# **📑 Machine Learning Project Report**

## **Title**

**The Friendship Blueprint of SNU: Clustering Students Based on Clubs, Hobbies, and Teamwork Preferences**

## **Author**

**Debasish Ghosh**Year 3, Section 09  
 B.Tech CSE (AI-ML)

## **Abstract**

Friendship and collaboration are essential for fostering student well-being and campus culture at **Student National University (SNU)**. However, planning seating arrangements and activities that maximize peer bonding is challenging in diverse student groups. This study leverages **unsupervised machine learning** techniques to identify **friendship clusters** based on students’ club memberships, hobbies, and teamwork preferences. Using **Agglomerative Clustering** with encoded interest vectors, we discovered **five meaningful friendship groups** that reflect natural patterns of collaboration. Although silhouette scores were relatively low due to overlapping interests, the results provide actionable insights for **event planning, seating design, and club management** at SNU.

## **1. Introduction**

Student life at SNU is shaped by clubs, hobbies, and collaboration preferences. Understanding how these factors influence friendships can help the **Student Union** design better seating layouts in common rooms, encourage collaborations, and strengthen student communities.

### **Problem Statement**

* Students share multiple hobbies and club memberships.
* Seating arrangements in **common rooms** are currently random, which may not maximize friendship opportunities.
* The Student Union seeks to identify **natural student clusters** to guide activity planning and seating.

**Research Question:** 👉 Can clustering students by hobbies, clubs, and teamwork preferences reveal meaningful friendship groups?

## **2. Dataset and Preprocessing**

### **2.1 Dataset Features**

* **club\_top1, club\_top2** → Primary & secondary clubs
* **hobby\_top1, hobby\_top2** → Primary & secondary hobbies
* **teamwork\_preference** → Scale from 1 (solo) to 5 (team-oriented)

### **2.2 Data Cleaning**

* Standardized column names (lowercase, underscores, removed symbols).
* Normalized text entries (lowercased, stripped whitespace).
* Replaced missing values with None.

### **2.3 Feature Engineering**

* Constructed **“interests” list** per student: [club\_top1, club\_top2, hobby\_top1, hobby\_top2].
* Encoded interests using **MultiLabelBinarizer** → binary matrix representation.
* Appended teamwork preference as a numeric feature.

## **3. Methodology**

### **3.1 Clustering Approach**

* **Algorithm:** Agglomerative Clustering (Ward linkage).
* **Distance Metric:** Euclidean, applied on encoded vectors.
* **Cluster Range Tested:** k = 2 to 7.

### **3.2 Evaluation Metrics**

* **Silhouette Score** → Measures cohesion and separation (range: –1 to 1).
* **Visualization:** Multidimensional Scaling (MDS) to reduce high-dimensional interests into **2D friendship maps**.

## **4. Results**

### **4.1 Silhouette Analysis**

### ➡️ **Optimal k = 5**

**Figure:1**

### 

### **4.2 Cluster Profiles**

| **Cluster** | **Dominant Interests** | **Avg. Teamwork Preference** | **Interpretation** |
| --- | --- | --- | --- |
| **0** | Sports club (27), Cricket (22), Coding club (14), Coding (8), Gym (8) | **3.79** | **Athletic Coders** – Students balancing sports with coding interests; moderately team-oriented. |
| **1** | Cultural club (16), Dance (13), Music club (12), Painting (12), Coding club (8) | **4.05** | **Performing Artists** – Strongly oriented toward cultural & artistic activities; highly collaborative. |
| **2** | Coding club (25), Music (13), Robotics club (11), Coding (10), Entrepreneurship cell (9) | **3.46** | **Tech Innovators** – Students driven by coding, robotics, and startups; lean toward moderate teamwork. |
| **3** | Music club (21), Music (18), Cricket (8), Photography (8), Sports club (7) | **3.55** | **Creative All-Rounders** – Blend of music, sports, and photography; balanced teamwork orientation. |
| **4** | Cultural club (8), Writing (8), Literary club (6), Photography (4), Gym (3) | **3.18** | **Literary Enthusiasts** – Focused on writing, literature, and creativity; relatively lower teamwork preference. |

### 

**Figure:2**

### **4.3 Visualizations**

* **Figure 1:** Silhouette Score vs. Number of Clusters (shows k=5 peak).
* **Figure 2:** MDS Friendship Map (students plotted in 2D, colored by cluster).

## **5. Discussion**

* **Low silhouette (~0.11):** Indicates overlapping interests, which is natural in student populations.
* **Clusters still meaningful:** Each cluster highlights distinct social groups (performers, intellectuals, techies, athletes, all-rounders).
* **Actionable insights:**
  + Design **common room seating** where Performers & Athletes share collaborative zones.
  + Provide **quiet spaces** for Intellectuals.
  + Assign Techies & All-Rounders to **innovation and hackathon projects**.

## **6. Conclusion and Future Work**

This project demonstrated how **unsupervised clustering** can reveal friendship structures at SNU:

* Identified **5 distinct student groups**.
* Validated results using **silhouette scores** and visualization.
* Suggested practical seating and activity arrangements.

### **Future Directions**

* Explore **Jaccard similarity** for set-based interests.
* Experiment with **DBSCAN or Spectral Clustering** for non-convex groups.
* Collect additional features (event participation, online activity) to strengthen cluster quality.