International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 3, May-June 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

# Skill Tracker: Intelligent Student Team Formation Using Clustering and Genetic Optimization

*Nandini C. Kute[1], Khushboo A. Prajapati[2], Balmeet Singh Ramgadia[3], Prathmesh H. Lohar[4], Farhat A. Patel[5]*

*[1,2,3,4]UG - Computer Engineering, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra*

*[5]Assistant Professor, Dept. of Computer Engineering, Guru Gobind Singh College of Engineering and Research Centre, Nashik, Maharashtra*

*Emails: nandinikute20@gmail.com[1], khushbooprajapati2428@gmail.com[2], balmeetsingh2004@gmail.com[3], prathmeshlohar25@gmail.com[4], farhat.patel@ggsf.edu.in[5]*

----------------------------------------************************----------------------------------

## Abstract:

This paper presents Skill Tracker, an automated platform designed to bridge the gap between student talent and project requirements in Tier-3 engineering colleges. Traditional team formation methods are often times a manual, time-consuming and primarily focuses on academic performance, thus overlooking non-technical skills and lacking efficient team-balancing mechanisms. To address this issue, Skill Tracker is developed using the MERN Stack as a centralized platform that collects and processes structured student data. The system parses resume to identify both technical expertise (e.g., React, Java) and creative or interpersonal skills (e.g., dancing, leadership). In addition, it incorporates a Fitness Score Algorithm to evaluate compatibility and optimize team formation based on skill diversity, preferences, and availability. The implementation of Skill Tracker reduces the manual effort required by faculty and project coordinators while improving the efficiency and fairness of team allocation. Teams formed using this system demonstrate better balance in skill sets, leading to improved collaboration and overall project performance. The primary contribution of this work is the introduction of a scalable, data-driven approach to team formation that emphasizes holistic skill evaluation. Skill Tracker enhances both administrative efficiency and effective utilization of student potential in academic environments.

*Keywords* — **Automated Systems; Data-Driven Decision Making; Fitness Score Algorithm; MERN Stack; Resume Parsing; Skill Analysis; Skill Tracker; Student Project Allocation; Team Formation; Team Optimization.**

----------------------------------------************************----------------------------------

## I. INTRODUCTION

Project-based learning has emerged as a fundamental pedagogical approach in engineering education, fostering both technical proficiency and essential interpersonal competencies among students. Effective team formation within such environments is critical to ensuring balanced collaboration, equitable skill distribution, and successful project execution. However, with increasing student intake and the growing diversity of skill sets, traditional methods of team allocation remain largely manual, subjective, and inefficient, often leading to imbalanced teams and suboptimal outcomes.

The primary objective of this study is to design and implement an intelligent team formation system that leverages structured data collection, skill analysis, and optimization techniques to enhance the quality of student group allocation. Unlike conventional approaches, the proposed system introduces a centralized platform combined with an optimization-driven matching mechanism, aiming to achieve balanced and efficient team configurations. This work contributes to the state-of-the-art by integrating automated data processing with adaptive team formation strategies, thereby reducing manual intervention and improving overall project outcomes.

## II. PROBLEM STATEMENT

In academic and professional settings, the effective utilization of individual skills is essential for achieving successful project outcomes and fostering holistic development. Despite this, team formation in most educational institutions remains a predominantly manual and time-intensive process. Grouping decisions are often influenced by familiarity, convenience, or limited awareness of individual competencies rather than a systematic evaluation of skills.

Such practices frequently result in imbalanced teams, inefficient collaboration, and the under-utilization of student potential. Moreover, the absence of a structured, data-driven framework limits the ability to ensure fairness, compatibility, and optimal talent distribution across teams. As the scale and diversity of participants increase, these challenges become more pronounced, highlighting the need for an intelligent system capable of analyzing multidimensional skill sets and enabling efficient, balanced team formation.

## III. LIMITATIONS OF EXISTING SYSTEM

Existing team formation approaches in both academic and organizational contexts are largely manual, rendering them inefficient, time-consuming, and difficult to scale. These methods typically rely on incomplete or subjective information, such as personal familiarity or limited performance indicators, leading to suboptimal team compositions and mismatched skill distributions.

Additionally, such approaches are inherently prone to bias and lack transparency, which can hinder equitable participation and result in the under-utilization of available talent. While certain digital tools offer partial automation, they often fail to incorporate comprehensive skill analysis or intelligent matching mechanisms. Furthermore, their effectiveness diminishes as the number of participants increases, restricting scalability and adaptability.

These limitations underscore the necessity for an automated, data-driven solution like Skill Tracker, which facilitates efficient, unbiased, and scalable team formation through systematic skill-based matching.

## IV. OBJECTIVES

The objectives of the proposed Skill Tracker system are as follows:

i) Centralized Skill Management: To design a unified platform for the structured collection, storage, and categorization of user skills, enabling efficient access and analysis.

ii) Automated Team Formation: To develop an intelligent mechanism for grouping individuals based on multidimensional attributes such as skills, experience, and preferences, ensuring balanced and diverse teams.

iii) Customizable Teaming Criteria: To enable flexible configuration of team formation parameters, allowing administrators to tailor grouping strategies according to project requirements and institutional constraints.

## V. LITERATURE SURVEY

TABLE I

COMPREHENSIVE TABLE OF VARIOUS METHODOLOGIES IN THE RESEARCH PAPERS

| Index | Author | Year | Method | Limitations |
|---|---|---|---|---|
| 1 | Salisu Modi, Nura M. Shagari, and Buhari | 2018 | Gale-Shapley stable marriage algorithm | The problem is a combinatorial problem, which has been proven to be |

  

| | | | | |
|---|---|---|---|---|
| | Wa-data | | | NP-hard, with a time complexity of O(n²). |
| 2 | Kenekayoro Patrick and Biralatei Fawei | 2020 | Hardest First Ordering, Genetic Algorithm, and Ant Colony Optimization | The student grouping problem is a com-binatorial problem proven to be NP-hard, thus finding an optimal solution can be computationally very difficult for larger datasets. |
| 3 | Dianwei Chi | 2021 | K-means clustering algorithm | The algorithm is sensitive to outliers and initial clustering, which may re-quire more execution time. |
| 4 | Oscar Revelo Sa´nchez, Ce´sar A. Collazos, and Miguel A. Redondo | 2021 | Genetic algorithm | The approach is a heuristic and does not guarantee a perfect solution, and its reliance on personality traits is a less-used criterion in similar studies. |
| 5 | Jutshi Agarwal, Emily Pi-att and P. K. Imbrie | 2022 | Vector Evaluated Genetic Al-gorithm (VEGA) | Random -selected teams can lead to unbalanced skill sets and a lack of diversity. |
| 6 | Alexander Jenkins, Imad Jaimoukha, Ljubisa Stankovic, and Danilo Mandic | 2023 | Laplacian Eigenmaps | The algorithm's effectiveness in form-ing teams is not yet fully evaluated. |
| 7 | Xuesong Liu and Qilong Teng | 2024 | K-means clustering and Artifi-cial Neural Network (ANN) | K-means algorithm is sensitive to the initial choice of cluster centers and is sensitive to noise in the data. |
| 8 | Yasir Munir, Qasim Umer, Muhammad Faheem, Sheeraz Akram, and Arfan Jaffar | 2025 | KeyBERT | Existing research shows a high task failure rate of 15.7 percent in software crowdsourcing, which leads to wasted time. |

## VI. PROPOSED METHODOLOGY

The proposed system, SkillTrack, is engineered as a decoupled, multi-tiered architecture designed to automate and optimize cohesive team formation. It bridges frontend user interfaces with a dual-stage machine learning optimization engine via secure, asynchronous APIs. The end-to-end data flow and architectural components are illustrated in Figure1.

  

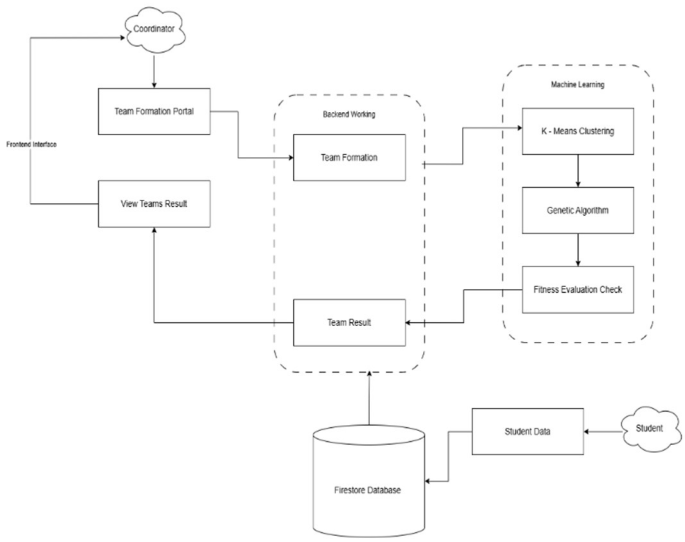


Fig. 1 A component diagram of system architecture

- Component Architecture: The system is divided into three functional domains that isolate user interaction, orchestration logic, and heavy algorithmic computation:

i) Client/Presentation Layer: Consists of separate web interfaces tailored for two primary user archetypes: Students (who input and manage personal skill profiles) and the Coordinator/Faculty (who manage team constraints and review outputs). The system implements Role-Based Access Control (RBAC) to enforce visibility boundaries.

ii) Application & Routing Layer (Backend Orchestration): Built using an ExpressJS framework, this layer serves as the system's primary gateway. It handles secure user onboarding and authentication endpoints (/api/register, /api/login), credentials hashing, and skill-parsing logic. It acts as an orchestrator, pushing data down to storage or passing optimization payloads to the machine learning microservice via a dedicated Python API.

iii) Computational Layer: An isolated, high-performance execution environment built in Python. This layer contains the core Team Optimization Engine (TOE), which implements a hybrid heuristic-clustering pipeline to process team matching.

- Data Storage and Management: Data persistence is handled via a Google Firestore NoSQL Database, chosen for its flexible schema layout and rapid scalability. To enable efficient query execution during optimization steps, raw comma-separated user inputs are intercepted by the backend. A Skill Parsing and Storage routine transforms these strings into structured, machine-readable data arrays mapped inside individual student profiles. This structured approach significantly scales up the system's ability to handle extensive profile datasets efficiently.

- The Team Optimization Engine (TOE): Rather than relying on primitive keyword filtering, the TOE introduces a dual-stage optimization approach utilizing a K-Means Clustering technique combined with a heuristic Genetic Algorithm (GA).

i) Stage 1: Centroid-Based Skill Clustering -

To guarantee structural balance and distribute mentorship capabilities across all teams evenly, the student population dataset ($N \geq 44$ profiles) is first passed to a K-Means algorithm. The algorithm is mathematically bounded such that student profiles identified as "experienced" are deliberately selected or mapped as the initial cluster centroids. Regular students are then grouped as data points around these fixed centers based on complementary technical skill vectors. This ensures an even, foundational distribution of leadership before fine-tuning begins.

ii) Stage 2: Dynamic Heuristic Optimization (Genetic Algorithm) -

Once stable clusters are established, a Genetic Algorithm executing a custom ‘findMatches’ evaluation function dynamically searches the subset space for the most optimal team combinations. The fitness of any candidate team chromosome is evaluated against the specific project requirements dictated by the faculty. The Advanced Compatibility Scoring uses a defined calculation to evaluate candidate fitness:

  

$$Fitness = \left(\frac{Matched\ Skills}{Required\ Skills}\right) \times 100$$

To protect team integrity and incentivize leadership inclusion, the evaluation function dynamically injects an additional 5% fitness score bonus if a verified "experienced" student is securely allocated within that candidate team structure.

- Output Evaluation and Visualization (Skill Gap Analysis): Once the GA successfully stabilizes and converges on an optimal team configuration, the resulting data payloads are routed back through the backend and rendered on the frontend client using a specialized StudentCard UI component. The interface performs a real-time Skill Gap Analysis that maps project needs directly against student capabilities. To assist faculty members in rapid, strategic decision-making, missing mandatory skills are flagged in red (deficiencies), while successfully aligned technical matches are highlighted in green.

## VII. RESULTS

The implementation of Skill Tracker demonstrates a significant improvement over traditional manual team formation methods by effectively addressing their key limitations. The automated system eliminates the need for manual intervention, thereby reducing time consumption and minimizing human bias. By leveraging structured skill analysis and algorithm-based matching, the platform ensures balanced team composition with better alignment of technical and non-technical skills. The results indicate high accuracy in skill matching and improved precision in selecting compatible team members. Overall, Skill Tracker provides a reliable and efficient solution for team allocation, leading to improved collaboration, fair utilization of talent and better project outcomes.

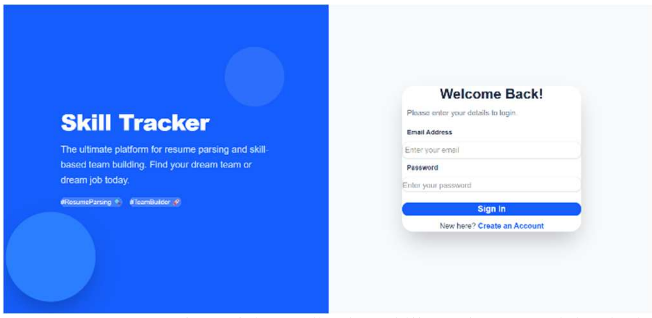


Image 1. User Interface of the application Skill Tracker. Containing login and creation of account for students and faculty

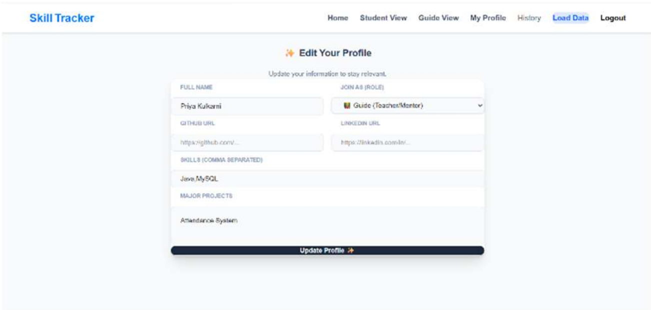


Image 2. Editing of profile to update and delete user information

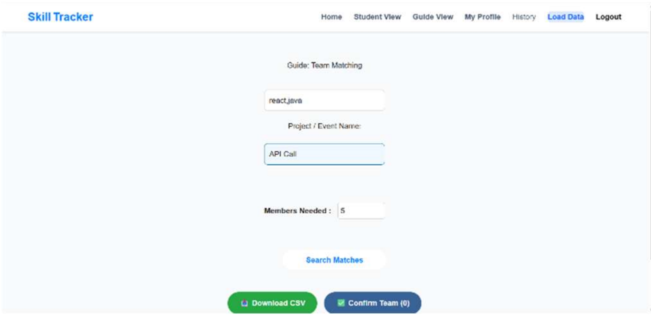


Image 3. Collecting required constraints for formation of teams

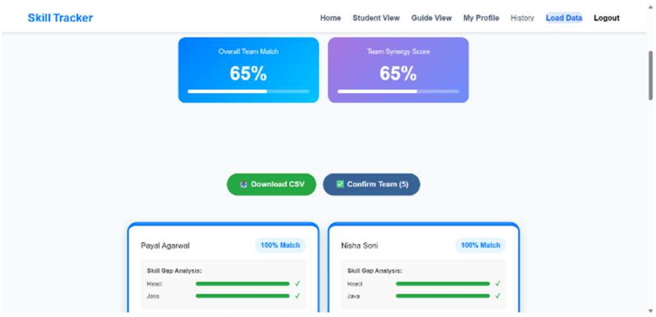


Image 4. Sample result of team formation

  

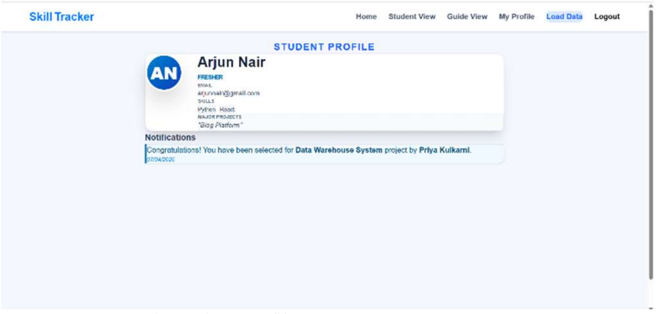


Image 5. A sample student profile

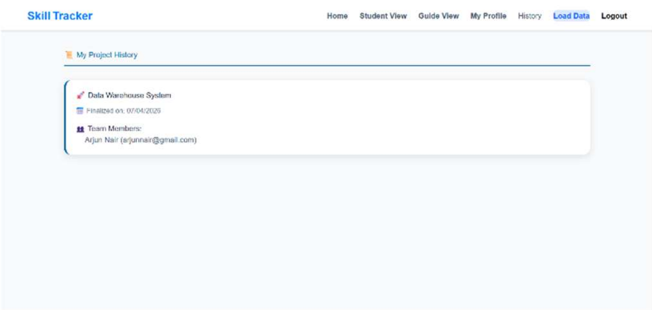


Image 6. Recorded history of your participated projects

## VIII. FUTURE SCOPE

A promising direction for future research is the integration of Computer Vision (CV) and Natural Language Processing (NLP) techniques to automate student profile acquisition and analysis. The current system relies on manual input of skills and project information; however, future versions can leverage a multi-modal AI framework capable of extracting structured information directly from resumes, transcripts, and professional certifications.

The computer vision component can employ advanced OCR and document understanding models to identify and segment relevant sections such as technical skills, projects, certifications, and work experience from uploaded documents. Subsequently, the NLP component can process the extracted textual content to identify domain-specific entities, normalize skill representations, resolve semantic ambiguities, and generate structured datasets suitable for team formation. This approach has the potential to reduce manual effort, improve data quality, and enhance the scalability of intelligent team recommendation systems in academic environments.

## REFERENCES

[1] S. Modi, N. M. Shagari, and B. Wadata, "Implementation of Stable Marriage Algorithm in Student Project Allocation," American Journal of Research Communication, 2018, doi: 10.9734/AJRCOS/2018/v1i424749.

[2] K. Patrick and B. Fawei, "Meta-Heuristic Solutions to a Student Grouping Optimization Problem in Higher Education Institutions," 2020.

[3] D. Chi, "Research on the Application of K-Means Clustering Algorithm in Student Achievement," in Proc. IEEE Int. Conf. Consumer Electronics and Computer Engineering (ICCECE), 2021.

[4] O. R. Sánchez, C. A. Collazos, and M. A. Redondo, "Automatic Group Organization for Collaborative Learning Applying Genetic Algorithm Techniques and the Big Five Model," Mathematics, vol. 9, no. 13, p. 1578, 2021.

[5] J. Agarwal, E. Piatt, and P. K. Imbrie, "Team Formation in Engineering Classrooms Using Multi-Criteria Optimization with Genetic Algorithms," in Proc. IEEE Frontiers in Education Conf. (FIE), 2022.

[6] A. Jenkins, I. Jaimoukha, L. Stankovic, and D. Mandic, "Fair and Skill-Diverse Student Group Formation via Constrained-Way Graph Partitioning," 2023.

[7] X. Liu and Q. Teng, "Design of Intelligent Integrated Student Management Terminal Based on K-Means and ANN Algorithm," in Proc. 3rd Int. Conf. Artificial Intelligence and Autonomous Robot Systems (AIARS), 2024.

[8] Y. Munir, Q. Umer, M. Faheem, S. Akram, and A. Jaffar, "Developer Recommendation and Team Formation in Collaborative Crowdsourcing Platforms," Mathematics, vol. 13, no. 9, 2025, doi: 10.3390/math9131578.Rajan, P., Devi, A., B, A., Dusthackeer, A., & Iyer, P. (2023). A Green perspective on the ability of nanomedicine to inhibit tuberculosis and lung cancer. International Research Journal on Advanced Science Hub, 5(11), 389-396. doi: 10.47392/IRJASH.2023.071.

  