

An Explainable Hybrid Machine Learning Approach for Data-Driven Career Recommendation and Skill Gap Analysis

Ms. Prachi Surve¹, Shravani Dahilkar²

¹Assistant Professor, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Belapur, Maharashtra, India.

²Student, Department of MCA, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Belapur, Maharashtra, India.

Abstract - Career choice is not always straightforward for students. Marks are important, but they are only one part of the decision. A student's interests, aptitude, awareness of available careers, expectations, and relevant skills can also influence the choice. This paper proposes an explainable hybrid machine learning approach that combines career recommendation with skill gap analysis. A questionnaire dataset was studied to understand students' career awareness, need for guidance, opinion about a recommendation application, and willingness to use such a system. Of the 40 collected records, one incomplete response was removed, leaving 39 usable responses for the preliminary analysis. The results indicate that many students would like career suggestions and that a considerable majority are open to using a technology-based recommendation tool. Based on these observations, the proposed framework brings together student profiling, machine learning, profile similarity, career matching, and skill gap analysis. Along with a recommendation, the system is designed to explain the main factors behind the result and point out skills that the student may need to strengthen. The work therefore provides a practical starting point for a more personalized and transparent career guidance system.

Keywords: career recommendation, explainable artificial intelligence, hybrid machine learning, skill gap analysis

I. INTRODUCTION

Career selection is an important decision in a student's academic life. Many students reach the stage of choosing a stream or profession without having a clear understanding of the options available to them. The difficulty becomes greater when several careers appear suitable on the surface but require different combinations of academic knowledge, technical skills, communication abilities, and personal interests.

Conventional career guidance can be useful, but it is often dependent on limited information or one-time counselling. A data-driven system can consider several student characteristics together and use patterns in collected data to support a more personalized recommendation. Recent work has explored machine learning datasets for career growth, hybrid career-path

recommendation, knowledge-graph-based career planning, and skill-gap analysis [1]-[4].

The main idea of this study is that recommending a career should not be the end of the process. A student also needs to know why the career is being suggested and what should be improved to move towards it. For this reason, the proposed approach combines career recommendation with skill gap analysis and an explainable output.

The present work uses questionnaire data as a preliminary study. The responses provide information about students' awareness, preferences, guidance requirements, and attitudes towards a career recommendation application. These findings are used to define the requirements of the proposed hybrid framework.

From a practical perspective, career recommendation systems should handle uncertainty rather than assume that one student attribute determines a career. An explainable system can present the main contributing factors so that students can review the recommendation instead of treating it as an unquestionable decision.

The reviewed studies also indicate the usefulness of hybrid recommendation methods when both student profile information and career-related relationships need to be considered. This provides a basis for combining machine learning with similarity-based matching in the proposed framework.

Career decisions may depend on interests, self-perceived abilities, personal preferences, and awareness of available opportunities in addition to academic performance. A recommendation system that combines these factors can provide a broader view of career suitability than a single academic indicator.

II. LITERATURE REVIEW

Acharya et al. introduced a student dataset designed for machine-learning-based career growth recommendation. The dataset brings together several student-related attributes and shows the value of structured educational data for career-oriented machine learning [1].

Almufareh and Almaleh proposed a hybrid career-path recommendation system that combines academic profiles, skills, and preferences using content and collaborative recommendation ideas [2]. Their work supports the use of a hybrid approach because career decisions depend on more than one type of student information.

Xu and Gao investigated a knowledge-graph-integrated approach for college student career planning. Their work represents relationships between students, skills, interests, and career information and demonstrates how connected information can support career planning [3].

Another recent study considered career recommendation together with skill gap analysis. Such an approach is useful because a recommendation becomes more practical when it is followed by information about the skills that a student needs to develop [4].

The reviewed work indicates that career recommendation is moving towards personalized and data-driven systems. However, there is still scope for a simple, explainable framework that connects student profiling, career matching, and skill-gap identification in one workflow. The present study focuses on this combined perspective.

The study is designed to move beyond simple career suggestions by considering the information needs of students and the skills required for a possible career path. The objectives therefore cover both recommendation and the practical identification of areas for improvement.

III. OBJECTIVES OF THE STUDY

- To study the factors that influence students while making career decisions.
- To analyze students' awareness, guidance requirements, and willingness to use a career recommendation system.
- To propose a hybrid machine learning framework for personalized career recommendation.
- To identify the skill gap between a student's present skills and the skills required for a selected career.
- To support transparent and student-friendly career guidance through explainable machine learning techniques.
- To develop a framework that connects career recommendation with a personalized skill-development plan.
- To identify the role of academic background, interests, aptitude, and preferences in career recommendation.
- To examine the common difficulties students face while selecting a suitable career path.

- To provide explanations for career recommendations so that students can understand the basis of the result.

IV. RESEARCH HYPOTHESES

H₀₁: There is no significant association between students' career awareness and their willingness to use a career recommendation tool.

H₁₁: There is a significant association between students' career awareness and their willingness to use a career recommendation tool.

H₀₂: There is no significant association between students' need for career suggestions and their willingness to use a career recommendation tool.

H₁₂: There is a significant association between students' need for career suggestions and their willingness to use a career recommendation tool.

H₀₃: There is no significant association between students' perception of whether their marks match their desired career and their willingness to use a career recommendation tool.

H₁₃: There is a significant association between students' perception of whether their marks match their desired career and their willingness to use a career recommendation tool.

For a future implementation, the methodology can be extended by collecting validated skill scores and confirmed career outcomes. Such labelled data would support model training, validation, and comparison of alternative algorithms without overstating the findings of the present exploratory dataset.

The research process consists of five broad stages: questionnaire preparation, data collection, data cleaning, exploratory analysis, and framework design. The questionnaire responses were first checked for completeness, after which the usable records were summarized using frequencies and percentages. The results were then used to identify the functional requirements of the proposed recommendation framework.

V. RESEARCH METHODOLOGY

The study follows a quantitative and exploratory survey-based design. Primary data were collected through a structured questionnaire. The supplied spreadsheet contained 40 records. One record was incomplete, so 39 usable responses were considered for the preliminary analysis.

The questionnaire included questions related to age, gender, class, academic stream, career awareness, aptitude-test experience, factors influencing career choice, perceived match between marks and desired career, need

for career suggestions, existing guidance, expected usefulness of an application, difficulties in career selection, and willingness to use a recommendation tool.

The analysis involved data cleaning, descriptive statistics, percentage calculations, and preliminary association analysis. The proposed machine learning stage is described as a framework because the present questionnaire does not contain a confirmed career outcome label or detailed individual skill-level measurements. Therefore, predictive accuracy is not claimed from this dataset.

The descriptive analysis is used to summarize the composition of the sample and the respondents' perceptions. Bar graphs are used for the academic-stream and career-awareness distributions, while pie charts are used for the need for career suggestions and willingness to use a recommendation tool. These visualizations make the distribution of responses easier to compare and interpret.

VI. DATA ANALYSIS AND INTERPRETATION

The cleaned dataset contains 39 usable responses. The respondents represent Science, Commerce, and Arts streams. The preliminary results indicate that students have different levels of career awareness and different expectations from career guidance technology.

TABLE I. DISTRIBUTION BY ACADEMIC STREAM

Stream	Responses	Percentage
Science	16	41.0%
Commerce	13	33.3%
Arts	10	25.6%
Total	39	100%

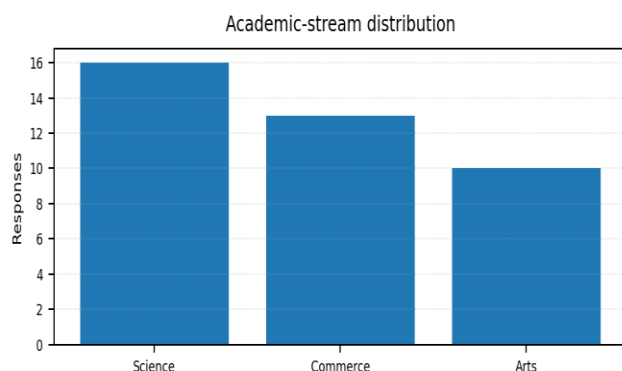


FIGURE 2. Bar graph of academic-stream distribution.

TABLE II. CAREER AWARENESS

Rating	Responses	Percentage
1	7	17.9%
2	8	20.5%

3	8	20.5%
4	8	20.5%
5	8	20.5%

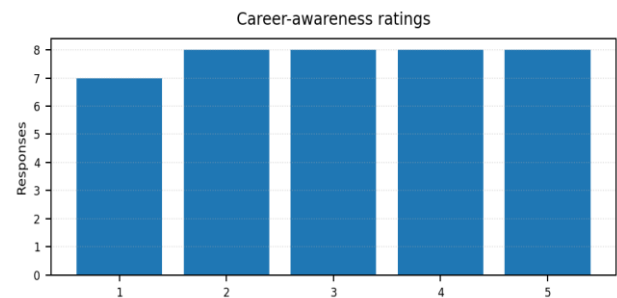


FIGURE 3. Bar graph of career-awareness ratings.

TABLE III. NEED FOR CAREER SUGGESTIONS

Response	Responses	Percentage
Yes	23	59.0%
No	15	38.5%
Incomplete/other	1	2.6%

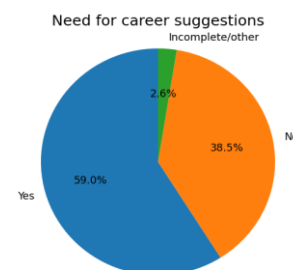


FIGURE 4. Pie chart of need for career suggestions.

TABLE IV. WILLINGNESS TO USE A CAREER RECOMMENDATION TOOL

Response	Responses	Percentage
Yes	31	79.5%
No	7	17.9%
Incomplete/other	1	2.6%

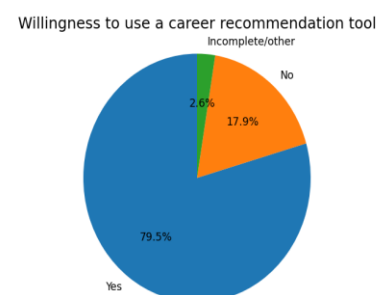


FIGURE 5. Pie chart of willingness to use a career recommendation tool.

The average career-awareness score in the usable sample was approximately 3.05 out of 5. This indicates a moderate level of awareness overall, with responses spread across the rating scale. About 59.0% of the usable responses indicated a need for career suggestions. The willingness to use a career recommendation tool was higher: 31 of the 39 usable responses, or approximately 79.5%, selected yes.

Students also reported different difficulties while selecting a career. The responses included having too many choices, fear of choosing the wrong path, confusion, family pressure, and not knowing enough options. These responses suggest that career guidance needs to address uncertainty as well as information gaps.

A preliminary chi-square analysis can be used to examine associations among categorical variables. For the present sample, the observed relationships should be treated as exploratory because the sample is small and is not necessarily representative of the wider student population.

VII. PROPOSED EXPLAINABLE HYBRID MACHINE LEARNING APPROACH

The proposed framework combines machine-learning-based student profiling with similarity-based career matching and skill-gap analysis. The objective is not simply to predict a career label but to produce a recommendation that can be understood by the student.

First, the system creates a student profile using academic information, aptitude, interests, preferences, awareness, and available skill information. Next, a supervised model such as Random Forest can be used to learn patterns from a sufficiently large labelled dataset. A content-based component can then compare the student profile with career profiles containing required skills and attributes.

The final stage performs skill gap analysis. For each relevant skill, the system compares the current skill level with the expected level for the selected career. The output can show the student's matching strengths and the skills that need improvement.

Explainability is included in the final output. Instead of displaying only a career name, the system can show the major factors that contributed to the recommendation, the skills already aligned with the career, and the skills that should be developed.

The skill-gap component can also support personalized learning recommendations. After identifying a gap, the system can suggest relevant courses, practice activities, projects, or other learning resources. This makes the output more actionable than a career name alone and

creates a clear connection between recommendation and skill development.

The hybrid design can be implemented as a sequence of connected modules. A preprocessing module can encode categorical information and normalize numerical attributes, followed by a prediction or ranking module. The recommendation score can then be combined with profile similarity to produce a ranked set of career options. An explanation module can translate the important contributing attributes into simple statements for the student.

VIII. SKILL GAP ANALYSIS

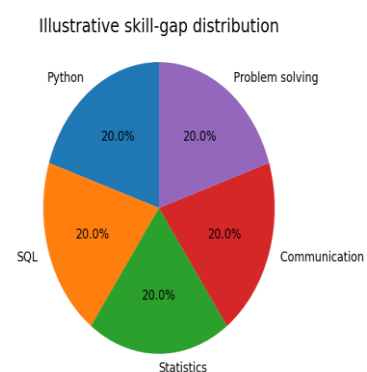


FIGURE 1. Skill gap distribution for the illustrative career profile.

Skill gap analysis is an important part of the proposed system. For example, if a student is matched with a data-oriented career, the system may compare current levels in Python, SQL, statistics, communication, and problem solving with the expected levels for that career.

TABLE V. EXAMPLE OF SKILL GAP ANALYSIS

The example shows how the recommendation can be converted into an actionable learning plan. A gap of zero indicates that the current level meets the assumed requirement, while a positive gap indicates an area for development. In an actual implementation, the required levels should be obtained from validated career profiles and industry or job-market data.

IX. KEY FINDINGS

- The usable survey sample contains 39 responses after removing one incomplete record.
- Science students form the largest stream group in the sample.
- The average career-awareness rating is approximately 3.05 out of 5.
- About 59.0% of respondents indicated that they wanted career suggestions.

- About 79.5% indicated that they would use a career recommendation tool.
- Students reported confusion, too many choices, fear of choosing the wrong path, family pressure, and limited awareness as career-selection difficulties.
- The findings support the need for a personalized career guidance approach.
- Combining career recommendation with skill-gap analysis can make the recommendation more useful and actionable.
- An explainable output can help students understand why a recommendation was produced.

X. LIMITATIONS OF THE STUDY

- The sample size is limited to 39 usable responses.
- The respondents represent a limited age and educational group.
- The questionnaire responses include subjective perceptions.
- The present dataset does not contain detailed skill-level measurements for every respondent.
- There is no confirmed career outcome label in the supplied data; therefore, model accuracy cannot be reported honestly from this dataset alone.
- The findings should not be generalized to all students without a larger and more diverse sample.

XI. FUTURE SCOPE

Future work can extend the dataset by collecting detailed technical and soft-skill scores, interests, aptitude scores, preferred careers, and actual career outcomes. A larger labelled dataset would allow supervised models to be trained and evaluated using suitable performance measures. Job descriptions can also be processed using natural language processing to identify current skill requirements. The recommendation engine could combine content-based filtering, collaborative filtering, and machine learning. Explainable artificial intelligence techniques such as feature importance and SHAP can be used to make individual recommendations easier to understand. Finally, the framework can be implemented as a web or mobile application that provides career suggestions together with a personalized skill-development plan.

Future development can also include periodic updates of career profiles so that recommendations reflect changing technologies and job requirements. A feedback mechanism could record whether students found a recommendation useful and use that feedback to improve future ranking,

while privacy controls would be needed when handling personal educational data.

XII. CONCLUSION

This study presents an explainable hybrid machine learning framework for data-driven career recommendation and skill gap analysis. The preliminary survey analysis shows that students in the sample have varied levels of career awareness and that many are interested in receiving career suggestions through a technology-based system. The proposed framework responds to this need by combining student profiling, machine learning, career similarity matching, and skill-gap analysis.

A major strength of the proposed approach is that it does not stop at recommending a career. It also aims to explain the recommendation and identify the skills that may require improvement. The current survey data provides a useful starting point, but a larger labelled dataset with detailed skill information and confirmed career outcomes is required before the predictive model can be trained and evaluated. The proposed framework can therefore serve as a foundation for a future personalized and transparent career guidance application.

XIII. REFERENCES

- 1) S. Acharya, S. Shetty, N. N. Prabhu, and N. Shetty, "A novel student dataset for ML based effective career growth recommendation," Scientific Reports, vol. 16, 2026.
- 2) T. Almufareh and A. Almaleh, "Hybrid career path recommendation system for computer science students: Integrating academic profiles and personal skills with a content-collaborative model," Electronics, vol. 15, no. 14, 2026.
- 3) X. Xu and H. Gao, "A knowledge graph-integrated recommendation method for college student career planning," Discover Artificial Intelligence, vol. 6, 2026.
- 4) R. A. MabelRose, G. LakshmiPriya, M. P. HarshiniSwetha, and V. Dhanya, "Career recommendation system with skill gap analysis," International Research Journal on Advanced Engineering Hub, 2026.