iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.70417

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Graph and Similarity-Based Approaches to On Campus Roommate Matching

A. Gurudev, M. Vanistelroy Dobgima, H. Raza Khan, E. Rudasunikwa, A. Kumar, T. Singh Walia
Lovely Professional University, Phagwara, Punjab, India, 144411

Abstract: Having the right roommate could be a game changer in a student's university life by influencing academic achievements, health, and social activities. Roommate assignments are often carried out randomly or sometimes based on arrival time which can often result in suboptimal pairings. To address this, we suggest a machine learning solution that automates roommate assignment based on similarities shared between individuals in on-campus accommodation. The model implements similarity-based roommate matching, using K-Nearest Neighbours algorithm and a hybrid graph model of Louvain and Spectral Clustering, which considers students' preferences, behaviours and traits to determine compatibility with the K-Nearest Neighbours algorithm performing better and achieving similarity scores of 0.72. The target users for which the outcomes of this study are most useful are universities, their student accommodation departments
Keywords: On-campus Roommate Matching, K-Nearest Neighbors (KNN), Graph-Based Clustering, Louvain Algorithm, Spectral Clustering, Student Housing, Compatibility Prediction, On-campus Accommodation, Machine Learning.

I. INTRODUCTION

It is time for college, and you move to a new city, state, region or even country to pursue university studies. Settling in is never easy, but one thing that could certainly make the process better is meeting people who share similarities with you. There could not be better people to share similarities with you than your roommates if you opt for a common on-campus shared space. While a student's academic performance may not be heavily affected by their roommates(McEwan and Soderberg, 2006) unless the students spend multiple years together and not just a semester(Cao et al., 2024), a student's social life, and even mental health can benefit from comfortable living arrangements within university housing. Unfortunately, the random or traditional methods authorities use in selecting roommates often result in unsatisfactory matches and conflict amongst roommates and are outperformed by modern statistical methods(Adeniyi et al., 2024). This study attempts to overcome these limitations and propose improvements using Machine Learning algorithms for automated structured roommate matching. The system will improve compatibility in on-campus housing by evaluating the student's behavioural patterns, lifestyle choices, and personality traits.

A. Research Gaps

Most organizations in this space either utilize randomized assignment methods or rely on self-selection models. Where the goal is to foster cross-race pairings, randomized assignments work well(Albuja et al., 2024). If the goal is to make the students comfortable, self-selection models are great. However, for students who are new, solo, or would like to meet new people like them, there is need for a better approach. The currently available roommate search solutions only rely on personality types and do not utilize machine learning models that could analyze multiple compatibility factors like sleeping schedules, social activities, and even financial responsibility like we do in this study. Also, they build application solutions that students can use in off-campus housing. Moreover, this problem goes beyond the lack of accurate datasets as standard datasets for studying roommate compatibility do not exist, making it difficult to perform any research in the area. To meet this challenge, we constructed a test dataset that contains various lifestyles and personality traits to enable precise and scalable roommate matching. Amongst the goals of this study, is an attempt to address these gaps by developing two machine learning powered roommate matching systems for colleges.

II. LITERATURE REVIEW

We would like to give credit to the numerous studies conducted prior to this. The work they did was the building block for this study.
Aditi Gupta et al. 2022 (Gupta et al., 2022) - A personality matching application did an incredible work to build a roommate matching application using K-means clustering algorithm. They used a dataset of collected personality traits of individuals to build a platform that would allow individuals find potential roommates.

Aditi Bornare et al. 2023 did a study “Troomate - Finding a perfect roommate”(Bornare et al., 2023) . Using filters like social traits, diet habits and sleeping schedules, they employed algorithms like Gale – Shapley, Elo rating score and clustering to match roommates.

Rahman & Manoj Kumar (2021), “Optimal Room and Roommate Matching System Using Nearest Neighbours Algorithm With Cosine Similarity Distribution”(Rahman, n.d.). They build a system where users could be able to apply filters like gender, location, amenities and find those with which they share the closest similarities. This solution is great in off campus housing and job transfer situations. The above mentioned studies did an excellent job in pairing individuals up in an off-campus setting. This study does not tackle this problem. It is rather centred around on-campus housing and how colleges could improve their already existing room allocation systems.

Abhishek Sharma &Amandeep Kaur (2021), “Hostel’s Room Allocation System: A framework Using Single-Layer Fuzzy Logic”(Sharma and Kaur, n.d.). This study used the input parameters Citizenship, State, Religion, Temperament, Personality and Time to study which passed through a single layer fuzzy system to get allocated rooms for students.

K. Zahran et al. 2024, "Autoencoder-Enhanced Roommate Recommendation System"(Zahran et al., 2024), the authors introduced a new method for enhancing roommate matching using autoencoder neural networks. The system gathered information about students' likes and dislikes and their behaviours, which were fed into an autoencoder to learn compressed latent features of these characteristics. By identifying the hidden patterns in the students' answers, the system can better gauge compatibility among potential roommates and provide more compatible living situations as well as higher student satisfaction.

The contributions made by these authors are truly remarkable, they provided different approaches to tackle the problem of finding similar roommates. We provide yet another approach that combines both clustering (a common technique used in the aforementioned studies) and KNN-based similarity matching. This approach provides two algorithm and directs colleges to implementing in their already existing systems. This makes our contribution simple, yet unique

III. PROPOSED METHODOLOGY

This section describes the methods used to develop this on-campus roommate assignment system. It would cover aspects of the dataset, preprocessing the dataset, machine learning algorithms used and implementation of the system. Figure 1 demonstrates the working of the two proposed approaches in this study.

A. Data Collection

Given the nature of the problem, with data being key to building machine learning models, readily available datasets were not available. So, an artificial dataset comprising of students and their characteristics, behavioural patterns and preferences. The variables used in this study are easily available to on-campus residential authorities.

The variables included: Student ID, Demographic Attributes, Behavioural Preferences, and Lifestyle variables.

B. Data Variables Preprocessing

By the end of this process, the twenty-four initial variables became sixty-four. The techniques we used included: Binary encoding, One-Hot encoding, Ordinal encoding, normalization.

C. Model One: K-Nearest Neighbour Based Incremental Matching Model

1) Algorithm

- Step 1: Start.
- Step 2:
  Get a dataset containing n students. Each student is represented as a feature vector . $\overrightarrow{x_i} \in \mathbb{R}^d$
- Step 3:
  Compute pairwise similarity. $Similarity(i,j) = \frac{1}{1+\|\vec{x}_i - \vec{x}_j\|_2}$
  Store similarity scores in a matrix. $S \in \mathbb{R}^{n \times n}$
- Step 4:
  Construct Similarity Graph
  - Undirected graph G, where each node is a student. $G = (V, E)$
  - Assign the similarity values as the edge weight.

- Step 5:
  Match students into rooms. For each student *i*, find their *k* most similar neighbours based on similarity scores.. Form groups up to 4 and assign room ID.
- Step 6:
  Handle remaining students. Either assign a new room or match with their next most similar neighbours whose room is below capacity.
- Step 7:
  Return an updated dataset with assigned rooms.
- Step 8: End

2) Model Rationale:

For matching to take place, there needed to be a pool of students. This pool of students could be a pool of all the students or just part of the students. In rare cases would universities have a pool of all students prior to students lodging in. In practice, the pool increases gradually through the semester as new students lodge in. As such, the model works by creating an initial pool of students (the early comers), this could be sixty percent or any portion of the residential facility's capacity. Once this initial pool is ready, new incoming students would be matched with the existing pool of students.

D. Model Two: Hybrid Louvain-Spectral Graph-Based Model

1) Algorithm

- Step 1: Start
- Step 2: Construct Similarity Grap G. $G = (V, E)$
- Step 3: Apply Louvain Community Detection. Use the graph G to identify communities. $C = \{C_1, C_2, \dots, C_k\}$ maximize modularity: $Q = \frac{1}{2m} \sum_{i,j} [A_{i,j} - \frac{K_i K_j}{2m}] \delta(c_i, C_j)$
  where: A_{ij} is the similarity between nodes i and j. k_i and k_j are node degrees. $m = \frac{1}{2} \sum_{i,j} A_{ij}$. $\delta(c_i, c_j) = 1\ if\ i\ and\ j\ are\ the\ same\ community$
- Step 4:
  Check Community Sizes. For each community $C \in C$:
  - If $|C| \leq r$, assign all members to the same room.
  - If $|C| > r$, proceed to step 5.
- Step 5:
  Spectral clustering withing large communities.
  For communities where $|C| > r$:
  - Extract the subgraph and construct its adjacency matrix A:

$$A_{ij} = \begin{cases} Similarity_{i,j}\ if (i,j) \in E_C \\ 0\ otherwise \end{cases}$$

  - Compute the degree matrix D and the Laplacian $L = D - A$
  - Use eigen decomposition on L_{norm} and select the bottom K eigenvectors
  - Apply k-means clustering on the eigenvectors to form smaller subgroups.
- Step 6: Assign room numbers
- Step 7: End

2) Model Rationale

The KNN approach (Peterson, 2009) was one way to get the job done. However, it was not enough to just know that it got the job done without any comparative model to measure its performance against. Then comes a second model, a hybrid combination of graph-based clustering algorithms the Louvain and Spectral algorithms.

3) Implementation Details

In a nutshell, the first algorithm here, The Louvain Algorithm is first applied to detect communities. For any of the detected communities exceeding four members, Spectral clustering is employed to split the community into smaller, balanced clusters. As such, the resulting room assignments are then merged back with the original dataset. Finally, the updated dataset is sorted by Room_Number for clarity and operational use.

Actual coding implementation of this models can be found on the data science platform Google Colab at the address Link .

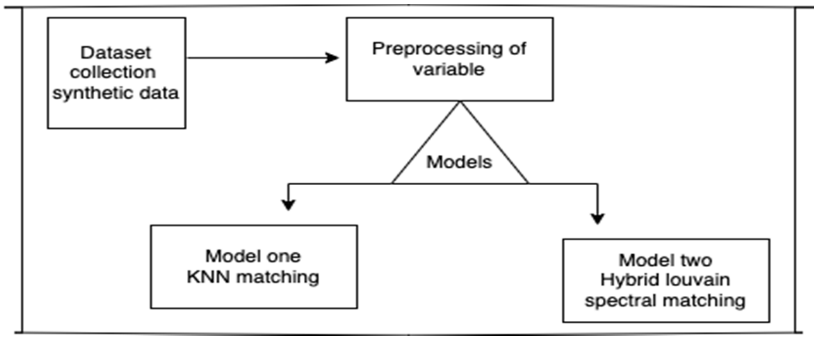


Figure 1 - Working of proposed methodology

IV. RESULT AND ANALYSIS

This section presents the results and explains the findings of this research. After developing the models, the results obtained were analysed from two perspectives: Quantitative and Qualitative. Table 1 summarises the results obtained by these models.

Table 1. Model performance scores.

Metric	Model	Description	Value
Matching Accuracy	KNN-Based	Average Similarity Score	0.72
	Hybrid Louvain-Spectral	Silhouette Score (Cluster Cohesion)	0.0248
		Davies-Boulding index	1.8446
Room Utilization	KNN-Based	Percentage of rooms that reached max occupancy (4 per room)	72%
	Hybrid Louvain and Spectral		61%

A. *Quantitative Results*

1) *Matching Accuracy*

For the KNN-based incremental matching model, we computed similarity scores that measure how closely matched room mates were (the distance between their attribute). An average of these similarity scores tells how close room mates are on average. The average similarity score turned out to be 0.72.

For the hybrid Louvain-Spectral model, we measured the clustering quality using the metrics: Silhouette Score and Davies-Bouldin Index. Silhouette Score measures the cohesiveness of clusters (i.e. room assignments). The higher the score, the better the separation between different rooms. Lower values of the Davies-Boulding Index suggest that the clusters are more distinct. We had a Silhouette score of 0.0248 and a Davies-Bouldin Index of 1.8446. Consider Fig. 2 which shows how the broad clusters were created using the Louvain algorithm but with a reduced dimensionality for visualization purposes.

2) *Room Utilization*

This performance metric measured the percentage of rooms that reached their maximum occupancy (of 4 students per room in this case). In addition to wanting students to be matched up with their closest possible peer, a college would also want their residential facilities efficiently filled up. We got a room utilisation percentage of 72%.

$$Room\ Utilization\ =\ Number\ of\ fully\ occupied\ rooms/Total\ Number\ of\ rooms\ \times\ 100\%$$

B. Qualitative Evaluations

Manual inspection (domain expertise) of the roommate assignments (from our synthetic dataset) confirmed that students within the same room shared similar attributes. This suggested that the models in this approach were effective in enhancing roommate compatibility. This was by far the most satisfying evaluation metric, given that we are college students, seeing the matches that were made based on student attributes was truly fulfilling

V. DISCUSSION

In this study, we have developed a machine learning driven system for roommate assignment in on-campus housing that uses both traditional similarity-based techniques and advanced graph clustering methods. Our contribution to this field helps solve the problem of roommate compatibility, where students are matched with roommate that they share similarities with. Recall that in the methodology section under the working of the hybrid Louvain and spectral algorithm, the process of matching was ran through two algorithms, first Louvain and then Spectral respectively. However, it did not always return assignments of a maximum of four students. In rare cases, assignments went beyond four, the maximum number of students per room decided to be used in this study. For this reason, the KNN algorithm is preferred over the hybrid Louvain and Spectral algorithm.

REFERENCES

[1] Adeniyi, O.J., Adekola, O.D., Akwaronwu, B.G., Abiodun, A.G., Eweoya, I.O., 2024. Exploring the Link Between Roommate Compatibility and Academic Outcomes: A Systematic Review of Personality-Based Matching Systems. Asian Journal of Computer Science and Technology 13, 29–40. https://doi.org/10.70112/AJCST-2024.13.2.4275

[2] Albuja, A.F., Gaither, S.E., Sanchez, D.T., Nixon, J., 2024. Testing intergroup contact theory through a natural experiment of randomized college roommate assignments in the United States. J Pers Soc Psychol 127. https://doi.org/10.1037/PSPA0000393

[3] Bornare, A., Dubey, A., Dherange, R., Chiddarwar, S., Deshpande, P., 2023. Troomate—Finding a Perfect Roommate a Literature Survey. Lecture Notes in Networks and Systems 662 LNNS, 3–17. https://doi.org/10.1007/978-981-99-1414-2_1

[4] Cao, Y., Zhou, T., Gao, J., 2024. Heterogeneous peer effects of college roommates on academic performance. Nature Communications 2024 15:1 15, 1–11. https://doi.org/10.1038/s41467-024-49228-7

[5] Gupta, A., Almeida, I., Balaji, H., Tiwari, M., 2022. DORMMATE-A Room-Mate Personality Matching Application. 2022 2nd International Conference on Computer Science, Engineering and Applications, ICCSEA 2022. https://doi.org/10.1109/ICCSEA54677.2022.9936173

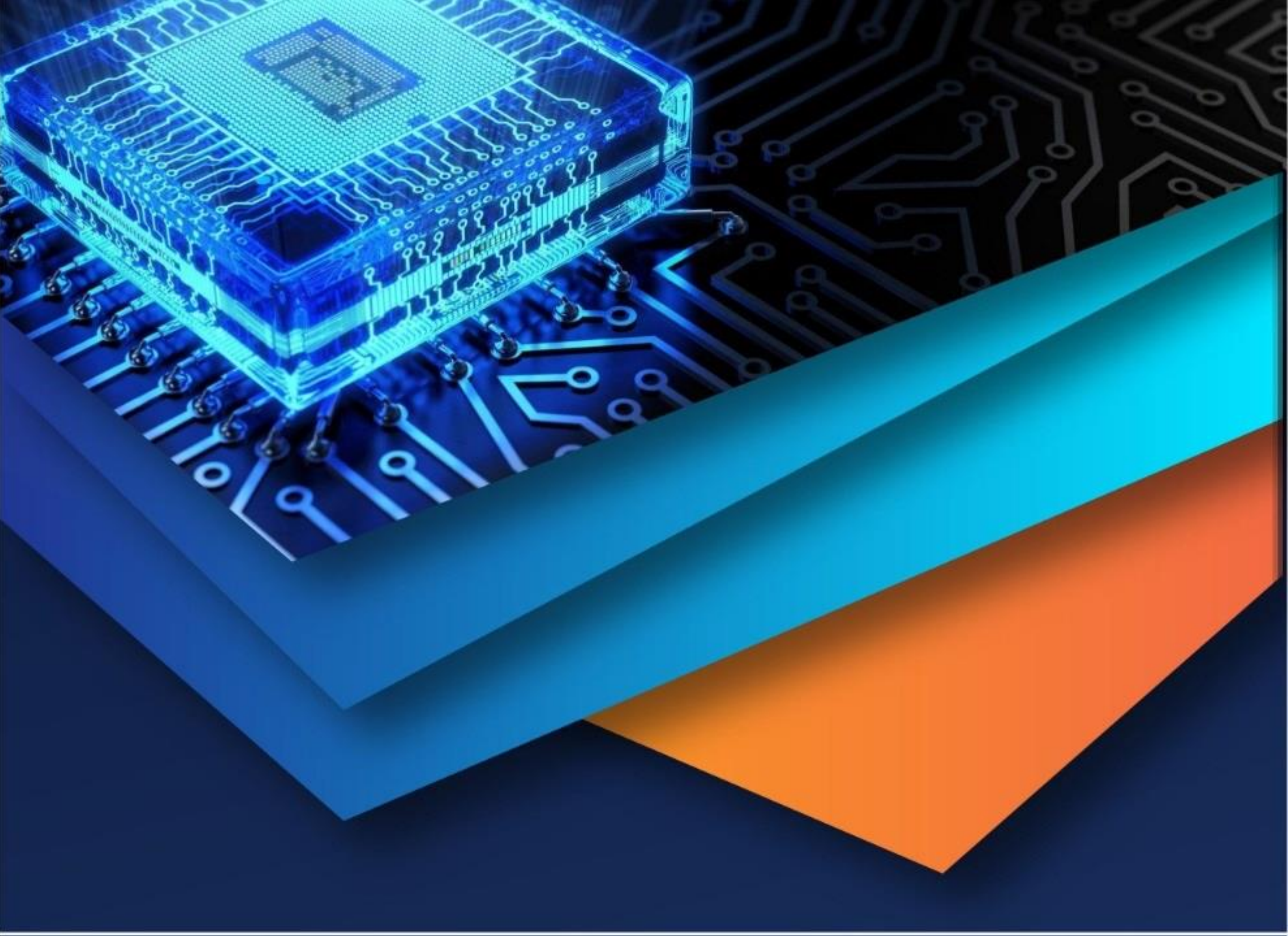
[6] McEwan, P.J., Soderberg, K.A., 2006. Roommate effects on grades: Evidence from first-year housing assignments. Res High Educ 47, 347–370. https://doi.org/10.1007/S11162-005-9392-2/METRICS

[7] Peterson, L., 2009. K-nearest neighbor. Scholarpedia 4, 1883. https://doi.org/10.4249/SCHOLARPEDIA.1883

[8] Rahman, S., n.d. Optimal Room and Roommate Matching System Using Nearest Neighbours Algorithm with Cosine Similarity Distribution.

[9] Sharma, A., Kaur, A., n.d. International Conference on Innovative Computing and Communication Hostel's Room Allocation System: A framework using single-layer fuzzy logic.

[10] Zahran, K., Amir, A., Elbanhawy, M., Ghanem, I., El-Ghamry, A., Fouad, K.M., Moawad, I.F., 2024. Autoencoder-Enhanced Roommate Recommendation System. NILES 2024 - 6th Novel Intelligent and Leading Emerging Sciences Conference, Proceedings 367–372. https://doi.org/10.1109/NILES63360.2024.10753185.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY