International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(13s) 2026 pp. 53-59
https://doi.org/10.70917/ijcisim-2026-4033
Article

Unsupervised Learning-Based Behavioural Analysis of IoT-Enabled Air Conditioner Systems Using Environmental Sensing

Shubhangi Mathe[1], Pooja Kulkarni[2]

[1] Department of Computer Science, Vishwakarma University, Kondhwa, Pune, Maharashtra, India.
Email: shubhangi.mathe-438@vupune.ac.in
[2] Department of Computer Science, Vishwakarma University, Kondhwa, Pune, Maharashtra, India.
Email: pooja.kulkarni@vupune.ac.in

Abstract: The growing demand for energy-efficient smart indoor environments demands the development of intelligent systems that is capable of real-time monitoring, behavioural analysis, along with autonomous decision-making. This study comprises of the ESP32 based IoT model that supports intelligent monitoring and autonomous decision - making smart indoor system. The environmental data collected using various sensors. This data was logged on to the cloud. To identify the implicit operating behavioural pattern of the data, K-means clustering algorithms along with Elbow method was utilised. For graphical pattern visualization and interpretation, Cluster centroid heatmaps and PCA (Principal Component Analysis) were used. This system enables the transition towards an adaptive and intelligent control system. The smart system is comprised of environmental sensors that results into real time data acquisition and data logging on cloud using Firebase. This research study, helped to monitor the pattern behaviour of the data which is useful for the transition of unsupervised data into supervised data.

Keywords: PCA (Principal Component Analysis), K-Means, Elbow method, IoT, Behavioural pattern, Unsupervised learning..

1. Introduction

This study particularly focuses on the behavioural patterns generated from the data collection by IoT based model which controls working of Air conditioner. This model fully works on remote less mode. This IoT based smart AC controller model operates under various indoor environmental conditions. Ideally, such models should follow reconcilable consumption of energy. The data set contains various parameters such as inner room and outer room temperature, inner and outer room humidity and people count. The real time data is logged on the the cloud depicting the behaviour of appliance under various environmental conditions. If the person is present inside the room and inner room temperature along with inner humidity values exceeds some threshold value then AC will be on otherwise AC will remain off. This model purely works on the demand basis and thus saving the consumption of power [1-3]. Some of the research studies have been done on the controlling of lights, sounds and other devices; controlling based on user’s emotion. This study carefully analyses user needs and conditions of the surroundings to predict the future actions with least human interventions. For this author highlighted various machine learning algorithms and reviews on multiple machine learning algorithms [4-5].

Irregular operational data values may lead due to the unexpected environmental variations or insufficient appliance usage. If data set is used directly for the learning purpose without pre-processing and cleaning, application of machine learning algorithms may result in less accuracy and showing low quality of data collected. To reveal the real useful characteristics of logged real time data, unsupervised learning algorithms are used. The unsupervised learning algorithms automatically discover the hidden patterns and intrinsic characteristics of the collected sensor data [6]. It does not rely on the label’s outputs. This unsupervised learning, drives the mechanism to unhide the hidden patterns of the data and detect the anomalies. K means clustering, an unsupervised approach is useful to group similar data points [7]. This grouping can happen based on intrinsic similarities in data values. Clustering, as an unsupervised learning approach, enables the automatic grouping of similar data instances based on inherent similarities in environmental and operational characteristics [8]. By applying clustering techniques to the collected IoT sensor data, the operational behaviour of the air conditioning system can be segmented into distinct usage patterns without requiring prior knowledge of normal or abnormal conditions. It helps in analysing behavioural patterns of the use of air conditioner by keeping separate unusual usage patterns. Hence, the responsible factor behind the energy inefficiency can be traced out [9-10]. The use of unsupervised data and the machine learning algorithms are widespread in each domain. The use of clustering algorithms for finding out the anomalies among data can found in steel industries also. Fen Jin and et.al. in 2023, explains the importance of data acquisition for the prediction analysis and scheduling operations in the steel industry. Author focuses on the HCA-DBSCAN model. The applied model proven to be the best for anomaly detection in multi-dimensional data[11].G.R. Venkatakrishnan and et al put an approach of smart meter that smartly and electronically records the data of electricity consumption. To avoid the financial losses and meter operational inefficiencies, it is very important to find the energy theft and defective meters. Author proposed the application of linear regression algorithm, random forest, extra trees algorithms along with the k- means algorithm. To balance the data SMOTE technique was also discussed [12]. Data acquisition plays very important role in any area. If the data is clean and clear, it gives the accurate and exact prediction results. Inas Ismael has focused the the intrusion detection in military data. Anomaly detection and identifying major breaches among data helps to predcit the possible attack. This research focuses on the application of Random forest and DBSCAN along with the K means clustering algorithms [13]. For the better performance of home appliances, anomalies of the data pattern should be found out. Jesmeen Mohd Zebaral worked on practical model for the data analysis. Moving Average approach identifies and removes the abnormal energy consumption of the data. This research also focuses on the Auto regressive integrated moving average i.e. ARIMA method [14]. Ana Lorena Jiménez-Preciado had done a very vast study on the CO2 emissions among 208 countries. Author elaborates the application of K means clustering algorithm in fusion with PCA and t-SNE visualization approaches. The findings are useful to reveal the significant patterns in worldwide CO2 emissions. From this, the author implied the that there are major climate changes [15].

2. Methodology

The AC controlling model based on IoT, works in different environmental conditions. These different environmental conditions are the real time data values. Table 1 depicts the data logged with the sensors on the cloud.

Table 1. The real time data set logged on to the cloud

Outer Temp	Inner Temp	Inner Humidity	Outer Humidity	No of persons	AC status	Time
23.5	23	54.8	52.6	0	OFF	8/4/2025 0:19:08
23.5	23	54.7	52.5	0	OFF	8/4/2025 0:20:15
24.5	24.5	54.6	52.4	0	OFF	8/4/2025 0:21:09
25.0	25.8	54.5	52.4	1	ON	8/4/2025 0:22:08
25.5	25.8	54.5	52.3	1	ON	8/4/2025 0:23:08
25.8	25.8	54.7	52.6	1	ON	8/4/2025 0:24:10
26.0	25.8	54.6	52.5	1	ON	8/4/2025 0:25:11
26.0	25.8	54.6	52.4	1	ON	8/4/2025 0:26:08
26.0	26.0	54.5	52.3	1	ON	8/4/2025 0:27:08
25.0	26.0	54.3	52.1	2	ON	8/4/2025 0:28:08
26.5	26.5	54.3	52.2	2	ON	8/4/2025 0:29:09
26.8	26.8	54.5	52.3	2	ON	8/4/2025 0:30:08
26.8	26.8	54.7	52.5	3	ON	8/4/2025 0:31:08

The entire system workflow involves-

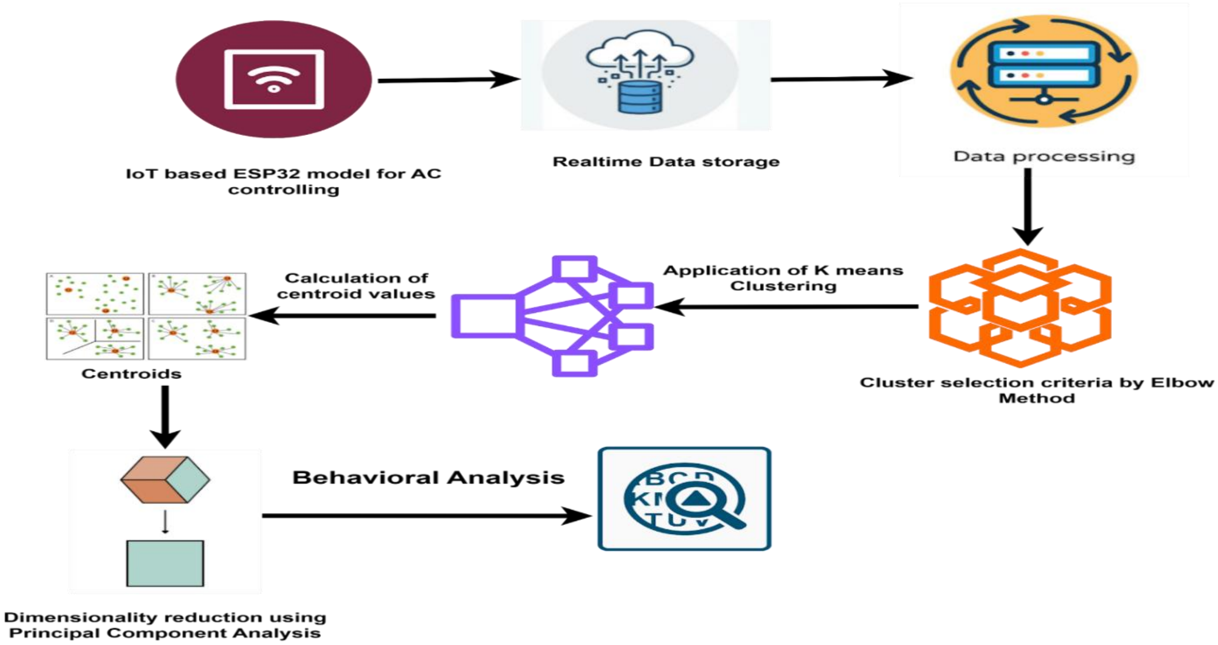

Fig. 1. (a) The entire system workflow

The working of the proposed IoT model can be interpreted in mathematical form as-

$S = \{ V_1, V_2, V_3 \ldots.. V_n \}$

where S= proposed IoT model

Each data point V_i contains features such as- temperature, humidity and person count.

We are dealing with unsupervised learning to find the anomalies among the data, AC status as ON or OFF is omitted.

Once the data set is ready, it should be scaled up to form the proper clusters. In the data processing scaling of data values would be done. Min-Max normalization is applied as a part of data pre-processing to rescale the data values within a fixed range of 0 to 1 using the following transformation:

$V_{\mathbf{processed}} = V_{\mathbf{i}} - V_{\mathbf{i\ min}} \;/\; V_{\mathbf{i\ max}} - V_{\mathbf{i\ min}}$

where

$V_{\mathbf{i}}$ = original data value

$V_{\mathbf{processed}}$ = processed data value

$V_{\mathbf{i\ min}}$ = minimum data values of respective data

$V_{\mathbf{i\ max}}$ = **maximum** data values of respective data

After the data pre-processing, cluster formation would be done. For the cluster formation, number of clusters that must be formed can be automatically decided with the help of Elbow method. Appropriate number of clusters gives appropriate results by clustering algorithms. Hence, elbow method is more suitable and given as- $WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$

Where

WCSS means within clusters sum of series k= number of clusters

Cj = j^{th} cluster

μj = centroid of j^{th} cluster

From this we get optimal k value. It can be well illustrated from below graph -

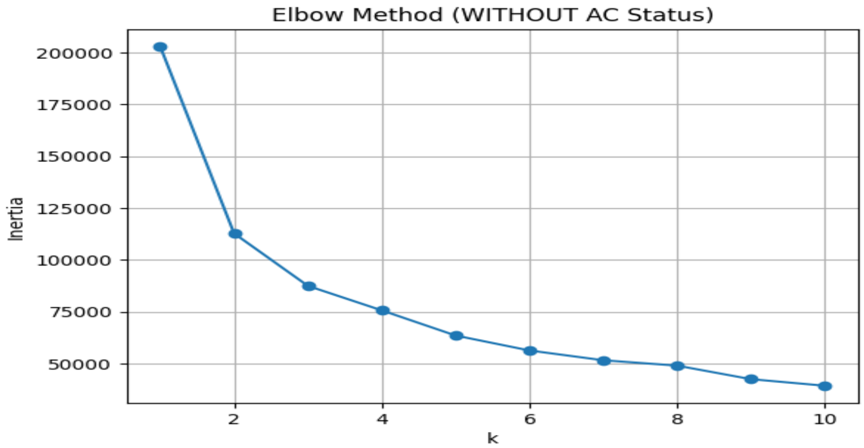

Fig. 1. (b) The Elbow Method

Silhouette Scores for unsupervised data (WITHOUT AC Status):

k=2 → score=0.4029

k=3 → score=0.4066

k=4 → score=0.3193

k=5 → score=0.2860

k=6 → score=0.2897

k=7 → score=0.2888

k=8 → score=0.2846

k=9 → score=0.3158

k=10 → score=0.2902

Therefore, automatically selected best k = 3 (number of clusters)

Best Silhouette Score = 0.4066

Above graph, Silhouette score was calculated as 0.4066. It interprets that clusters are separated on moderate level, that clarifies recognizable behavioural usage patterns in the data values. Also the environmental changes and variations in the readings results in the partial overlap between clusters.

From the Elbow method, we automatically get an ideal value of number of clusters on the dataset that is 3.

The graph interpreted the elbow point at 3 at the X axis and Y axis contains inertia range from 50000.

3. Results and discussions

The maximum cluster count obtained from the Elbow Method. Each data point is assigned to a cluster. Centroids are computed for each cluster. Centroids represent the mean of all points in a cluster across the data points. Finally, centroid heatmaps for analysing the operational characteristics of each behavioural usage pattern.

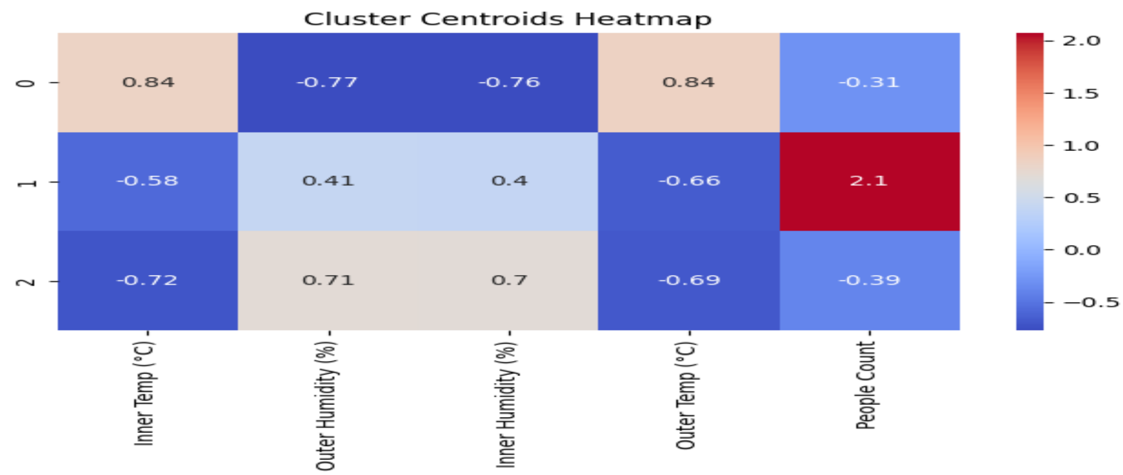

Fig. 1. (c) The Elbow Method

The mean of data values that corresponds to each other and belongs to one cluster pictorates the Centroid Heat Map. In above Centroid heat map, total clusters are 3 viz; 0,1,2 (shown in row) and columns are showing the data features such as inner temperature, outer humidity, inner humidity, outer humidity and people count. Following interpretations can be drawn from the heatmap, as -

Cluster 0 and Cluster 2 shows less people count whereas Cluster 1 shows very high level of people count.

On the contrary, Cluster 0 and Cluster 2 shows low level of people count. Also, the humidity levels in Cluster 2 are slightly higher than other clusters, showing that unnecessary cooling for less people count.

Principal Component Analysis (PCA) was applied to the clusters that are obtained from K-Means. Application of PCA was done to further authenticate the clusters and to reduce the multi-dimensional feature space into two principal components. Each data point was projected onto the PCA space. The clusters were envisage based on K-Means assignments. This visualization praises the cluster centroid heatmap by providing 2D and 3D representation of the cluster separation, confirming that the centroid differences correspond to different operational patterns of the AC system.

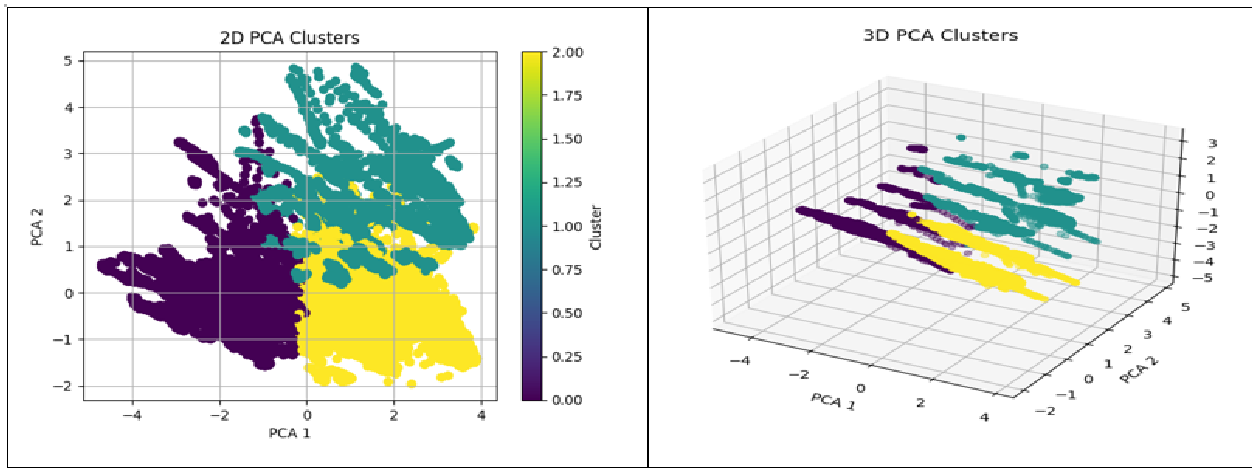

Fig. 2. (a) 2D PCA (b) 3D PCA.

The dataset includes: indoor temperature, outdoor temperature, indoor humidity, outdoor humidity and people count. Principal Component Analysis (PCA) was applied to the dataset to reduce dimensionality and identify the dominant patterns.

Fig 1 demonstrates the 2-dimensional PCA, PCA1 is influenced by indoor and outdoor temperature values, indoor and outdoor humidity values, whereas PC2 is influenced by people count. In PC2, high values demonstrating that highly crowded room. Moderate values indicate partially occupied rooms whereas low values indicate empty room or room occupied with one or two persons. PCA1 contains some negative values that interprets cool environment, near to zero values indicates equivalence in indoor and outdoor conditions and the positive value shows the equivalent values of inner temperature and inner humidity.

Cluster 1 —Location: Negative PC1 region (purple colour) represents a <u>*balanced indoor environment.*</u>

Cluster 2 — Location: Positive PC1 and low PC2 region (yellow colour) indicates low person count.

Cluster 3 — Location: Positive PC1 and high PC2 region (green colour) shows increase in person count and increase in indoor temperature causing load on air conditioner.

Fig 2 demonstrates the three-dimensional PCA as PCA1, PCA2 and PCA3. It gives a clearer visualization of cluster separation as compared to 2D PCA. The data points spread over along the PCA1 axis specifies that there are variations in person count data points. This variation affects the usage of air conditioner. PCA2 axis interprets the moderate differences in indoor environmental datapoints. This difference results in variety of AC usages. PCA2 axis suggests that moderate differences in indoor environmental parameters contribute to distinguishing different AC usage patterns. PCA3 identifies the minor differences in AC usage patterns happened due to changes in environment and person count.

4. Conclusion

This study focuses on the operational behaviour of air conditioner system. This system is fully controlled by IoT based model that used ESP32 microcontroller. This model records real time environmental data along with the person count. The operational behaviour of air conditioner is fully depending on the environmental temperature and humidity conditions in combination with person count. Clusters should be formed to analyse the data properly. Cluster count and formation can be done with the help of elbow method and k means clustering algorithms respectively. Cluster centroid heatmap provides insights on variations in data points. Principal component analysis was performed for visualization. These findings help to find out the operational behaviour of the air conditioner and helps to find out the factors that are responsible for energy consumption. This approach fully works on unsupervised data as it does not require labelled data.

References:

1. Mena A. R., Ceballos H. G.and Alvarado-Uribe J. (2022) Measuring Indoor Occupancy through Environmental Sensors: A Systematic Review on Sensor Deployment, MDPI https://doi: 10.3390/s22103770.
2. Wilhelm, Sebastian, Jakob, Dietmar, Jakob and Dietmar (2020) Human presence detection by monitoring the indoor CO2concentration https://doi: 10.1145/3404983.3409991
3. Mendell, Mark J. et. al. (2024) Carbon dioxide guidelines for indoor air quality: a review Journal of Exposure Science and Environmental Epidemiologyhttps://www.nature.com/articles/s41370-024-00694-7
4. John Jaihar; Neehal Lingayat (2023); Patel Sapan Vijaybhai; Gautam Venkatesh; K. P. Upla (2020) IEEE Xplore 10.1109/INCET49848.2020.9154007
5. Tsvetelina Mladenova; Irena Valova A Review of the Application of Machine Learning in Home Automation IEEE XploreDOI 10.1109/HORA58378.2023.10156796
6. Cheng Fan, Xiao Fu, Zhengdao Li, Jiayuan Wang (2017) Unsupervised Data Analytics in Mining Big Building Operational Data for Energy Efficiency Enhancement: A Review, Energy buildings, https://doi.org/10.1016/j.enbuild.2017.11.008
7. Dingsheng Deng (2020) Research on Anomaly Detection Method Based on DBSCAN Clustering Algorithm,IEEE Explore,2020, 10.1109/ISCTT51595.2020.00083
8. Marcello F; Virginia P(2023) Energy Consumption Profiling Of Appliances Inside Smart Buildings Based On k-means Clustering, IEEE Xplore, 10.1109/MeditCom58224.2023.10266646
9. Srinivasan A; Parmar V; Tom Oh; Jungwoo Ryoo; Mark Viglione (2021) Anomaly Detection System for Smart Home using Machine Learning, IEEE Xplore,10.1109/ICSSA53632.2021.00018
10. Modi, Krishna (2016) Outlier Analysis Approaches in Data Mining International Journal of Innovative Research in Technology
11. Feng Jin(2023) Varying-scale HCA-DBSCAN-based anomaly detection method for multi-dimensional energy data in

steel industry, sciencedirect https://doi.org/10.1016/j.ins.2023.119479
12. Venkatakrishnan G.,R. Rengaraj R (2025) Anomaly Detection in Smart Metering: Clustering-Based Identification of Energy Theft, IEEE Xplore, 10.1109/ICCCT63501.2025.11019827
13. Inas I(2024) Exploiting Anomalies with Data Mining Techniques to Enhance Cloud Security, Mathematical Modelling of Engineering Problems https://doi.org/10.18280/mmep.120227.
14. Jesmeen H(2022) Anomalies detection for smart-home energy forecasting using moving average, International Journal of Electrical and Computer Engineering (IJECE), 10.11591/ijece. v12i6.pp5808-5820
15. Ana L(2024) Identification of Patterns in CO2 Emissions among 208 Countries: K-Means Clustering Combined with PCA and Non-Linear t-SNE Visualization, MDPI https://doi.org/10.3390/math12162591.