AI-Based Student Club Recommendation System

Prof. Aboli More[1] Aditya Laxmikant Birhade[2] Manthan Niranjan Jamale[3] Ridima Ratnakar Naik[4]

[1,2,3,4]Department of Artificial Intelligence and Machine Learning

[1,2,3,4]Zeal Polytechnic, Narhe, Pune, Maharashtra, India

Abstract — Students can participate in college clubs to explore their interests, develop skills, and interact with other students. However, choosing a suitable club can be difficult when information about club activities and student interests is scattered across different sources. This paper proposes an AI-based Student Club Recommendation System that suggests suitable college clubs according to students' interests, skills, and activity preferences. The proposed system uses content-based filtering, Term Frequency–Inverse Document Frequency (TF-IDF), and cosine similarity to compare student preference profiles with club descriptions. The system can be developed using Python and its machine-learning libraries, with a simple web interface for collecting preferences and displaying recommendations. The aim is to make club discovery more personalized and accessible. The proposed approach will require testing with student preferences and club data to evaluate recommendation relevance and usability.

Keywords: Artificial Intelligence, Student Club Recommendation, Content-Based Filtering, TF-IDF, Cosine Similarity, Personalization

I. Introduction

College clubs provide opportunities for students to explore technical, cultural, sports, social, and creative activities. Participation in these clubs can help students develop practical skills, teamwork, communication, and leadership abilities. However, students may not always know which clubs match their interests and abilities.

A recommendation system can help solve this problem by analyzing student preferences and comparing them with available club information. Instead of manually searching through every club, students can receive a ranked list of relevant options.

This paper proposes an AI-based Student Club Recommendation System using content-based filtering. The system compares student interests with club descriptions and recommends clubs according to their similarity. The proposed system is intended to be simple, affordable, and suitable for implementation as a diploma-level project.

A. Problem Statement

Students may find it difficult to identify college clubs that match their interests, skills, and preferred activities because club information may be spread across different sources. A personalized recommendation system can help students discover relevant clubs more efficiently.

B. Objectives

1) To collect student interests, skills, and activity preferences.
2) To maintain descriptions of available college clubs.
3) To apply content-based filtering to match students with suitable clubs.
4) To rank club recommendations using cosine similarity.
5) To provide recommendations through a simple and user-friendly interface.

II. Literature Review

A. Academic Community Recommendation System Using Content-Based Filtering (2025)

Ashshiddiqi Zuhri and colleagues studied a recommendation system for academic communities at UIN Malang. The approach used content-based filtering, TF-IDF, and cosine similarity. The study involved 48 students and 10 communities and reported a System Usability Scale (SUS) score of 76. The work provides a relevant example of applying content-based recommendations to academic communities. Its reported scope was limited to a relatively small number of students and communities.

B. A Holistic Recommendation System for Higher Education Academic Activities (2025)

Rafael Alves da Silva and Lucas N. Ferreira proposed a recommendation approach for higher-education academic activities using Jina/BERT-based embeddings and linear regression on historical records. The study used a dataset containing approximately 1.6 million transcripts and 4,164 courses, covering records from 2010 to 2024. The authors reported that their approach performed favorably against collaborative filtering on the measures evaluated. This work demonstrates the use of text representations and historical data in educational recommendations. Its focus is academic activities rather than student clubs.

C. Enhancing Student Collaboration in Academic Projects Through a Content-Based Filtering Recommender System (2025)

Aldian Faizzul Anwar and colleagues investigated a content-based filtering system for matching students with academic projects. The approach used TF-IDF, cosine similarity, Python, Flask, and HTML. The study considered 198 completed student projects and reported a SUS score of 79, along with mean average precision (MAP) values of 0.92, 0.79, and 0.98 for different scenarios. This study is relevant because it demonstrates content-based filtering in a student-focused recommendation context. However, its primary purpose was project-team matching rather than club recommendation.

D. Research Gap

The reviewed studies demonstrate that content-based filtering and text similarity can be used for recommendations in educational settings. However, their main focuses include academic communities, higher-education activities, and project collaboration. The proposed system focuses specifically on matching students' interests and skills with college club descriptions through a simple, accessible

All rights reserved by www.ijsrd.com

interface. Its recommendation quality and usability must be evaluated through future testing.

III. Proposed Methodology

The proposed system uses content-based filtering to compare student preference information with descriptions of available clubs.

A. Student Preference Collection

Students enter their interests, skills, preferred activities, and areas they want to explore. Examples may include programming, robotics, public speaking, music, sports, design, and social service.

B. Club Database

The system stores information about each club, including its name, category, activities, required skills, and description. This information forms the basis for matching clubs with student preferences.

C. Text Preprocessing

The student preference text and club descriptions are converted into a consistent textual representation. Basic preprocessing may include converting text to lowercase and removing unnecessary words or symbols.

D. TF-IDF Vectorization

Term Frequency–Inverse Document Frequency (TF-IDF) converts text into numerical vectors. Words that help distinguish one preference profile or club description from another receive suitable weights. These vectors allow the system to compare textual information mathematically.

E. Cosine Similarity

Cosine similarity measures the similarity between the student preference vector and each club-description vector. A higher similarity score indicates a closer textual match.

F. Recommendation Ranking

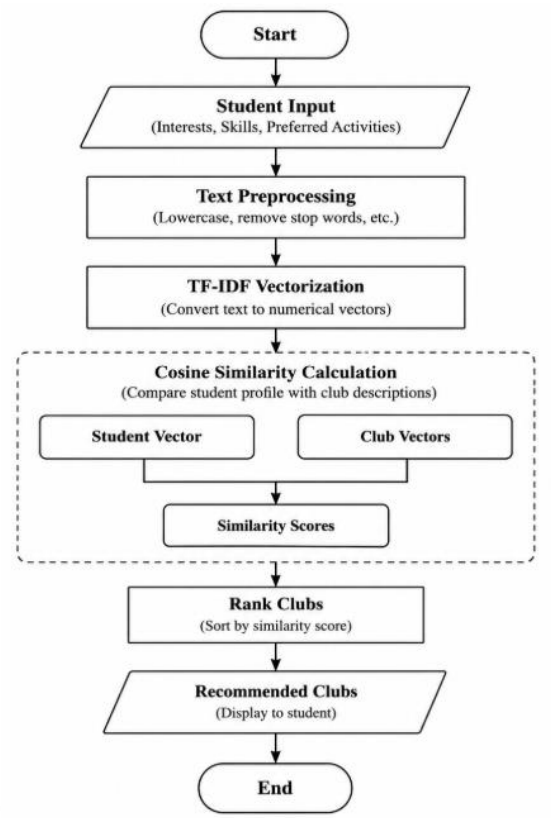
The system ranks clubs in descending order of their similarity scores and displays the most relevant options. Recommendations should be presented as suggestions rather than guarantees that a student will enjoy a particular club.

IV. System Architecture and Workflow

The proposed system consists of five main modules:

1) Student Interface: Collects student interests and preferences.
2) Club Database: Stores club names, categories, activities, and descriptions.
3) Preprocessing Module: Prepares preference text and club descriptions.
4) Recommendation Engine: Applies TF-IDF and cosine similarity.
5) Results Module: Displays ranked club recommendations.

V. Workflow:



A. Proposed Technologies

- Programming language: Python
- Machine-learning library: Scikit-learn
- Web framework: Flask
- Frontend: HTML and CSS
- Data storage: CSV files or SQLite database

These technologies are proposed choices for implementation and have not been presented as an already completed system.

VI. System Designation and Numerical Expressions

A. System Design:

1) Student Profile and Club Data

The proposed system uses two types of information: the student's profile and the details of available college clubs. The student profile contains interests, skills, and preferred activities. Club records contain the club name, category, activities, and a short description. Both types of text are prepared in a consistent format before comparison.

For example, a student interested in Python, robotics, and problem-solving may be matched with a coding club or robotics club. The recommendation depends on the information stored in the club descriptions, so accurate and updated data is important.

2) Text Representation Using TF-IDF

The system needs to convert text into a numerical form before comparing student preferences with club descriptions. TF-IDF assigns weights to words based on their frequency in a document and how common they are across the collection of documents. This helps distinguish meaningful terms from words that appear everywhere.

All rights reserved by www.ijsrd.com

The resulting vectors represent the student profile and each club description. These vectors are then passed to the similarity calculation stage.

3) Similarity-Based Recommendation

Cosine similarity is used to estimate how closely the student profile matches each club description. A similarity score is calculated for every available club. The clubs are then sorted by score, and the highest-ranked clubs are presented first.

This method is suitable for a diploma-level prototype because it can be implemented using Python and Scikit-learn without training a large, complex AI model. However, the recommendations depend on the quality of the input data and may not capture interests that a student has not mentioned.

B. Mathematical equations to add

Equation 1: TF-IDF calculation

$$TFIDF(t,d) = TF(t,d) \times IDF(t)$$

Where:

t = a word or term

d = a student profile or club description

$TF(t,d)$ = frequency of the term in the document

$IDF(t)$ = importance of the term across all documents

A commonly used IDF formula is:

$$IDF(t) = \log\left(\frac{N}{1+DF(t)}\right)$$

Here, N is the total number of documents and $DF(t)$ is the number of documents containing term t. The exact smoothing formula may vary depending on the implementation.

Equation 2: Cosine similarity

$$\text{sim}(A,B) = \frac{A \cdot B}{\|A\|\|B\|}$$

Where:

A = student preference vector

B = club description vector

$A \cdot B$ = dot product of the vectors

$\|A\|\|B\|$ = vector magnitudes

For non-negative TF-IDF vectors, the similarity generally ranges from 0 to 1. A higher score indicates a closer match between the student's preferences and a club's description.

Equation 3: Ranking clubs

Let S_i be the similarity score for club i. The recommended clubs are ordered from the highest score to the lowest:

$$S_1 \geq S_2 \geq S_3 \geq \cdots \geq S_n$$

The system displays the top $k$clubs, where k is the selected number of recommendations.

C. Add a worked example

Suppose a student enters the interests Python, coding, and robotics. The database contains a Coding Club, Robotics Club, Music Club, and Sports Club.

The system converts the student profile and club descriptions into TF-IDF vectors. It calculates a cosine similarity score for each club and sorts the clubs by score.

The following values are illustrative only, not actual experimental results:

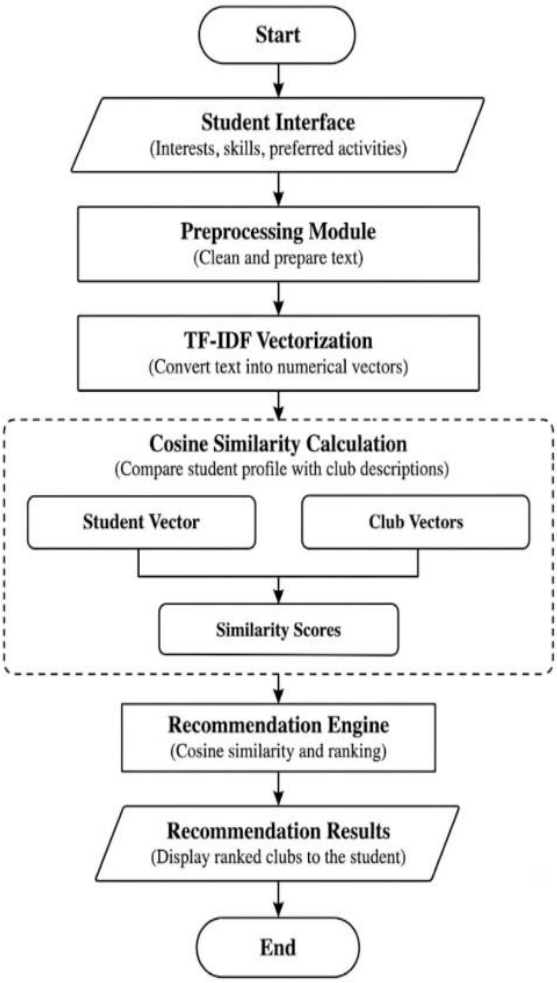
Club	Example similarity score
Coding Club	0.86
Robotics Club	0.78
Music Club	0.21
Sports Club	0.14

In this example, the Coding Club is ranked first because its description is the closest match to the student's entered interests. Actual scores depend on the collected text and the TF-IDF implementation.

Suggested caption: Table 1. Illustrative example of similarity-based club ranking.

VII. System Architecture and Algorithm

A. System Architecture:



B. Algorithm

1) Start the system.
2) Collect the student's interests, skills, and preferred activities.
3) Retrieve the available club descriptions from the database.
4) Preprocess the student profile and club descriptions.
5) Convert the text into TF-IDF vectors.
6) Calculate cosine similarity between the student profile and every club.
7) Sort the clubs in descending order of similarity scores.
8) Display the top-ranked clubs.
9) Stop.

All rights reserved by www.ijsrd.com

VIII. Evaluation Plan and Discussion

The proposed system should be evaluated after implementation using a sample of student preference profiles and verified club descriptions.

A. Recommendation Relevance

Students can review the recommended clubs and indicate whether the suggestions match their interests. The relevance of the displayed recommendations can then be assessed.

B. Usability

A small group of students can test the interface and provide feedback on ease of use, clarity, and usefulness. A standard usability questionnaire may be used.

C. Limitations

The recommendation quality will depend on the accuracy and completeness of student preferences and club descriptions. Content-based filtering may also recommend similar types of clubs repeatedly and may not discover interests that students have not entered. The proposed system does not replace students own decisions or official club information. The proposed system depends on the accuracy of club descriptions and the information entered by students. If a student's interests are unclear or a club has an incomplete description, the recommendations may be less relevant. Content-based filtering may also repeatedly suggest similar clubs and may not discover new interests automatically.

D. Expected Outcome

The expected outcome is a system that presents a ranked list of clubs based on textual similarity between student preferences and club descriptions. Actual recommendation accuracy, response time, and usability scores can only be reported after implementation and testing.

IX. Conclusion and Future Scope

This paper proposes an AI-based Student Club Recommendation System that uses content-based filtering, TF-IDF, and cosine similarity to match students with relevant college clubs. The approach offers a practical way to personalize club discovery using student preferences and structured club information. The proposed system can be implemented using Python, Scikit-learn, Flask, HTML, CSS, and a lightweight database. Future work may include collecting real student feedback, testing recommendation relevance, updating club information automatically, and exploring hybrid recommendation techniques to improve personalization.

Acknowledgment

The author would like to acknowledge the guidance and support provided by the faculty members of the Department of Artificial Intelligence and Machine Learning, Zeal Polytechnic, Narhe, Pune, for their academic guidance.

References

[1] A. A. S. Zuhri et al., "Development of Academic Community Recommendation System Using Content-Based Filtering at UIN Malang Informatics Engineering Study Program," 2025. Available: https://repository.uin-malang.ac.id/24093/

[2] R. A. da Silva and L. N. Ferreira, "A Holistic Recommendation System for Higher Education Academic Activities," Proceedings of the Educational Data Mining Conference, 2025. Available: https://educationaldatamining.org/EDM2025/proceedings/2025.EDM.short-papers.161/index.html

[3] A. F. Anwar et al., "Enhancing Student Collaboration in Academic Projects Through a Content-Based Filtering Recommender System," 2025. Available: https://repository.uin-malang.ac.id/25558/

All rights reserved by www.ijsrd.com