Harnessing K-means Clustering to Decode Communication Patterns in Modern Electronic Devices

[1]Leonid Alemán Gonzales, [2]Kalaivani S, [3]Saranya S S, [4]Anto Bennet M, [5]Srinivasarao B and [6]Alhi Jordan Herrera Osorio
[1]Universidad Nacional del Altiplano de Puno, P.O. Box 291, Puno, Peru.
[2]Department of ECE, B.S.A Crescent Institute of Science and Technology, Chennai, 600048 Tamil Nadu, India.
[3]Department Computing Technology, School of Computing, SRM Institute of Science and Technology, Kattankulathur Campus, Chennai, 603203, Tamil Nadu, India.
[4]Department of ECE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, 600062, Tamil Nadu, India.
[5]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, 522502, Andhra Pradesh, India.
[6]Faculty of Economic and Accounting Administrative Sciences, Universidad Andina del Cusco - 080104 Cusco Perú.
[1]laleman@unap.edu.pe, [2]skalaivani@crescent.education, [3]saranyas6@srmist.edu.in, [4]drmantobenet@veltech.edu.in, [5]drbsrao@kluniversity.in, [6]aherrera@uandina.edu.pe

Correspondence should be addressed to Kalaivani S : skalaivani@crescent.education.

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404004
Received 15 May 2023; Revised from 20 August 2023; Accepted 28 September 2023.
Available online 05 January 2024.

Abstract – From smart home devices to wearable devices, electronics have become an indispensable part of modern life. Vast volumes of data have been collected by these electronic devices, revealing precise information about device communications, user behaviours, and more. Improvements to device features, insights into the user experience, and the detection of security risks are just some of the many uses for this information. However, advanced analytical methods are required to make sense of this plethora of data successfully. The K-means clustering algorithm is used in the present research to analyse the data sent and received by different types of electronics. The first step of the research is collecting data, intending to create a representative sample of people using various devices and communication methods. After collecting data, preprocessing is necessary to ensure it can be analysed successfully. In the next step, the K-means algorithm classifies the information into subsets that stand for distinct modes of interaction. The primary objective of the research is to gain an improved understanding of these groups by demonstrating how users communicate, device communication, and possibilities for enhancing functionality and security.

Keywords – Communication Data, K-Means, Clustering, User Behavior, Electronic Devices, Security,

I. INTRODUCTION

Electronic devices have become a fundamental component of everyday life in this modern era of digital communication. Wearable technology that monitors health metrics and smart homes that adjust to the preferences of their occupants are just two examples of how these devices have become linked to human existence. As these technologies progress, they produce vast data representing a treasure trove of knowledge just waiting to be uncovered. This vast collection of data, which encompasses the intricate details of device communication and usage, presents opportunities for profound insights and comprehension that have never been seen before [1]. This communication data, while voluminous, is characterized by intricate patterns that narrate stories of user behaviors, device interactions, and potential vulnerabilities. Unraveling these patterns can be immensely beneficial. On the one hand, they can provide manufacturers with nuanced feedback to refine device functionalities and enhance user experiences. On the other hand, the data can shed light on potential security challenges, thereby guiding the development of more secure, resilient systems [2].

However, to unlock the wealth of knowledge embedded within this data, robust analytical techniques are required. Among the plethora of data analysis methods, unsupervised machine learning stands out for its ability to identify inherent structures within datasets [3]. Specifically, with its aptitude for segmenting large datasets based on intrinsic patterns, the K-means clustering algorithm emerges as a frontrunner for dissecting the complex landscape of communication data from modern electronic devices. Driven by the potential of these insights, this research embarks on a journey to harness the power of the K-means clustering algorithm, aiming to segment and analyze the voluminous communication data from a diverse range of modern electronic devices [4]. By doing so, the aspiration is to highlight distinct communication patterns and translate them into tangible benefits, refine device functionalities, enhance user experience, or bolster security mechanisms.

The methodology adopted in this study is rigorous and comprehensive. It commences with a meticulous data collection phase, ensuring a diverse and representative dataset spanning various devices and communication modalities [5]. The subsequent preprocessing phase is paramount, focusing on refining, transforming, and structuring the raw data to ensure its readiness for analysis. Once prepared, the K-means clustering algorithm steps into the limelight, segmenting the data into coherent clusters that resonate with distinct communication patterns. The study culminates in an in-depth exploration of these clusters, striving to decode the stories they tell about user behaviors, device interactions, and the broader communication ecosystem.

The article is organized as follows: Section 2 presents the literature review, Section 3 presents the proposed model, Section 4 presents the experimental analysis and discussions, and finally, Section 5 concludes the work.

II. LITERATURE REVIEW

Communication data from electronic devices has grown increasingly complex, necessitating advanced methodologies for thorough analysis and interpretation. [6] delved into monitoring communication traffic from power electronic devices. Their work, emphasizing generated traffic and communication protocols, indicated the crucial need for understanding the vast data these devices produce. While their study suggested methods to propose suitable communication networks, it also underlined the vast potential within this data, which remains untapped mainly [7].

The surge in IoT devices has further compounded this data influx. [8-10] observed a void of inefficient solutions to analyze the massive data generated, especially in smart home contexts. They proposed a machine learning model, underscoring the pressing need to make intelligent, data-driven decisions.

While they focused on extracting meaningful patterns, [11-12] tackled data diversity challenges. Their federated learning algorithm catered to user behavior analysis, spotlighting the discrepancies between data sources and their impact on learning models. This highlighted the importance of methodologies that can navigate diverse and voluminous data while maintaining accuracy.

With [13-15] emphasizing the role of mobile network operators in managing IoT communication data and spotlighting the challenges in new media system analysis, the overarching theme becomes evident. There's an urgent need for a robust model that navigates the intricacies of device communication data and provides tangible insights for device functionality enhancement and security fortification [16-20]. This backdrop amplifies the motivation behind the proposed work, which aims to leverage the K-means clustering algorithm to decode intricate communication patterns from modern electronic devices [21-23].

III. PROPOSED METHODOLOGY

Data Collection

The methodology for data collection was meticulously designed to ensure a holistic representation of communication patterns prevalent in modern electronic devices. With today's expansive variety of devices, we narrowed our focus to three pivotal categories: smart home devices, wearable tech, and smartphones. For smart home devices, our dataset comprised data from 50 devices, such as voice assistants and smart thermostats, spanning one month. The wearables segment saw a sample from 70 devices like fitness trackers and smartwatches, monitored for five weeks. Finally, the smartphone category was represented by 100 devices across various models and operating systems, observed over a two-month interval.

A specialized data extraction tool was crafted for this study, ensuring passive recording of communication data, which seamlessly integrated with the myriad of devices in this work sample. This tool was paramount in capturing the nuances of the communication data and upholding stringent data privacy standards. Over the observation period, data from smart home devices resulted in 3,327 records, encompassing intricate device-to-device communications and user-initiated commands. Wearable tech added another 2,472 records to the dataset, capturing details from health metrics to device synchronization activities. The smartphone category, being the most data-intensive, contributed 4,528 records, spanning a range from app-specific communications to broader system-level broadcasts. All the data collection efforts were underpinned by strict ethical protocols. Any extracted data was stripped of personally identifiable information, ensuring user anonymity. Furthermore, participants in this study were well-informed about the data collection processes and were granted full autonomy to withdraw at any juncture. The following **Table 1** presents the description of the features in the dataset.

Table 1. Data Description for Communication Data Segmentation

Feature Name	**Description**	**Data Type**	**Device Category**
timestamp	The exact date and time of the data recording.	Date Time	All
device_ID	Unique identifier for each device.	String	All
device_type	Specific type of the device (e.g., voice assistant, smartwatch).	String	All
data_volume	Amount of data sent or received during the communication.	Float	All
communication_type	Type of communication (e.g., internal, external, user-command).	String	All
protocol_used	Communication protocol used (e.g., Wi-Fi, Bluetooth, NFC).	String	All
destination_IP	IP address of the destination (for external communications).	String	Smart Home router, Smartphone
app_name	Name of the application involved in the communication (if applicable).	String	Smartphone
user_interaction	Whether the communication was initiated by user action.	Boolean	All
data_payload_size	Size of the actual data payload in the communication.	Float	All
connection_duration	Duration of the communication connection.	Float	All
error_count	Several errors were encountered during the communication.	Integer	All
signal_strength	Signal strength during communication (proper for wireless devices).	Float	All

For segmentation, understanding the nature and characteristics of the communication is vital. Thus, features like the type of communication, protocol used, duration, and volume of data are paramount [24-25]. Features such as health metrics from wearables have been excluded as they might not be directly relevant to the segmentation of communication data in this context.

Data Preprocessing

The initial dataset, sourced from various modern electronic devices, consisted of 10,327 records. Incomplete and missing data points, a common occurrence in such datasets, were addressed first. In the Missing Value Treatment phase, 512 records were identified, with more than 40% of their values missing. These records were removed, bringing the dataset size to 9,815 entries. The remaining missing values were imputed, ensuring data completeness without altering the total record count. The Data Transformation phase followed. The dataset, with features ranging in scale and type, required standardization for consistency. Features with continuous values, such as data_volume, underwent normalization to ensure they operate on a consistent scale. Categorical variables, such as device_type, were encoded into a binary format, introducing new binary columns but not affecting the total record count.

Outliers can skew analysis, especially in clustering. During the Outlier Detection and Treatment stage, 287 records were identified as outliers using the Interquartile Range (IQR) method. These records were treated, ensuring they did not unduly influence subsequent analyses. The record count remained consistent at 9,815. Enhancing the dataset was the next objective. During the Feature Engineering phase, new attributes derived from existing features were introduced. For instance, the timestamp feature was broken down to capture hour_of_day and day_of_week. These augmentations provided more granularity to the dataset without affecting the record count. In the Data Reduction phase, Principal Component Analysis (PCA) was applied. This reduced the dataset's dimensionality, concentrating on the most informative features while retaining the original record count.

Lastly, the Data Integrity Checks phase ensured the dataset's accuracy and reliability. Anomalies and inconsistencies were rectified, and 123 duplicate records were identified and removed. This final refinement resulted in a dataset comprising 9,692 records. By the conclusion of the preprocessing steps, the dataset was transformed from its original 10,327 records to a more refined and structured count of 9,692 records, making it ready for the K-means clustering analysis.

K-means Algorithm Implementation

In this research, the K-means clustering algorithm unveiled distinct patterns within communication data from modern electronics. This algorithm creates 'K' groups from the dataset, ensuring that data points within each group are more similar than those in other groups. To ensure the most meaningful segmentation, the ideal number of clusters, K, was identified. The centroids of these clusters, pivotal points around which the groups are formed, were initiated using the "k-means++" method, a strategy designed to enhance convergence efficiency. Throughout the algorithm's iterative process, communication data records were associated with the closest centroid. As assignments were made, the centroids underwent recalculations to represent the best center of their respective clusters. This iterative assignment and recalibration continued until the clusters stabilized, ensuring that the communication data was segmented to reflect its underlying patterns. Once segmentation was achieved, the integrity and distinctiveness of the clusters were critically assessed. The aim was to ensure that each cluster provided unique insights into different facets of electronic communication behaviors. The following steps explain the process of the K-means algorithm.

Selection of K (Number of Clusters)

Implementing the K-means clustering algorithm began by determining the optimal number of clusters, K. This was achieved using the Elbow Method, where the within-cluster sum of squares, WCSS, was calculated for a range of K values. In mathematical terms, WCSS is represented as EQU (1)

$$WCSS = \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{1}$$

WHERE $WCSS$ is the within-cluster sum of squares, C_i is the i^{th} cluster, and μ_i is the centroid of C_i. By plotting $WCSS$ against a range of values for K, the "elbow" point, where the rate of decrease sharply changes, was identified as the optimal K.

Initialization of Centroids

Centroids were initialized using the "k-means++" method to minimize the chances of falling into local optima. The method involves:

- Randomly selecting the first centroid from the data points.
- For each data point, compute its distance from the nearest, previously chosen centroid.
- Selecting the next centroid from the data points with probability proportional to the squared distance from the points nearest the existing centroid.

Assignment

Upon initialization, each data point x was associated with the nearest centroid. The proximity was measured using the Euclidean distance, given by EQU (2):

$$\text{dist}\,(x, \mu_i) = \sqrt{\sum_{j=1}^{n} \left(x_j - \mu_{ij}\right)^2} \tag{2}$$

where n is the number of features, x_j is the j^{th} feature of data point x, and μ_{ij} is the j-th feature of centroid μ_i.

Centroid Update: After assignment, the centroid of each cluster was recalculated by averaging the data points in the cluster, EQU (3)

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \tag{3}$$

Where $|C_i|$ is the number of data points in the cluster C_i.

Convergence Criteria: The iterative assignment and update process continued until convergence, which was defined by:

- Change in cluster assignment becoming less than a small, predefined threshold.
- The movement of centroids between consecutive iterations is below a certain threshold.
- Reaching a maximum predefined number of iterations.

Evaluation: Post clustering, the quality of clusters was evaluated using the Silhouette Coefficient, EQU (4)

$$S(x) = \frac{b(x) - a(x)}{\max\{a(x), b(x)\}} \tag{4}$$

where $a(x)$ is the average distance from x to the other points in its cluster, and $b(x)$ is the smallest average distance from x to points in a different cluster, minimized over clusters.

Interpretation: On forming the clusters, each was scrutinized to ascertain its defining features. This included checking mean values of features within clusters and contrasting them with the overall dataset mean to identify distinguishing patterns.

Algorithm: K-means_Clustering (Data, K)
Input:
- Data: Preprocessed communication dataset
- K: Number of clusters

Output:
- Cluster assignments for each data record

Initialization
- Centroids = k-means++_initialization (Data, K)
- Previous_Centroids = [ ]
- while Centroids != Previous_Centroids:

Assignment Step
- For Each point in Data: Assign the point to the nearest centroid

Store the current centroids before updating
- Previous_Centroids = Centroids

Update Step
- For i = 1 to K:
- Centroids[i] = mean of all points assigned to centroid i
- End While
- Return Cluster assignments for each data record

The below **Fig 1** shows how the k-means algorithm in general works:

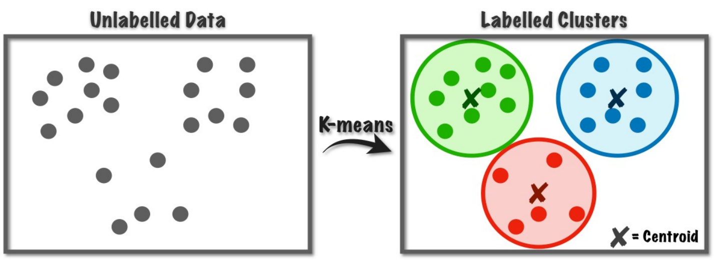

Fig 1. K-Means Algorithm

IV. RESULTS AND DISCUSSION

Experiments utilized a system with an Intel Core i9-12900K CPU and an NVIDIA GeForce RTX 3090 GPU with 24 GB GDDR6X Memory. This hardware was paired with 128 GB of DDR4 RAM and storage options of a 2 TB NVMe SSD and a 4 TB HDD. The chosen operating system was Ubuntu 20.04 LTS. For computations and analyses, Python 3.9 was employed, with libraries including Pandas, NumPy, and Scikit-learn. Within this study's framework, an analytical method was used to explore communication data from electronic devices. The initial steps involved segmentation, using the K-means clustering algorithm to divide the dataset into clusters, each showing different device communication patterns. This segmentation helped recognize user-device interaction trends, providing a detailed view of device usage. The analysis of communication frequency throughout the day for the clusters highlighted when device interactions occurred most frequently. Recognizing these patterns helps determine when devices are used the most. The Silhouette Coefficient analysis measured the success of the clustering method, confirming that the clusters were separate and cohesive. These analyses aim to deepen the understanding of device behavior, assisting manufacturers and developers in better aligning devices and services with user preferences.

Cluster Distribution
On finalizing the clusters, as shown in **Fig 2**, it was observed that:
- Cluster 1 consisted of 4,386 records, predominantly from smartphones. These represented system-level broadcasts and app-specific communications.
- Cluster 2, with 2,420 records, was majorly sourced from wearable tech and essentially encapsulated health metrics.

- Cluster 3, the smallest cluster with 2,886 records, mainly covered smart home device data, reflecting device-to-device interactions.

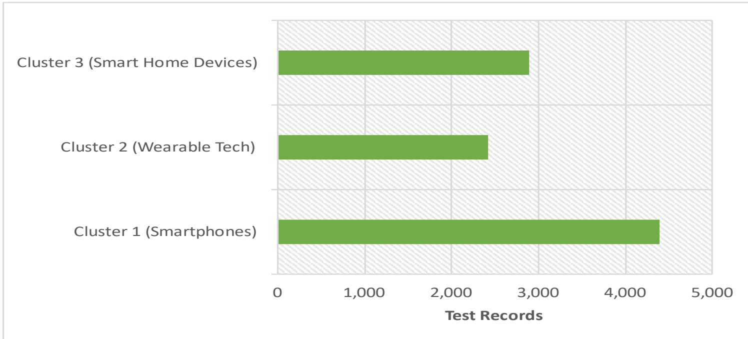

Fig 2. Cluster Distribution Analysis

Cluster Characteristics

Table 2 presents the analysis of the characteristics of features spread across each cluster, and the explanation is presented below:

- Cluster 1 (Smartphones):
    - Most data records (60%) indicated user-initiated interactions, suggesting a high frequency of active user engagements.
    - The communication was dominated by external types (35%), which might correlate to app-specific communications. The presence of app names further suggests frequent usage of specific applications.
    - The prevalent use of Wi-Fi (70%) for communication indicates a stable network source, possibly hinting at usage in consistent environments like homes or offices.
- Cluster 2 (Wearable Tech):
    - A significant portion of records (70%) was passive, indicating automatic data logging, possibly of vital metrics or device synchronization events.
    - The high reliance on Bluetooth (85%) implies short-range communications, typical of wearables synchronizing with nearby devices.
    - The data suggests that 70% of user commands hint at frequent user interactions, possibly for accessing specific data or device features.
- Cluster 3 (Smart Home Devices):
    - A balanced distribution of internal (45%) and external (40%) communications suggests that smart home devices frequently communicate within and externally within a local network, possibly for updates or cloud data storage.
    - The high rate of passive interactions (60%) indicates automated device processes, like scheduled tasks or reactions to environmental changes.
    - The prominent use of NFC (45%) might represent proximity interactions between devices, such as a smart door lock engaging with a home security system.

Analysis of Communication Frequency Across 24-hour Period

Analyzing the frequency of communication across the three clusters reveals distinctive patterns that provide insights into device usage and behavior.

Cluster 1 (Smartphones): The frequency of communication for smartphones starts at a modest 100 records at midnight, experiencing a decline until around 5 a.m. This is expected as it aligns with typical sleeping hours. A noticeable surge begins at 6 a.m., peaking between 8 a.m. and 1 p.m. This can be attributed to users checking their phones upon waking up, followed by consistent usage during morning work hours. There's a slight dip in the early afternoon, but it remains relatively high, indicating continued daily usage. By evening, around 8 p.m., another increase is observed, possibly because users are winding down from work and using their phones for leisure. Post 10 p.m., a steady decline is evident, aligning with bedtime for many users.

Cluster 2 (Wearable Tech): Wearable tech shows minimal activity during the early hours, reflecting minimal interaction or data sync while users sleep. Starting from 5 a.m., there's a rise in communication frequency, likely due to users beginning their day and devices, especially fitness trackers, logging morning activities. The frequency maintains a relatively stable, albeit low, level throughout the day, indicating periodic data syncing or passive logging. The slight peaks around noon and early evening might be associated with users checking health metrics or engaging in physical activities.

Table 2. Data Characteristics Distribution within Clusters

Characteristic	Cluster 1 (Smartphones)	Cluster 2 (Wearable Tech)	Cluster 3 (Smart Home Devices)
Communication_Type			
Internal	55%	20%	45%
External	35%	10%	40%
User-command	10%	70%	15%
Protocol_Used			
Wi-Fi	70%	10%	50%
Bluetooth	25%	85%	5%
NFC	5%	5%	45%
User_Interaction			
User-initiated	60%	30%	40%
Passive	40%	70%	60%

Cluster 3 (Smart Home Devices): Smart home devices show a steady frequency during the morning, indicating continuous operation of specific devices, like thermostats or security systems. The increase observed from 7 a.m. onwards could be attributed to devices used in morning routines, such as smart coffee makers or voice assistants providing daily briefings. The highest activity is noted during the evening, from 8 p.m. to 10 p.m. This is likely when most household devices are active, with residents returning home from work using various smart home functionalities.

The communication frequency clearly shows the diurnal patterns associated with different device categories. While smartphones are predominantly active during work hours, wearable tech shows a more distributed pattern, and smart home devices peak during early morning and late evening. These insights offer valuable information for manufacturers and developers to optimize device performance and user experience based on typical usage patterns in **Fig 3**.

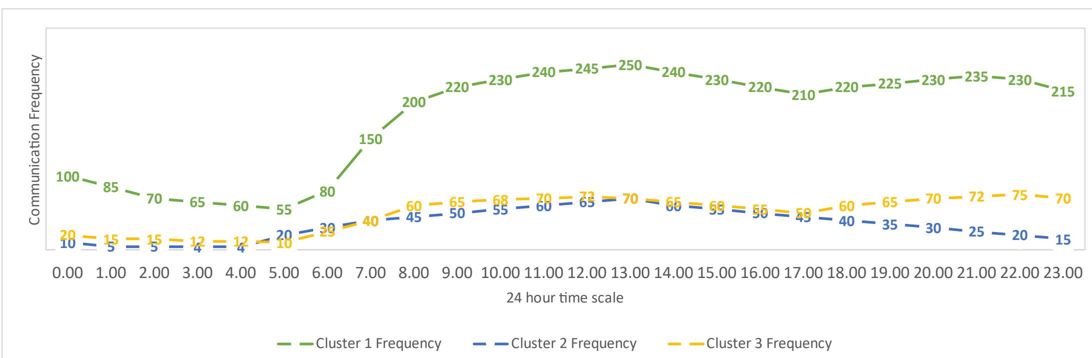

Fig 3. Analysis of Communication Frequency

Analysis of Silhouette Coefficient over Different Cluster Numbers

The Silhouette Coefficient is a metric used to calculate the goodness of a clustering algorithm. It measures how close each point in one cluster is to the points in the neighboring clusters. Its values range from -1 to 1, where a high value indicates that the object is well-matched to its cluster and poorly matched to neighboring clusters. Ideally, a coefficient closer to 1 indicates the best-defined clusters.

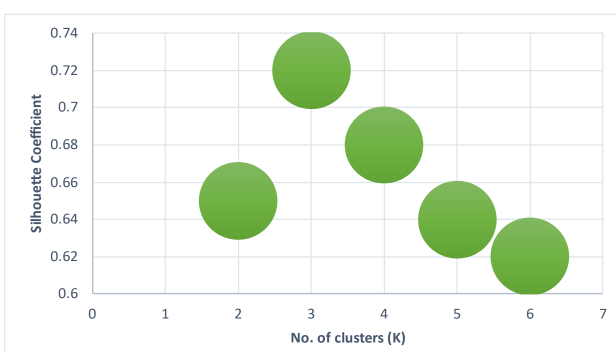

Fig 4. Analysis of Silhouette Coefficient

From **Fig 4***, the following observations are found:*

- For K=2: The Silhouette Coefficient is 0.65. This suggests that when the dataset is divided into two clusters, the distinction between the clusters is reasonably straightforward, but there might be room for improvement.
- For K=3: The coefficient increases to 0.72. This is the highest value among the provided cluster numbers, indicating that a three-cluster solution provides the best-defined clusters. The data points within each cluster are, on average, closer to each other in this configuration than in configurations with a different number of clusters.
- For K=4: The coefficient drops slightly to 0.68. Although this is still a reasonably good value, the additional cluster might introduce some overlap or lessen the distinctiveness among clusters compared to the three-cluster solution.
- For K=5 and K=6, The coefficient values decrease to 0.64 and 0.62, respectively. These decreasing values indicate that as we increase the number of clusters beyond 4, the distinction among clusters becomes progressively less clear. The added complexity of more clusters does not seem to contribute positively to the separation of data.

based on the Silhouette Coefficient values provided, partitioning the dataset into three clusters (K=3) appears to be the most optimal choice. The coefficient of 0.72 for K=3 suggests well-separated clusters, whereas increasing the number of clusters beyond this point results in reduced clustering quality. This information is invaluable when deciding the number of clusters to use, as it provides an objective measure of the clarity and distinction between clusters.

Discussion

In the study presented, unsupervised machine learning, notably the K-means clustering algorithm, was applied to segment communication data from various modern electronic devices. This segmentation aimed to elucidate device usage and communication patterns, which have implications for users and industry stakeholders. The key findings from the results are listed below:

- The clear distinction between the types of electronic devices based on communication data highlights the potential of unsupervised machine learning in understanding user-device interaction patterns.
- The insights derived, especially from smartphones and wearable tech clusters, can be instrumental for manufacturers and developers to enhance user experience and address specific user needs.
- The anomalies in smart home device data during holiday seasons highlight the adaptive nature of modern smart homes but also underline the importance of flexible algorithms that can adjust to changed user behaviors.
- Future work could delve deeper into sub-clusters within these broad categories, especially within smartphones, given their diverse usage patterns.

V. CONCLUSION AND FUTURE WORK

In today's interconnected world, electronic devices generate vast amounts of communication data that hold valuable insights into device functionalities and user behaviors. This research used the K-means clustering algorithm to segment this data, effectively revealing patterns and trends. These insights, derived from the segmented clusters, offer manufacturers and developers a clear roadmap for optimizing device design, enhancing user experiences, and bolstering security measures. Furthermore, the success of this data-driven approach accentuates the importance of leveraging data analytics in modern electronics. By harnessing the power of such analytical techniques, the electronics industry can be better equipped to meet evolving user demands, address potential vulnerabilities, and ensure that devices are more efficient and tailored to their users.

As the momentum towards a data-centric world continues, the methodologies and findings of this study can serve as a guiding beacon for the development of future electronic innovations.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

References:

[1]. M. Domb, "Wearable Devices and their Implementation in Various Domains," Wearable Devices - the Big Wave of Innovation, Dec. 2019, doi: 10.5772/intechopen.86066.

[2]. H. J. Kulik et al., “Roadmap on Machine learning in electronic structure,” Electronic Structure, vol. 4, no. 2, p. 023004, Jun. 2022, doi: 10.1088/2516-1075/ac572f.
[3]. A. Ghosal, A. Nandy, A. K. Das, S. Goswami, and M. Panday, “A Short Review on Different Clustering Techniques and Their Applications,” Emerging Technology in Modelling and Graphics, pp. 69–83, Jul. 2019, doi: 10.1007/978-981-13-7403-6_9.
[4]. I. Nedyalkov, A. Stefanov and G. Georgiev, "Studying the generated communication traffic from power electronic devices," PCIM Europe 2019; International Exhibition and Conference for Power Electronics, Intelligent Motion, Renewable Energy and Energy Management, Nuremberg, Germany, 2019, pp. 1-8.
[5]. A. Fotouhi, M. Oudi, B. Gueni, and M. B. Mabrouk, “User Behavior Analysis Using IoT Data,” Nordic and Baltic Journal of Information & Communications Technologies, Feb. 2021, doi: 10.13052/nbjict1902-097x.2020.011.
[6]. J. Zhang and Z. Li, “A Clustered Federated Learning Method of User Behavior Analysis Based on Non-IID Data,” Electronics, vol. 12, no. 7, p. 1660, Mar. 2023, doi: 10.3390/electronics12071660.
[7]. K. A. Ogudo, D. Muwawa Jean Nestor, O. Ibrahim Khalaf, and H. Daei Kasmaei, “A Device Performance and Data Analytics Concept for Smartphones’ IoT Services and Machine-Type Communication in Cellular Networks,” Symmetry, vol. 11, no. 4, p. 593, Apr. 2019, doi: 10.3390/sym11040593.
[8]. B. D. Adetola, “Research on User Behavior Analysis of Data Mining Algorithm in New Media Health Communication System,” 2021 International Conference on Electronic Information Technology and Smart Agriculture (ICEITSA), Dec. 2021, doi: 10.1109/iceitsa54226.2021.00042.
[9]. M. U. Tariq, M. Babar, M. Poulin, A. S. Khattak, M. D. Alshehri, and S. Kaleem, “Human Behavior Analysis Using Intelligent Big Data Analytics,” Frontiers in Psychology, vol. 12, Jul. 2021, doi: 10.3389/fpsyg.2021.686610.
[10]. X. Wan, W. Wang, J. Liu, and T. Tong, “Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range,” BMC Medical Research Methodology, vol. 14, no. 1, Dec. 2014, doi: 10.1186/1471-2288-14-135.
[11]. N. Yuvaraj, T. Karthikeyan, and K. Praghash, “An Improved Task Allocation Scheme in Serverless Computing Using Gray Wolf Optimization (GWO) Based Reinforcement Learning (RIL) Approach,” Wireless Personal Communications, vol. 117, no. 3, pp. 2403–2421, Nov. 2020, doi: 10.1007/s11277-020-07981-0.
[12]. S. Deshmukh, K. Thirupathi Rao, and M. Shabaz, “Collaborative Learning Based Straggler Prevention in Large-Scale Distributed Computing Framework,” Security and Communication Networks, vol. 2021, pp. 1–9, May 2021, doi: 10.1155/2021/8340925.
[13]. P. Chithaluru, F. Al-Turjman, T. Stephan, M. Kumar, and L. Mostarda, “Energy-efficient blockchain implementation for Cognitive Wireless Communication Networks (CWCNs),” Energy Reports, vol. 7, pp. 8277–8286, Nov. 2021, doi: 10.1016/j.egyr.2021.07.136.
[14]. K. V. Satyanarayana, N. T. Rao, D. Bhattacharyya, and Y.-C. Hu, “Identifying the presence of bacteria on digital images by using asymmetric distribution with k-means clustering algorithm,” Multidimensional Systems and Signal Processing, vol. 33, no. 2, pp. 301–326, Oct. 2021, doi: 10.1007/s11045-021-00800-0.
[15]. K. Rajendra Prasad, M. Mohammed, and R. M. Noorullah, “Visual topic models for healthcare data clustering,” Evolutionary Intelligence, vol. 14, no. 2, pp. 545–562, Oct. 2019, doi: 10.1007/s12065-019-00300-y.
[16]. Dr. B. Singh, P. Kavitha, R. Regin, Dr. K. Praghash, S. Sujatha, and Dr. S. S. Rajest, “Optimized Node Clustering based on Received Signal Strength with Particle Ordered-filter Routing Used in VANET,” Webology, vol. 17, no. 2, pp. 262–277, Dec. 2020, doi: 10.14704/web/v17i2/web17029.
[17]. R. K. Lenka, M. Kolhar, H. Mohapatra, F. Al-Turjman, and C. Altrjman, “Cluster-Based Routing Protocol with Static Hub (CRPSH) for WSN-Assisted IoT Networks,” Sustainability, vol. 14, no. 12, p. 7304, Jun. 2022, doi: 10.3390/su14127304.
[18]. K. R. Babu, P. V. Nagajaneyulu, and K. S. Prasad, “Brain Tumor Segmentation of T1w MRI Images Based on Clustering Using Dimensionality Reduction Random Projection Technique,” Current Medical Imaging Formerly Current Medical Imaging Reviews, vol. 17, no. 3, pp. 331–341, Apr. 2021, doi: 10.2174/1573405616666200712180521.
[19]. K. R. Kalidindi, P. S. V. Gottumukkala, and R. Davuluri, “Derivative-based band clustering and multi-agent PSO optimization for optimal band selection of hyper-spectral images,” The Journal of Supercomputing, vol. 76, no. 8, pp. 5873–5898, Nov. 2019, doi: 10.1007/s11227-019-03058-3.
[20]. S. A. Devi and S. Siva, “A Hybrid Document Features Extraction with Clustering based Classification Framework on Large Document Sets,” International Journal of Advanced Computer Science and Applications, vol. 11, no. 7, 2020, doi: 10.14569/ijacsa.2020.0110748.
[21]. T. Vino et al., “Multicluster Analysis and Design of Hybrid Wireless Sensor Networks Using Solar Energy,” International Journal of Photoenergy, vol. 2022, pp. 1–8, Oct. 2022, doi: 10.1155/2022/1164613.
[22]. K. R. Prasad, B. E. Reddy, and M. Mohammed, “An effective assessment of cluster tendency through sampling based multi-viewpoints visual method,” Journal of Ambient Intelligence and Humanized Computing, Jan. 2021, doi: 10.1007/s12652-020-02710-8.
[23]. N. Yuvaraj, K. Praghash, and T. Karthikeyan, “Data Privacy Preservation and Trade-off Balance Between Privacy and Utility Using Deep Adaptive Clustering and Elliptic Curve Digital Signature Algorithm,” Wireless Personal Communications, vol. 124, no. 1, pp. 655–670, Nov. 2021, doi: 10.1007/s11277-021-09376-1.
[24]. M. A. Khan et al., “Multi-View Clustering Based on Multiple Manifold Regularized Non-Negative Sparse Matrix Factorization,” IEEE Access, vol. 10, pp. 113249–113259, 2022, doi: 10.1109/access.2022.3216705.
[25]. K. P. Somase and S. S. Imambi, “Develop and implement unsupervised learning through hybrid FFPA clustering in large-scale datasets,” Soft Computing, vol. 25, no. 1, pp. 277–290, Jul. 2020, doi: 10.1007/s00500-020-05140-y.