Of Electrical Energy a Literature
Review", International Journal of Pure and Applied Mathematics, Volume 119 No. 15 2018, 289- 293.

[2] Data Mining: Concepts And Techniques 3rd Edition Authors: Jiawei Han Micheline Kamber Jian PeiHardcover ISBN: 9780123814791 eBook ISBN: 9780123814807Imprint: Morgan Kaufmann Published Date: 22nd June 2011Page Count: 744.

[3] Jain Amit, B. Satish, "Clustering based short term load forecasting using support vector machines", *roceedings of 2009 IEEE Bucharest PowerTech*, 2009.

[4] Tianhe Sun, Tieyan Zhang, Yun Teng, Zhe Chen, and Jiakun Fang 2019, 'MONTHLY ELECTRICITY CONSUMPTION FORECASTING METHOD BASED ON X12 AND STL DECOMPOSITION MODEL IN AN INTEGRATED ENERGY SYSTEM' Hindawi Mathematical Problems in Engineering Volume 2019, Article ID 9012543, 16 pages.

[5] Vaishali.S, Mrs. Rajni N. Pamnani, 2017 'Effective Prediction of Electricity Consumption based on Efficient Analysis of Household Characteristic ',IJCSMC, Vol. 6, Issue. 5, May 2017, pg.323 – 328.

[6] N. Kaur and A. Kaur, *2016* 'Predictive modelling approach to data mining for forecasting electricity consumption' , *6th International Conference - Cloud System and Big Data Engineering (Confluence)*, Noida, 2016, pp. 331-336.

[7] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, and M. E. Webber, "Clustering analysis of residential electricity demand profiles," Appl. Energy, vol. 135, pp. 461–471, Dec. 2014.

[8] I. Benítez, A. Quijano, J.-L. Díez, and I. Delgado, "Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers," Int. J. Electr. Power Energy Syst., vol. 55, pp. 437–448, Feb. 2014.

[9] V. R. Niveditha et. al, "Detect and Classify Zero Day Malware Efficiently In Big Data Platform", International Journal of Advanced Science and Technology, Vol. 29, No. 4s, **(2020)**, pp. 1947-1954.

[10] Natrayan, L., Yogeshwaran, S., Yuvaraj, L., & Kumar, M. S. Effect of graphene reinforcement on mechanical and microstructure behavior of AA8030/graphene composites fabricated by stir casting technique. AIP Conference Proceedings, Vol. 2166 (1), **(2019)**, pp. 020012.

[11] D. Usha, Niveditha VR, Sharadha Priya K., "Anomaly Detection on Android System", International Journal of Innovative Technology and Exploring Engineering, 9(1), (2019) , pp. 4310-4313.

[12] Unnati R. Raval , Chaita Jani ,2016 'Implementing & Improvisation of K-means Clustering Algorithm', IJCSMC, Vol. 5, Issue. 5, May 2016, pg.191 – 203.

[13] Fahim, A.M., Salem, A.M., Torkey, ,2006 F.A. et al. 'An efficient enhanced k-means clustering algorithm'. J. Zhejiang Univ. - Sci. A 7, 1626–1633.

[14] Josephine, M. S., and V. Jeyabalaraja. "Expert System and Knowledge Management for Software Developer in Software Companies." *International Journal of Information and Communication Technology Research* 3.2 (2012): 243.
[15] S. Velliangiri, P. Karthikeyan & V. Vinoth Kumar (2020) Detection of distributed denial of service attack in cloud computing using the optimization-based deep networks, *Journal of Experimental & Theoretical Artificial Intelligence*, DOI: 10.1080/0952813X.2020.1744196
[17] Jayasuruthi L,Shalini A,Vinoth Kumar V.,(2018) ” Application of rough set theory in data mining market analysis using rough sets data explorer” Journal of Computational and Theoretical Nanoscience, 15(6-7), pp. 2126-2130