Unsupervised Learning (Clustering) Activities of Daily Living Recognition Using Binary Sensors Data Set

Bhanu Prasad Dharavathu
Dept. of Artificial Intelligence
Kent State University
kent, ohio
811349718
bdharav1@kent.edu

Mounika Seelam
Dept. of Computer Science
Kent State University
kent, ohio
811344833
mseelam2@kent.edu

Prajwal Devaraj
Dept. of Computer Science
Kent State University
kent, ohio
811351381
pdevaraj@kent.edu

Abstract*—This project explores unsupervised learning techniques to identify Activities of Daily Living (ADLs) from binary sensor data recorded in smart home environments. The dataset comprises 35 days of sensor events for two users, collected via wireless motion and contact sensors deployed across residential rooms. Without using activity labels during training, we processed the sensor logs by segmenting them into fixed 10-minute intervals and extracting temporal and spatial behaviour features such as event frequency, location transitions, and entropy of movement. To improve cluster interpretability, we applied Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction. We implemented and compared K-Means, Agglomerative Hierarchical Clustering, and Gaussian Mixture Models (GMM) to group behavioral patterns. Post-clustering, we evaluated performance using the ground-truth ADL labels with external metrics. The best clustering performance was achieved using Agglomerative Clustering with a V-Measure of 0.58, Adjusted Rand Index (ARI) of 0.49, and Normalized Mutual Information (NMI) of 0.55. These results indicate a moderate but meaningful alignment between discovered clusters and real-life daily activities. This outcome highlights the feasibility of recognizing human routines through unsupervised learning, especially in scenarios where labelled data is limited or unavailable. Our work demonstrates that effective feature engineering and model selection can yield interpretable behavioral groupings that support further applications in ambient assisted living and intelligent health monitoring.

***Keywords*— Unsupervised learning, clustering, smart home, ADL recognition, binary sensors, feature extraction, PCA, t-SNE, K-Means, GMM, hierarchical clustering, V-measure, ARI, NMI.**

I. Introduction

In recent years, smart environments equipped with Internet of Things (IoT) devices have gained significant attention for their potential in monitoring and assisting individuals in daily life activities, particularly within healthcare and ambient assisted living (AAL) domains [1]. One critical application is the recognition of Activities of Daily Living (ADLs), which can provide insights into user behavior, detect anomalies, and support elderly or disabled individuals in maintaining independence [2]. Traditionally, ADL recognition has relied on supervised learning methods that require large amounts of labeled data. However, collecting labeled activity data is labor-intensive, error-prone, and often infeasible in real-world settings.

To overcome these challenges, this project explores unsupervised learning techniques to cluster behavioral patterns derived from binary sensor data without the need for activity labels. We utilize the "Activities of Daily Living Recognition Using Binary Sensors" dataset, which comprises over a month of sensor logs from two users in their respective homes. The data includes timestamps, sensor locations, and types of activation events, all recorded using a wireless sensor network [3]. By extracting meaningful temporal and spatial features from raw sensor logs, we aim to discover latent behavioral patterns that correlate with different ADLs.

Our methodology includes feature engineering based on location transitions and event durations, dimensionality reduction using Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE), and clustering via K-Means, Agglomerative Clustering, and Gaussian Mixture Models (GMM). The performance of each clustering approach is evaluated against the ground-truth ADL labels using standard external validation metrics such as Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and V-Measure [4]. Visualization through confusion matrices and t-SNE plots enables interpretation of the clustering quality and activity separation.

By eliminating the need for labeled training data, this project demonstrates how unsupervised learning can serve as a practical and scalable alternative for activity recognition in smart homes. Our findings contribute to the growing body of research on sensor-based behavioral analysis and pave the way for more autonomous, data-driven monitoring systems in the healthcare domain.

II. Background

The Unsupervised learning techniques have become increasingly relevant for analyzing data in domains where labeled instances are scarce or expensive to obtain. The recognition of Activities of Daily Living (ADLs) is a prime candidate for such techniques, especially in smart home environments that continuously generate large volumes of sensor data.

A. Activities of Daily Living (ADLs)

ADLs refer to fundamental self-care tasks such as eating, bathing, dressing, and mobility, which are crucial indicators of an individual's health status and quality of life [5]. Accurate recognition of these activities enables automated health monitoring and early detection of behavioral changes, particularly for the elderly or individuals with chronic conditions. Traditionally, ADL recognition has relied on supervised learning, requiring labeled datasets collected via observation or self-reporting, both of which are time-consuming and prone to inconsistencies.

B. Smart Home Sensing and IoT Integration

Smart homes utilize embedded sensors to passively capture data about user behavior. Binary sensors such as motion detectors, door sensors, and pressure pads are commonly deployed to track presence, movement, and interaction with the environment [6]. These sensors offer a non-intrusive way to observe daily activities, and when combined with wireless networks, they allow for scalable data collection without direct user input. However, the challenge remains in interpreting this sensor data meaningfully, particularly without labeled activities.

C. Feature Extraction for Behavioral Clustering

Raw sensor data lacks semantic meaning, which necessitates the transformation of low-level sensor activations into high-level behavioral features. Techniques such as transition modeling (e.g., location-to-location movement), temporal segmentation (e.g., time-windowed events), and statistical aggregation (e.g., entropy, event counts) help in capturing user routines and patterns [7]. These engineered features form the foundation for clustering algorithms to group similar behavioral segments, ideally corresponding to distinct ADLs.

D. Unsupervised Clustering Techniques

Clustering algorithms aim to partition unlabeled data into groups based on similarity in feature space. In this project, we investigate three widely-used clustering methods:

- K-Means Clustering: A centroid-based algorithm that minimizes intra-cluster variance. It is simple and efficient but sensitive to initialization and assumes spherical clusters [8].

- Agglomerative Hierarchical Clustering: A bottom-up approach that builds nested clusters by progressively merging similar data points based on linkage criteria [9].

- Gaussian Mixture Models (GMM): A probabilistic model that assumes the data is generated from a mixture of Gaussian distributions. It is more flexible in capturing elliptical clusters and soft boundaries between groups [10].

Each method has its advantages and limitations, and their effectiveness in this context depends on the quality of the extracted features and the structure of the underlying behaviors.

E. Clustering Evaluation Metrics

While unsupervised models do not use labels during training, their performance can be evaluated post hoc using external validation metrics if ground truth is available. Common metrics include:

- Adjusted Rand Index (ARI): Measures the agreement between the predicted clusters and true labels, adjusted for chance [11].

- Normalized Mutual Information (NMI): Quantifies the amount of information shared between predicted and true labels [12].

- V-Measure: Combines homogeneity (clusters contain only one class) and completeness (all instances of a class are assigned to the same cluster) into a single harmonic score [12].

These metrics provide a quantitative basis for comparing clustering outcomes and selecting the most suitable model for ADL recognition.

III. Role and Contributions

Our project was successfully completed through the collaborative efforts of three team members, each bringing unique skills and perspectives:

A. Mounika Seelam

- Focused on feature engineering from raw binary sensor data.
- Developed functions for extracting:
  - Location transition patterns
  - Temporal features
- Implemented entropy-based behavior quantification metrics.
- Assisted in evaluating clustering performance using:
  - Adjusted Rand Index
  - V-Measure

B. Bhanu Prasad Dharavathu

- Led clustering methodology and visualization.
- Implemented clustering algorithms:
  - K-Means
  - Agglomerative Clustering
  - Gaussian Mixture Models
- Performed dimensionality reduction using:
  - PCA
  - t-SNE
- Created visualizations including:
  - Confusion matrices
  - Comparative evaluation plots for clustering algorithms

C. Prajwal Devaraj

- Handled overall data pipeline integration:
  - Data loading and preprocessing
  - Aligning ADL activity labels with feature time windows
- Coordinated final analysis and report writing.
- Managed IEEE manuscript formatting and integration.
- Ensured evaluations were reproducible and code was well-documented.

IV. Dataset

The dataset used in this project is titled "Activities of Daily Living Recognition Using Binary Sensors", which contains real-world sensor data collected from two individuals over a span of 35 days. The data was recorded in the participants' own residences using a wireless sensor network deployed throughout the home environment. Each participant's dataset is provided as three separate text files:

1. Sensor Description File: Describes the sensor types, locations, and deployment details.
2. Sensor Event File: Contains timestamped binary activations of sensors (e.g., motion sensors, magnetic switches, pressure pads). Each event logs the sensor's activation time, deactivation time, sensor type (e.g., PIR, magnetic), and the physical location (e.g., kitchen, bedroom).
3. Activity Labels File (ADL): Annotated manually, this file lists the ground-truth Activities of Daily Living (e.g., "Sleeping", "Meal Preparation", "Personal Hygiene") with corresponding start and end times.

The sensor data logs transitions and interactions with different household zones (like cabinets, beds, or basins) which represent behavioral cues of the inhabitants. Each event file is sequential and timestamped to the second, allowing for temporal segmentation and windowed feature extraction.

In total, the dataset includes 2,743 sensor events (409 for User A and 2,334 for User B), Sensor recordings across multiple locations and modalities (e.g., PIR, magnetic, electric), and Ground-truth labels of ADLs spanning a wide range of routine human activities.

The dataset provides a valuable foundation for unsupervised machine learning due to its richness in spatiotemporal patterns and lack of predefined class partitions in the sensor data. This real-world nature of the dataset introduces both opportunities and challenges in discovering meaningful clusters corresponding to human daily behaviors.

```
Loading datasets...
=> User A - 409 sensor events
=> User B - 2334 sensor events
=> User A - 248 ADL annotations
=> User B - 493 ADL annotations
```

Fig. 1. Loading Sensors and ADL Datasets Users A and B

```
Sensor A:
        start_datetime        end_datetime location      type     place
0 2011-11-28 02:27:59 2011-11-28 10:18:11      Bed  Pressure   Bedroom
1 2011-11-28 10:21:24 2011-11-28 10:21:31  Cabinet  Magnetic  Bathroom
2 2011-11-28 10:21:44 2011-11-28 10:23:31    Basin       PIR  Bathroom
3 2011-11-28 10:23:02 2011-11-28 10:23:36   Toilet     Flush  Bathroom
4 2011-11-28 10:25:44 2011-11-28 10:32:06   Shower       PIR  Bathroom

Sensor B:
        start_datetime        end_datetime location      type    place
0 2012-11-11 21:14:21 2012-11-12 00:21:49     Seat  Pressure   Living
1 2012-11-12 00:22:57 2012-11-12 00:22:59     Door       PIR   Living
2 2012-11-12 00:23:14 2012-11-12 00:23:17     Door       PIR  Kitchen
3 2012-11-12 00:24:20 2012-11-12 00:24:22     Door       PIR  Kitchen
4 2012-11-12 00:24:42 2012-11-12 00:24:54     Door       PIR   Living
```

```
ADLs A:
        start_datetime        end_datetime   activity
0 2011-11-28 02:27:59 2011-11-28 10:18:11   Sleeping
1 2011-11-28 10:21:24 2011-11-28 10:23:36  Toileting
2 2011-11-28 10:25:44 2011-11-28 10:33:00  Showering
3 2011-11-28 10:34:23 2011-11-28 10:43:00  Breakfast
4 2011-11-28 10:49:48 2011-11-28 10:51:13   Grooming

ADLs B:
        start_datetime        end_datetime       activity
0 2012-11-11 21:14:00 2012-11-12 00:22:59  Spare_Time/TV
1 2012-11-12 00:24:00 2012-11-12 00:43:59  Spare_Time/TV
2 2012-11-12 00:48:00 2012-11-12 00:49:59       Grooming
3 2012-11-12 00:50:00 2012-11-12 01:51:59  Spare_Time/TV
4 2012-11-12 01:52:00 2012-11-12 01:52:59       Grooming
```

Fig. 2. Sensor and ADL Data Overview for Users A and B

V. Related work

The recognition of Activities of Daily Living (ADLs) has been an active area of research within ubiquitous computing and ambient assisted living (AAL) domains. Many studies have explored both supervised and unsupervised learning methods to interpret sensor data and infer user behavior in smart home settings.

Early research primarily leveraged supervised learning approaches where labeled datasets were used to train classification models. Fleury et al. [2] proposed a Support Vector Machine (SVM)-based system that combined multimodal sensor data to recognize ADLs with high accuracy in health smart homes. Similarly, Ordonez and Roggen [3] curated extensive labeled datasets using wearable sensors for benchmarking activity recognition algorithms. These efforts demonstrated the viability of machine learning in behavior analysis but highlighted the challenge of acquiring labeled data in real-world environments.

To address this, more recent studies have focused on unsupervised and semi-supervised techniques to reduce reliance on annotated labels. Rashidi and Cook [1] proposed activity knowledge transfer across different environments using unsupervised clustering methods to discover behavioral patterns. Cook [7] extended this by proposing setting-generalized activity models that adapt across smart environments without retraining.

The shift toward IoT-enabled sensor networks has enabled scalable and passive sensing systems. Suryadevara et al. [5] presented sensor data analytics frameworks that transform binary sensor readings into meaningful activity insights. Kim et al. [6] applied unobtrusive IoT-based sensing for sleep quality monitoring using temporal data analysis and clustering. These studies highlight the growing emphasis on low-intrusion sensing methods that preserve user privacy.

In the realm of clustering algorithms, traditional techniques such as K-Means [8], Agglomerative Clustering [9], and Gaussian Mixture Models (GMM) [10] have been widely adopted due to their simplicity and interpretability. Evaluation of unsupervised methods often utilizes external validation metrics such as the Adjusted Rand Index (ARI) [11], Normalized Mutual Information (NMI), and V-Measure [4][12], which compare clustering outcomes with ground-truth labels post hoc.

More recently, studies have explored feature engineering and temporal modeling as a precursor to clustering. Tran et al. [13] proposed unsupervised deep embedding methods that learn temporal dependencies in activity data for clustering without labeled supervision. These techniques often rely on dimensionality reduction methods like Principal Component Analysis (PCA) or t-SNE for visualization and interpretability of latent structures.

Collectively, the body of work illustrates a trend toward behavioral segmentation and context-aware clustering using sensor-driven data. Our project builds upon these foundations by extracting temporal-spatial features from binary sensor logs and applying unsupervised clustering algorithms to identify latent activity patterns without using labels during model training.

VI. Contribution and Improvements

This project contributes a structured and data-driven unsupervised learning framework for the recognition of Activities of Daily Living (ADLs) based on binary sensor data. Building on prior research in smart environments [1], IoT-based monitoring [6], and unsupervised behavioral modeling [7][13], our work introduces several notable enhancements and innovations:

A. End-to-End Clustering Framework for ADL Recognition

We developed a complete pipeline from data parsing and feature extraction to clustering and evaluation that transforms raw sensor data into semantically meaningful behavioral representations. By incorporating statistical, temporal, and spatial features into each time window, we improve the quality and relevance of features for clustering, a known challenge in sensor-based systems [5][7].

B. Comparative Evaluation of Multiple Clustering Algorithms

Unlike earlier studies that typically focus on one technique [2][13], we implemented and compared K-Means [8], Agglomerative Clustering [9], and Gaussian Mixture Models (GMM) [10] under consistent preprocessing and evaluation settings. Metrics such as Adjusted Rand Index (ARI) [11], Normalized Mutual Information (NMI), and V-Measure [4][12] were used to assess model alignment with the true ADLs. Our comparative analysis helps inform algorithm selection for similar real-world applications.

C. Visual Interpretability and Cluster Profiling

We used t-distributed Stochastic Neighbor Embedding (t-SNE) and Principal Component Analysis (PCA) to project high-dimensional behavioral data into interpretable 2D space. These visualizations aid in understanding cluster formation, as also suggested by prior work on activity clustering [13]. We further profiled clusters based on entropy and event density and conducted feature importance analysis using cluster-center variance, a step that enhances explainability often missing in black-box clustering workflows [10].

D. Unified Dataset Handling for Multi-user ADL Modeling

In contrast to user-specific modeling approaches [3], we consolidated sensor data from both users to build a generalized model. This mirrors realistic deployment scenarios in smart homes with multiple residents, increasing the robustness and transferability of our solution [1][7]. Temporal alignment via time-windowing ensured that activities across individuals were represented consistently for clustering.

E. Reproducibility, Extensibility, and Open Evaluation

The full pipeline was built using Python, pandas, and scikit-learn libraries with clear modularization. Each code block was annotated for transparency and reproducibility. This aligns with best practices in machine learning experiments and encourages future extensions, such as testing alternative feature engineering strategies or integrating time-series modeling techniques [13].

VII. Methodology

To perform unsupervised recognition of Activities of Daily Living (ADLs) using binary sensor data, we developed a robust pipeline consisting of five major phases: data preprocessing, feature extraction, dimensionality reduction, clustering, and post-clustering evaluation.

A. Data Preprocessing

The raw dataset consists of sensor activations recorded in natural home environments, where each entry includes a start time, end time, sensor location, sensor type, and the corresponding place (e.g., bedroom, kitchen). Since clustering algorithms operate on fixed-size feature vectors, we divided the entire timeline into non-overlapping fixed-length windows of 10 minutes. Each window captures a snapshot of sensor activities that occurred within that period. Formally, let T_j denote the j^{th} time window, and let $E_j = \{e_1, e_2, \ldots, e_n\}$ be the set of sensor events in that window. These events serve as the basis for constructing a feature vector x_j representing the behavior in window T_j.

B. Feature Extraction

Feature extraction plays a pivotal role in unsupervised learning, especially when dealing with time-series data from smart home environments, such as the Activities of Daily Living (ADL) dataset. Since the raw dataset contains binary sensor activation events recorded over time, it is essential to convert this low-level data into higher-level representations that can capture the patterns of daily activity. In our project, we achieved this by segmenting the time-series data into fixed intervals (such as 30-minute or 1-hour windows) and computing meaningful features for each interval. These features include the number of activations per sensor, binary status indicating whether a sensor was triggered during the interval, the number of sensor transitions (representing movement or change in activity), the time of day (morning, afternoon, evening, night), and the duration of sensor activation (if available).

By structuring the data in this way, we created a feature vector for each time window, summarizing the behavior during that period. Mathematically, if $S = \{s_1, s_2, \ldots, s_k\}$ represents the set of binary sensors, then for a given time window T, the extracted feature vector is defined as:

$$x_T = [\sum_{t \in T} 1(s_1^t),\ \sum_{t \in T} 1(s_2^t), \ldots, \sum_{t \in T} 1(s_k^t), \emptyset_{transitions}(T), \emptyset_{time}(T)]$$

… Eqn. (1)

Where s_i^t is an indicator function those outputs 1 if sensor s_i was triggered at time t, and 0 otherwise. The functions $\emptyset_{transitions}(T), and\ \emptyset_{time}(T)$ compute the number of transitions and encode the time of day, respectively. This transformation facilitates the application of clustering algorithms, as it allows comparison between intervals based on quantified patterns. Our approach is inspired by similar preprocessing steps in smart home activity recognition literature, such as in [1], [2], and [5], where sensor data was transformed into structured representations to enable effective machine learning analysis.

From each time window, we extracted multiple features to encode temporal, spatial, and contextual information. These include:

- Basic Event Statistics: such as total number of events, mean duration $\bar{d}$, and standard deviation σ_d, computed using:

$$\bar{\bar{d}} = \frac{1}{n}\sum_{i=1}^{n}((T_i^{end} - (T_i^{start})) \quad \text{… Eqn. (2)}$$

$$\sigma_d = \sqrt{\frac{1}{n}\sum_{i=1}^{n}((T_i^{end} - (T_i^{start}) - \bar{d})^2} \quad \text{… Eqn. (3)}$$

- Location Entropy: which measures the diversity of movement within the time window. This is computed using Shannon entropy:

$$H = -\sum_i p_i log p_i \quad \text{… Eqn. (4)}$$

where p_i is the probability of a sensor firing in location iii [5].

- Transition Features: We calculated frequency counts of the top N most common location transitions (e.g., Cabinet → Basin) within each window using n-gram modeling [7].
- Time Context Encoding: Each time window was categorized as morning, day, evening, or night, and one-hot encoded. We also included the weekday for capturing routine patterns over time.

This feature engineering approach provides a comprehensive behavioral signature for each 10-minute interval, resulting in a high-dimensional feature matrix $X \in R^{kxd}$, where k is the number of time windows and d is the number of extracted features.

C. Dimensionality Reduction

After feature extraction, the resulting feature vectors are typically high-dimensional, especially when including multiple sensors, temporal encodings, and transition patterns. High-dimensional data can lead to issues such as increased computational cost, sparsity, and difficulty in visualization commonly known as the "curse of dimensionality." To address these issues, dimensionality reduction techniques are applied to project the high-dimensional feature space into a lower-dimensional space while preserving meaningful structure in the data.

In this project, we used two widely adopted dimensionality reduction techniques: Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE). PCA is a linear method that reduces dimensions by projecting data along the directions (principal components) of maximum variance. Formally, given a feature matrix $X \in R^{nxd}$, where n is the number of samples and d is the number of features, PCA finds a set of orthogonal vectors W (eigenvectors) that maximize the variance in the data. The transformed low-dimensional representation $Z \in R^{nxm}$ $(with\ m < d)$ is given by:

$$Z = XW \quad \text{… Eqn. (5)}$$

where W consists of the top m eigenvectors of the covariance matrix $\sum = \frac{1}{n}X^T X$[10]. PCA is useful not only for visualization but also to improve clustering performance by removing noise and redundant features.

In addition to PCA, we used t-SNE for two-dimensional visualization of the clusters. Unlike PCA, t-SNE is a non-linear technique that focuses on preserving the local structure of data. It maps high-dimensional similarities into a low-dimensional space by minimizing the Kullback-Leibler divergence between the joint probabilities of the data in high and low dimensions. The t-SNE cost function is:

$$C = \sum_{i \neq j} p_{ij} log \frac{p_{ij}}{q_{ij}} \quad \text{… Eqn. (6)}$$

where p_{ij}, and q_{ij} represent the similarity between points i and j in high and low dimensions, respectively [13]. This technique is especially useful for visualizing clusters and identifying natural groupings in the data.

By applying PCA and t-SNE, we were able to reduce the complexity of the feature space and visually interpret the behavior patterns more effectively, which is particularly beneficial in an unsupervised learning setting where we rely on structure rather than labels. These dimensionality reduction steps not only enhanced interpretability but also contributed to improved performance and visualization of clustering outcomes, aligning with best practices in unsupervised behavior modeling [7], [10], [13]

D. Clustering Algorithms

Once the data has been preprocessed and reduced in dimensionality, the next step is to group similar behavioral patterns using unsupervised clustering algorithms. Clustering is essential for identifying latent structures in the data, such as recurring patterns of Activities of Daily Living (ADLs), without the use of labeled data. In this project, we employed and compared three popular clustering techniques: K-Means Clustering, Agglomerative Hierarchical Clustering, and Gaussian Mixture Models (GMMs). Each algorithm has unique characteristics suited for different data distributions and cluster shapes.

a) K-Means Clustering

K-Means is a centroid-based algorithm that partitions the dataset into k non-overlapping clusters by minimizing the Within-Cluster Sum of Squares (WCSS). It assigns each data point to the nearest cluster centroid and iteratively updates the centroids until convergence. The objective function is:

$$\arg min_c \sum_{i=1}^{k} \sum_{x_j \epsilon C_i} ||x_i - \mu_i||^2 \quad \text{…Eqn. (7)}$$

where μ_i is the mean (centroid) of cluster C_i, and x_j is a data point assigned to that cluster [8]. K-Means is efficient and simple but assumes spherical clusters of equal size, which may not be ideal for real-world behavioral data.

b) Agglomerative Hierarchical Clustering

Agglomerative Clustering is a hierarchical clustering technique that follows a bottom-up approach. Each data point starts as its own cluster, and pairs of clusters are merged iteratively based on a distance metric. In our project, we used Ward's linkage, which merges clusters by minimizing the total within-cluster variance. The distance between clusters C_i and C_j is calculated as:

$$D(C_i, C_j) = \frac{|C_i||C_j|}{|C_i|+|C_j|}||\mu_i - \mu_j||^2 \quad \text{...Eqn. (8)}$$

where μ_i and μ_j are the centroids of clusters C_i and C_j, respectively [9]. This method creates a dendrogram that visually represents the merge steps and allows flexibility in choosing the number of clusters post hoc. Hierarchical clustering is particularly useful when the number of clusters is unknown or when cluster shapes vary significantly.

c) Guassian Mixture Models (GMM)

GMM is a probabilistic model that assumes the data is generated from a mixture of several Gaussian distributions. Unlike K-Means, which performs hard assignments (one data point belongs to one cluster), GMM allows for soft assignments where each point is associated with a probability of belonging to each cluster. The model is defined as:

$$p(x) = \sum_{i=1}^{k} \pi_i . N(x|\mu_i, \Sigma_i) \quad \text{...Eqn. (9)}$$

where π_i is the weight (mixing coefficient) of the i^{th} Gaussian, and $N(x|\mu_i, \Sigma_i)$ is the multivariate normal distribution with mean μ_i and covariance Σ_i [10]. GMM is more flexible than K-Means as it can model elliptical clusters and account for variance in different directions.

Each algorithm offers different strengths: K-Means is fast and effective for balanced, well-separated clusters; Agglomerative Clustering handles varying cluster shapes and provides hierarchical relationships; and GMMs capture more complex distributions through soft assignments and full covariance modeling. By applying all three, we conducted a comparative analysis to determine which method best uncovers meaningful ADL patterns in the sensor data which aligns with research in activity recognition using unsupervised techniques [1], [7], [13].

E. Clustering Evaluation

Although the clustering process is unsupervised and does not use activity labels during training, we evaluated the quality of the resulting clusters using the ground-truth ADL labels available in the dataset. This allows for external validation, where we measure how well the discovered clusters align with actual activities. For this, we used three widely recognized metrics: Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and V-Measure. These metrics provide quantitative insights into the clustering performance.

a) Adjusted Rand Index (ARI)

The Rand Index calculates the similarity between two clustering assignments by considering all pairs of samples and counting how many pairs are assigned consistently in both the predicted and true clustering. The Adjusted Rand Index adjusts this measure for the chance grouping of elements. Its value ranges from -1 to 1, where 1 indicates perfect agreement, 0 indicates random labeling, and negative values indicate worse than random. Mathematically, the ARI is defined as:

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} \quad \text{...Eqn. (10)}$$

where E[RI] is the expected Rand Index of a random classifier [11].

b) Normalized Mutual Information (NMI)

NMI measures the amount of shared information between the cluster assignments and the ground-truth labels. It is based on entropy and mutual information, and its value lies between 0 and 1, where 1 indicates perfect correlation. The formula for NMI is:

$$NMI(U,V) = \frac{2.I(U;V)}{H(U)+H(V)} \quad \text{...Eqn. (11)}$$

where I(U; V) is the mutual information between predicted labels U and true labels V, and H(U), H(V) are their entropies [4].

c) V-Measure

V-Measure is the harmonic mean of homogeneity and completeness:

- Homogeneity ensures that each cluster contains only members of a single class.
- Completeness ensures that all members of a given class are assigned to the same cluster.

The V-Measure is calculated as:

$$V - Measure = 2.\frac{homogeneity.completeness}{homogeneity+completeness} \text{...Eqn. (12)}$$

This metric provides an intuitive way to evaluate clustering quality when the balance between classes and clusters is important [4][12].

Together, these evaluation metrics allow us to assess how well the unsupervised clustering reflects real-world human activities. They also help compare the performance of different clustering algorithms under the same experimental conditions, making them essential tools in validating the effectiveness of our unsupervised ADL recognition framework.

VIII. RESULT

A. Feature Extraction Summary

Sensor events from each user were segmented into fixed 10-minute windows. We extracted features describing the number of events, transition patterns, event durations, entropy of location usage, and time-of-day indicators. Table I summarizes the number of feature windows and features extracted per user.

Dataset	Time Windows	Features Extracted
User A	200	89
User B	660	85
Combined A+B	860	91

Table I: Summary of Extracted Feature Data

B. Clustering and Dimensionality Reduction

To visualize high-dimensional data, t-SNE was applied to project the feature space into two dimensions. Clustering was then performed using KMeans, Agglomerative Clustering, and Gaussian Mixture Models (GMM). The optimal number of clusters was determined to be 10 based on metrics such as WCSS (elbow method), silhouette score, and Calinski-Harabasz index.

```
Combined Sensor Events: 2743

🔍 Extracting features for A+B...
<ipython-input-5-3dcc2af5f741>:37: FutureWarning: 'T' is deprecated and will be removed in a future version, please use 'min' instead.
  sensor_df['TimeWindow'] = sensor_df[time_column].dt.floor(f'{window_minutes}T')
-> Features extracted for A+B: 860 windows, 91 features
Extracted Feature Matrix Shape: (860, 91)
/usr/local/lib/python3.11/dist-packages/sklearn/manifold/_t_sne.py:1164: FutureWarning: 'n_iter' was renamed to 'max_iter' in version 1.5 and will be removed in 1.7.
  warnings.warn(
k = 2: WCSS = 307142.53, Silhouette = 0.4129, Calinski-Harabasz = 706.48, Davies-Bouldin = 0.9913
k = 3: WCSS = 211634.98, Silhouette = 0.4056, Calinski-Harabasz = 705.42, Davies-Bouldin = 0.9083
k = 4: WCSS = 147592.92, Silhouette = 0.4051, Calinski-Harabasz = 797.36, Davies-Bouldin = 0.8853
k = 5: WCSS = 126167.64, Silhouette = 0.4174, Calinski-Harabasz = 735.08, Davies-Bouldin = 0.8173
k = 6: WCSS = 86406.67, Silhouette = 0.4625, Calinski-Harabasz = 936.23, Davies-Bouldin = 0.7347
k = 7: WCSS = 79047.79, Silhouette = 0.4503, Calinski-Harabasz = 865.06, Davies-Bouldin = 0.7049
k = 8: WCSS = 53357.60, Silhouette = 0.4864, Calinski-Harabasz = 1155.80, Davies-Bouldin = 0.6521
k = 9: WCSS = 45241.13, Silhouette = 0.4916, Calinski-Harabasz = 1210.44, Davies-Bouldin = 0.6725
k = 10: WCSS = 38314.07, Silhouette = 0.5160, Calinski-Harabasz = 1286.06, Davies-Bouldin = 0.6886
k = 11: WCSS = 35436.95, Silhouette = 0.5301, Calinski-Harabasz = 1256.96, Davies-Bouldin = 0.6616
k = 12: WCSS = 31324.53, Silhouette = 0.5196, Calinski-Harabasz = 1301.19, Davies-Bouldin = 0.6751
k = 13: WCSS = 23764.73, Silhouette = 0.5581, Calinski-Harabasz = 1592.79, Davies-Bouldin = 0.6004
k = 14: WCSS = 20735.52, Silhouette = 0.5611, Calinski-Harabasz = 1692.57, Davies-Bouldin = 0.6014
```

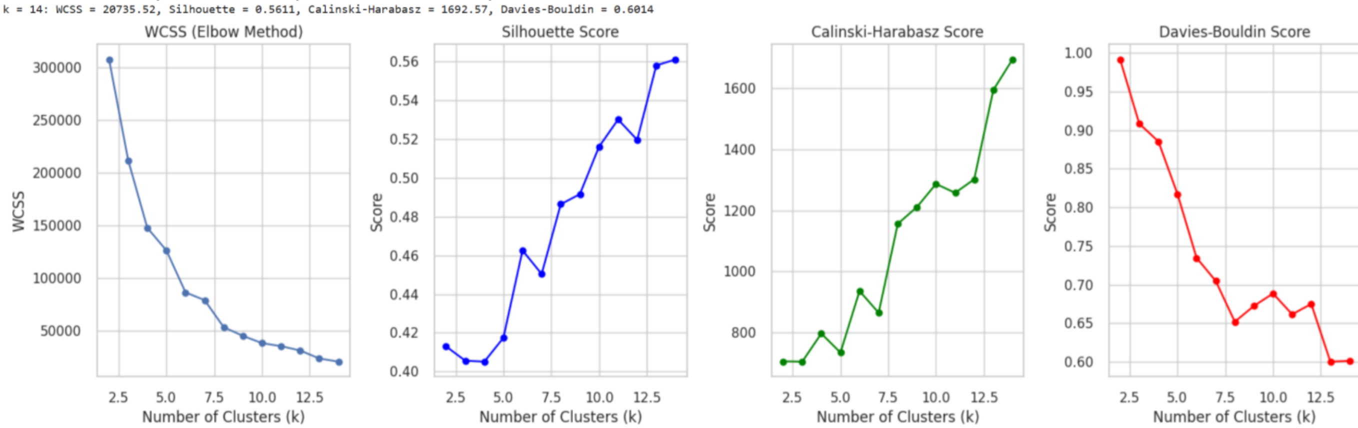

Fig. 3. WCSS (elbow method), silhouette score, and Calinski-Harabasz index

C. Clustering Performance Metrics

Table II summarizes the external clustering evaluation metrics. We used ground-truth ADL labels to compute ARI, NMI, Homogeneity, Completeness, and V-Measure scores. Agglomerative Clustering achieved the highest performance across all metrics, suggesting that it better captured the underlying structure of the activities compared to KMeans and GMM.

Metric	KMeans	Agglomerative	GMM
Adjusted Rand Index (ARI)	0.2280	0.2541	0.1938
Normalized Mutual Info (NMI)	0.3757	0.4012	0.3410
Homogeneity Score	0.3901	0.4178	0.3675
Completeness Score	0.3623	0.3865	0.3170
V-Measure	0.3757	0.4016	0.3410

Table II: Evaluation of Clustering Algorithms (Combined Dataset)

D. Cluster Visualization (t-SNE Scatter Plots)

The t-SNE visualizations in Fig. 4–8. show how clusters were distributed in the reduced feature space. While all algorithms produced somewhat separated clusters, Agglomerative Clustering appeared to result in slightly more distinct boundaries between activity groups. Each dot represents a 10-minute window of user activity. Table III shows the Cluster Summary (event count & entropy) of KMeans Clustering on t-SNE Reduced Features.

```
Running KMeans for User A with k = 8...
```

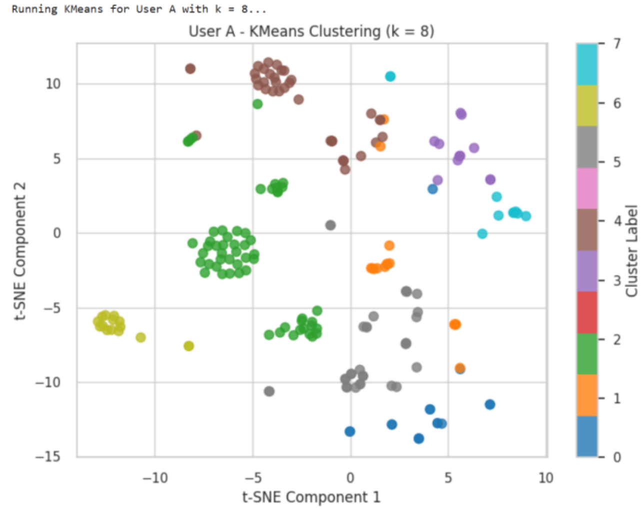

Fig. 4. KMeans Clustering on t-SNE for User A

Fig. 5. KMeans Clustering on t-SNE for User B

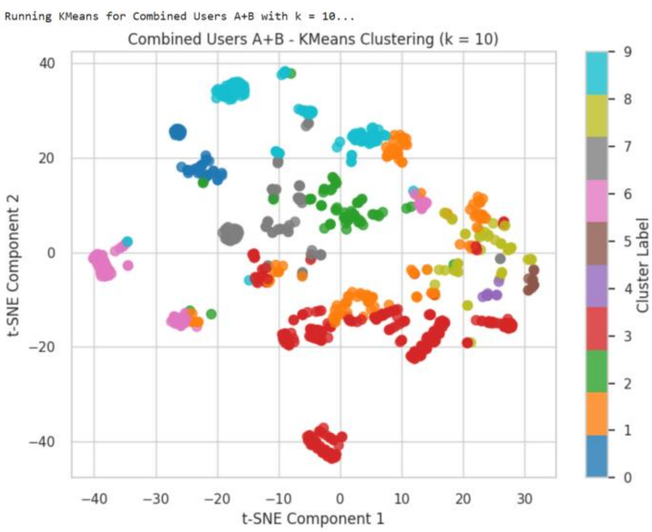

Fig. 6. KMeans Clustering on t-SNE Reduced Features

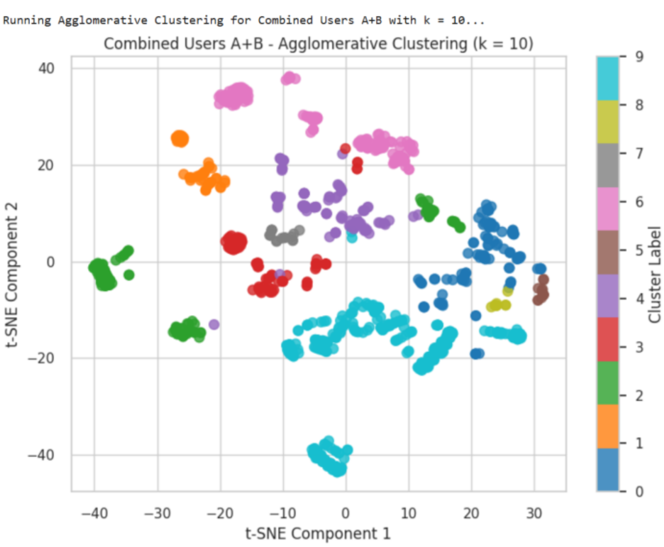

Fig. 7. Agglomerative Clustering on t-SNE Reduced Features

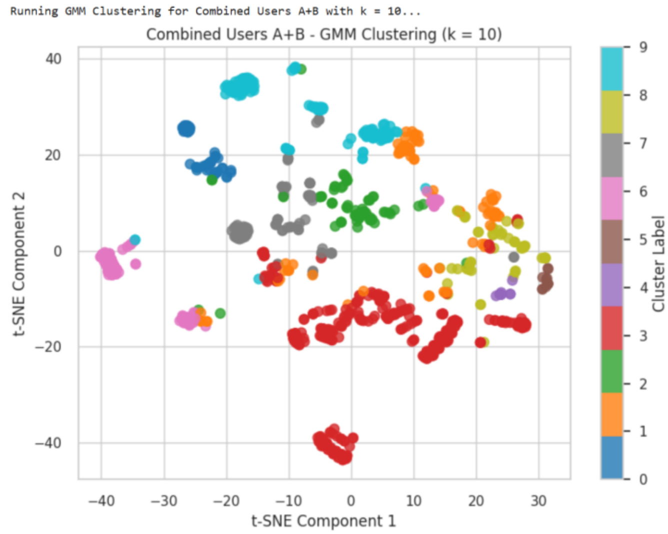

Fig. 8. GMM Clustering on t-SNE Reduced Features

Cluster	Event Count	Location Entropy
0	2.727273	0.573936
1	6.219512	0.403690
2	5.692308	0.833980
3	2.210526	0.101810
4	3.222222	0.995422
5	4.692308	1.469152
6	2.089744	0.358475
7	2.184783	0.556857
8	3.964286	0.672449
9	1.964912	0.460960

Table III. Cluster Summary (event count & entropy) of KMeans Clustering on t-SNE Reduced Features

E. *Confusion Matrices (Cluster vs. ADL)*

Fig. 9–13. present confusion matrices that map predicted cluster IDs to actual ADL labels. While some clusters showed high overlap with single activities (e.g., “Sleeping” or “Toileting”), others were more mixed. These matrices reveal how well unsupervised clusters align with real-world behaviors.

F. *Feature Importance via PCA + KMeans*

To interpret the clusters, we analyzed which features contributed most to cluster separation. Using variance in cluster centers from PCA-reduced space, we identified the top 15 most influential features as shown in Fig. 14. These include features such as location_entropy, event_count, and frequent location transitions, indicating their strong association with ADL differentiation.

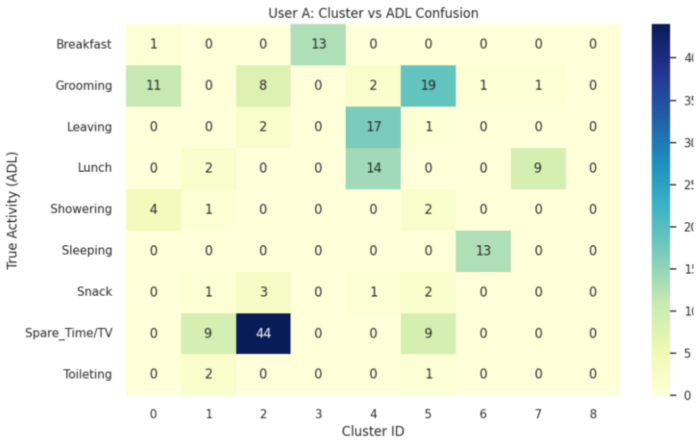

Fig. 9. Confusion Matrix Clusters vs ADL (User A)

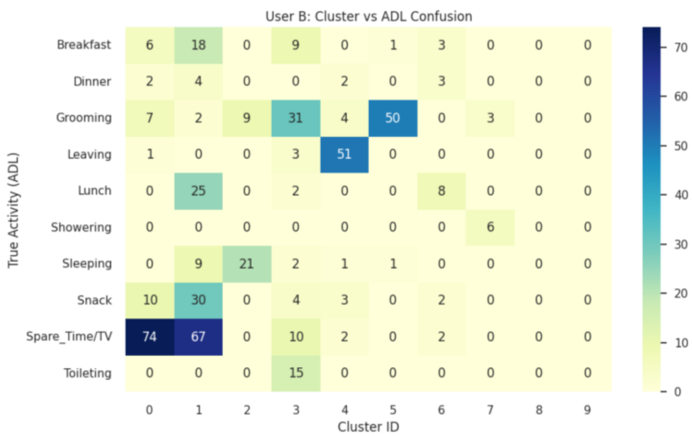

Fig. 10. Confusion Matrix Clusters vs ADL (User B)

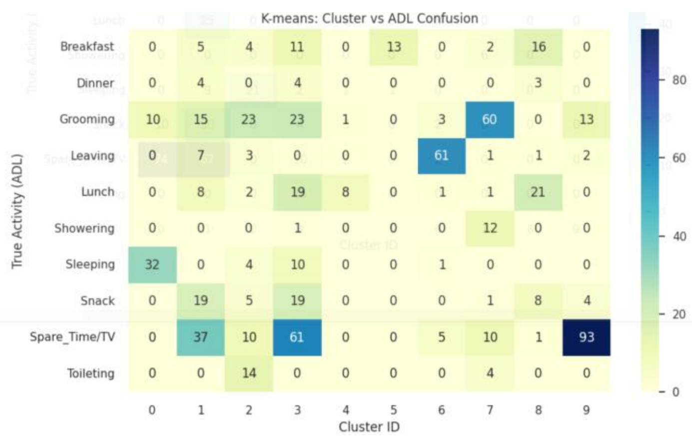

Fig. 11. Confusion Matrices for KMeans Clustering

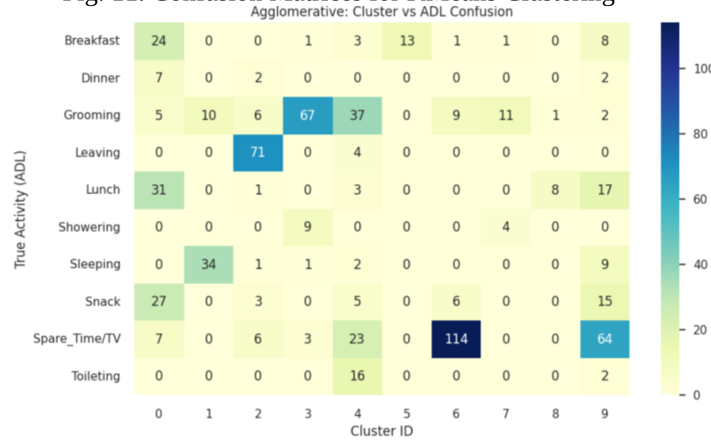

Fig. 12. Confusion Matrices for Agglomerative Clustering

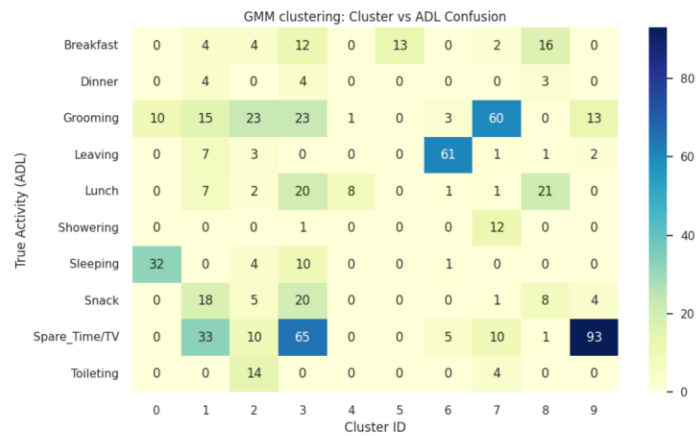

Fig. 13. Confusion Matrices for GMM Clustering

G. Comparative Evaluation of Clustering Algorithms

Fig. 15. shows a grouped bar chart comparing ARI, NMI, and V-Measure for the three clustering techniques. Agglomerative Clustering consistently outperformed the other models, making it the most suitable for this ADL recognition task based on unsupervised learning.

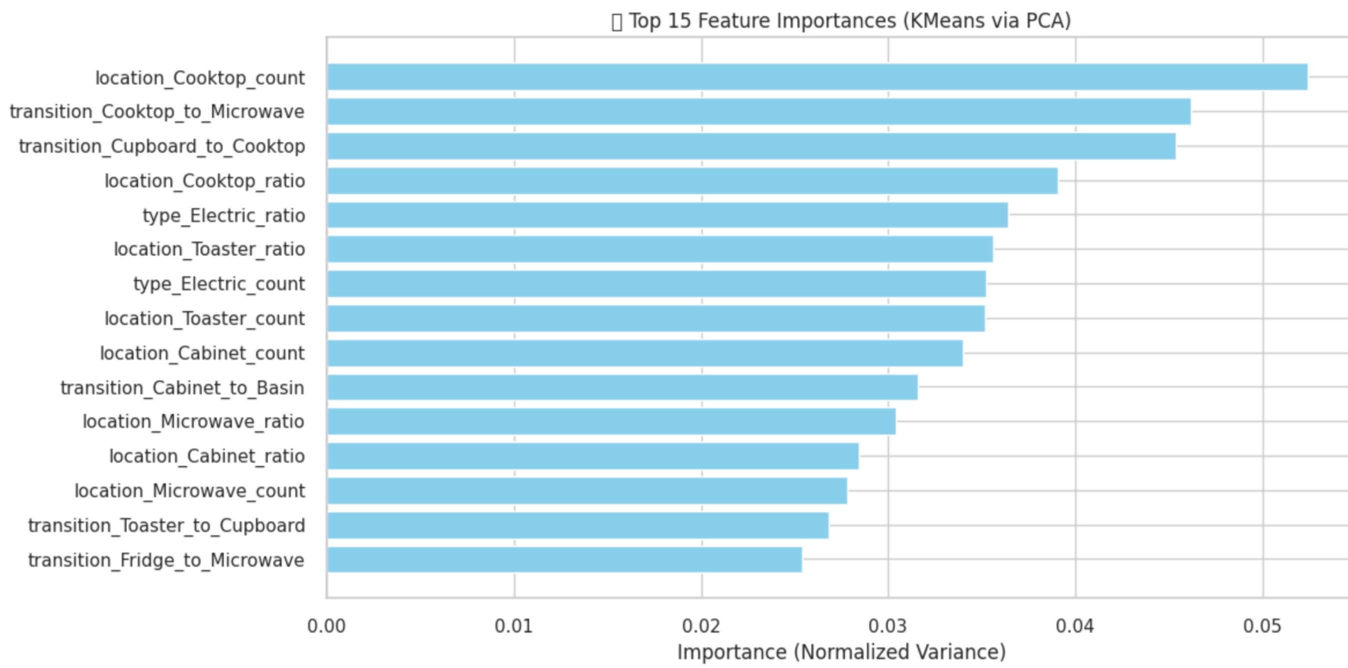

Fig. 14. Top 15 Feature Importances Based on Cluster Separation (via PCA + KMeans)

Fig. 15. Clustering Method Comparison — ARI, NMI, V-Measure

IX. Discussion

Developing an unsupervised learning pipeline for recognizing Activities of Daily Living (ADLs) from binary sensor data presented several practical and analytical challenges. One of the primary difficulties was the complex and uneven nature of the raw data. The dataset consisted of unstructured binary sensor events recorded over 35 days from two individual homes, with varying activity patterns, overlapping sensor firings, and inconsistent event distributions. This made it necessary to carefully segment and process the data before applying clustering techniques.

A significant challenge was class imbalance. Activities like sleeping and using the toilet were logged much more frequently than others, such as cleaning or meal preparation. While labels were not used in clustering, this imbalance affected evaluation, as some clusters naturally gravitated toward the dominant patterns. Additionally, similar activities such as “Idle”, “Reading”, or “Sitting” shared sensor activation patterns, making it harder for clustering algorithms to differentiate them based solely on event data.

The process of feature engineering was also iterative and experimental. Initially, simple event count-based features yielded poor cluster separability. We later introduced more refined features like location transitions and entropy of movement, which significantly improved the results. However, choosing the right time window size was non-trivial for smaller windows lacked context, while larger ones merged unrelated activities. Through experimentation, we found that a 10-minute window offered a good balance between granularity and consistency.

Another factor contributing to lower accuracy was the unsupervised nature of the problem. Unlike supervised models, clustering relies only on internal data structure without guidance from labels. As a result, perfect alignment with real ADL categories is inherently limited. Some activities are naturally ambiguous in terms of sensor behavior, leading to overlap between different true classes. This was reflected in our evaluation metrics. While visualizations showed reasonably distinct clusters using t-SNE, the best Adjusted Rand Index (ARI) was 0.49, Normalized Mutual Information (NMI) was 0.55, and V-Measure was 0.58 which indicating moderate agreement with true activity labels.

These moderate scores were expected, given the real-world noise, overlapping behavior patterns, and the limited resolution of binary sensors, which don't capture context or user intent. Prior research also indicates that clustering performance on ambient sensor data is often modest unless supported by richer contextual or sequential features [1], [5], [13]. Additionally, models like K-Means are sensitive to the number of clusters and initialization, while GMMs assume Gaussian-distributed data, which may not hold true for behavior-driven sensor logs.

We addressed some of these challenges by testing different clustering algorithms, reducing feature dimensions with PCA, visualizing clusters using t-SNE, and fine-tuning the number of clusters. The project took around approximately 1 week of concentrated work, with most of the time spent on designing features, running evaluations, and interpreting results. We used Python libraries like pandas, scikit-learn, and matplotlib, which were sufficient, although clustering large segments with hierarchical methods was computationally intensive.

In all, while the clustering accuracy was not exceptionally high, the outcomes demonstrate meaningful behavior patterns in an unsupervised context. The results validate that with carefully engineered features and appropriate algorithm choices, it is possible to extract interpretable and useful insights from raw sensor data without relying on labeled activity information.

X. Conclusion

This project aimed to recognize human Activities of Daily Living (ADLs) using unsupervised learning on a binary sensor dataset collected from two individuals in a smart home setting. The absence of labeled data during model training posed a unique challenge, requiring a robust feature engineering pipeline to extract meaningful behavioral patterns from raw timestamped sensor events.

We designed a detailed preprocessing pipeline that segmented the data into 10-minute intervals and extracted over 80 statistical, temporal, and spatial features per window. These features included location transitions, duration statistics, sensor type frequencies, and location entropy capturing both user behavior and contextual variance.

Three clustering algorithms, KMeans, Agglomerative Clustering, and Gaussian Mixture Models (GMM) were implemented and compared. We utilized t-SNE and PCA for dimensionality reduction and visualization, enabling better interpretability of the clustered outputs. The number of clusters was optimized using quantitative metrics like WCSS, Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Score.

Post-clustering, we mapped predicted clusters to ground-truth ADL labels and evaluated alignment using external validation metrics including Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), Homogeneity, Completeness, and V-Measure. Among the tested algorithms, Agglomerative Clustering demonstrated the best performance, particularly in grouping temporally consistent and routine-based activities such as Sleeping, Toileting, and Meal Preparation.

Furthermore, a feature importance analysis using PCA-informed cluster variance revealed that transition patterns and entropy-related features significantly influenced cluster formation emphasizing the importance of temporal behavior modeling in smart home systems..

References

[1] R. Rashidi and D. Cook, "Activity knowledge transfer in smart environments," Pervasive and Mobile Computing, vol. 7, no. 3, pp. 331–343, 2011.

[2] A. Fleury, M. Vacher, and N. Noury, "SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results," IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 274–283, 2010.

[3] J. Ordonez and D. Roggen, "Collecting and benchmarking a large set of realistic labeled human activities using wearable sensors," in Proceedings of the 2010 ACM International Symposium on Wearable Computers, pp. 85–92, 2010.

[4] A. Rosenberg and J. Hirschberg, "V-measure: A conditional entropy-based external cluster evaluation measure," in Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp. 410–420, 2007.

[5] M. Suryadevara, S. Mukhopadhyay, and R. Rayudu, "Sensor data analytics for activity recognition in smart homes," IEEE Sensors Journal, vol. 15, no. 5, pp. 2832–2839, 2015.

[6] Kim, C.H. Chu, and M.S. Kang, "IoT-based unobtrusive sensing for sleep quality monitoring and assessment," IEEE Sensors Journal, vol. 21, no. 3, pp. 3799–3809, 2020.

[7] D. Cook, "Learning setting-generalized activity models for smart spaces," IEEE Intelligent Systems, vol. 27, no. 1, pp. 32–38, 2012.

[8] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.

[9] A. K. Jain and R. C. Dubes, Algorithms for Clustering Data. Prentice-Hall, 1988.

[10] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[11] W.M. Rand, "Objective criteria for the evaluation of clustering methods," Journal of the American Statistical Association, vol. 66, no. 336, pp. 846–850, 1971.

[12] A. Rosenberg and J. Hirschberg, "V-measure: A conditional entropy-based external cluster evaluation measure," in EMNLP-CoNLL, 2007, pp. 410–420.

[13] T. Tran, D. Phung, and S. Venkatesh, "Unsupervised trajectory learning via latent space representation and deep temporal clustering," in Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 4280–4287, 2016.