Volume 11, Issue 5, May – 2026 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2010

An AI-Driven Data Mining Approach for Predicting Student Career Streams Using Academic and Behavioural Indicators

Dr. Sameer P. Patil[1]

[1]Shree Damodar College of Commerce and Economics Margao Goa

Publication Date: 2026/06/05

Abstract: This study explores how Artificial Intelligence (AI) and Machine Learning (ML) can be used to predict which career stream higher secondary students may choose. Many students today find it difficult to choose the right career path. This is because they do not have enough information about different careers. They may also feel confused or unsure about their future. Their choices can also be influenced by family, friends, and other people around them. This study checks whether school marks and personal factors like career awareness, family pressure, peer influence, and self-confidence affect these decisions. Data was collected from 160 higher secondary students using a simple questionnaire. The answers were checked for reliability using a method called Cronbach's Alpha. After that, the data was studied using basic statistical methods and tests like the Shapiro-Wilk test and the Kruskal-Wallis test. AI methods such as Decision Tree, Naive Bayes, Logistic Regression, and Support Vector Machine (SVM) were used to predict students' career choices. The results showed that students' marks and behaviour did not have a strong effect on their career stream choice. The AI models showed different levels of accuracy, but the differences were not statistically important. Among them, SVM gave the best accuracy. The study concludes that career choice depends on many factors. The factors used in this study are not enough to fully explain students' decisions. Future studies with more data and more factors may give better and clearer results.

Keyword: *Artificial Intelligence (AI), Machine Learning (ML), Career Stream Prediction, Higher Secondary Students, Academic Performance, Behavioural Factors, Career Choice, Decision Tree, Random Forest, Support Vector Machine (SVM), Logistic Regression, Educational Data Mining, Predictive Analytics.*

How to Cite: Dr. Sameer P. Patil (2026) An AI-Driven Data Mining Approach for Predicting Student Career Streams Using Academic and Behavioural Indicators. *International Journal of Innovative Science and Research Technology*, 11(5), 3252-3258. https://doi.org/10.38124/ijisrt/26may2010

I. INTRODUCTION

Artificial Intelligence (AI) and Machine Learning (ML) are changing many parts of our daily life, including education. Schools and colleges now use data to understand students better and help them learn more easily. One important use of AI in education is to predict how students may perform in studies and help them choose the right career path.

Choosing a career stream after higher secondary school is an important decision for students. This choice can affect their future studies, jobs, and career growth. Many students find it hard to choose the right stream because they may not have enough knowledge about careers or may feel unsure about their future. Their decisions are also influenced by family expectations, friends' opinions, and their academic performance.

Traditionally, students get career guidance through counselling sessions, aptitude tests, and their academic marks. However, these methods may not fully explain why students choose certain careers. Academic performance is important, but other factors also affect career decisions. These include career awareness, family influence, friends' opinions, and self-confidence.

With the growth of AI and ML technologies, researchers can now create smart systems that study student data and make predictions. Machine Learning methods like Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression can study student information and find patterns related to career choices. These methods can help teachers and career counsellors guide students in a better way.

This study focuses on higher secondary students and aims to create an AI-based system that predicts students' career stream choices using academic and behavioural data. The study also compares different Machine Learning models to find out which model gives more accurate prediction results. By understanding the factors that affect career choices, this research can help schools and colleges provide better career guidance and support students in making the right decisions for their future.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2010

II. RESEARCH OBJECTIVE

➢ *General Objective:*

To create and test an AI-based system that predicts the possible career streams of higher secondary students using their academic performance and behavioural factors.

➢ *Specific Objective*

- To study the connection between students' academic performance and their preferred career streams.
- To understand how behavioural factors like career awareness, family expectations, friends' influence, and self-confidence affect students' career stream choices.
- To find the most important academic and behavioural factors that influence career stream selection.
- To create and use prediction models such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Logistic Regression to predict career streams.
- To compare different AI and Machine Learning models using measures like accuracy, precision, recall, and F1-score to find which model performs better.
- To provide data-based insights that can help teachers, counsellors, and educational institutions guide students toward suitable career paths.

➢ *Hypothesis*

- *Academic Performance:*

✓ H01 (Null Hypothesis)::- There is no strong connection between students' marks and the career stream they choose.
✓ H11 (Alternative Hypothesis)::- There is a strong connection between students' marks and the career stream they choose.

- *Behavioural Factors:*

✓ H02 (Null Hypothesis): Things like career knowledge, family pressure, friends' opinions, and self-confidence do not greatly affect the career stream students choose.
✓ H12 (Alternative Hypothesis): Things like career knowledge, family pressure, friends' opinions, and self-confidence greatly affect the career stream students choose.

- *Career Awareness:*

✓ H03 (Null Hypothesis): Knowing about different careers does not greatly affect the career stream students choose.
✓ H13 (Alternative Hypothesis): Knowing about different careers greatly affects the career stream students choose.

- *Family & Peer Influence:*

✓ H04 (Null Hypothesis): Family and friends do not have a big effect on the career stream chosen by students.
✓ H14 (Alternative Hypothesis): Family and friends have a big effect on the career stream chosen by students.

- *Model Performance:*

✓ H05 (Null Hypothesis): The AI models used in the study do not show much difference in how well they predict results.
✓ H15 (Alternative Hypothesis): The AI models used in the study show a clear difference in how well they predict results.

III. LITERATURE REVIEW

Trujillo, Pozo, and Suntaxi (2025)[1], in their systematic literature review published in the *Journal of Technology and Science Education (JOTSE)*, studied many research papers on how Artificial Intelligence helps in student career prediction. They carefully selected 38 important studies from 1296 papers. Their study shows that machine learning is now widely used to help predict students' career choices in higher education. The authors explain that common methods used in these studies are Random Forest, Support Vector Machine (SVM), Neural Networks, and Logistic Regression. Among these, Random Forest and SVM work better because they give more accurate results and can handle large and complex student data. The review also shows that these systems use different types of student information such as exam marks, interests, and background details to make better predictions. To check how well the systems work, researchers use simple measures like accuracy, precision, recall, and F1-score. The authors also point out some problems in this field, such as the lack of common datasets, fewer long-term studies, and concerns about fairness, bias, and understanding how the AI makes decisions. In the end, they conclude that AI-based career prediction systems are very helpful for guiding students, but more work is needed to make them more clear, fair, and useful in real-life situations.

Sanjeev Jain (2024)[2], in the article *"AI-Driven Career Guidance System: A Predictive Model for Student Subject Recommendations Based on Academic Performance and Aspirations"* published in the *International Journal of Medical Informatics (IJMI)*, explains how Artificial Intelligence (AI) can help students choose the right subjects and career paths. The study shows that AI and Machine Learning (ML) systems look at student marks, interests, and goals to give better suggestions. The author explains that traditional career counseling can be slow and depends on human opinion, which may not always be fully accurate. In comparison, AI-based systems give faster and more data-based results. The study also says that machine learning methods can study student data and suggest suitable career options more effectively. This helps reduce mistakes where students choose subjects that do not match their skills or interests. However, the author also mentions some problems. The system depends on good quality data, and sometimes AI can show bias. It is also hard to understand how some AI models make decisions. In the end, the study concludes that AI-based career guidance systems are very useful for students because they help in better and more informed career choices.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2010

Recent research in Educational Data Mining (EDM) and Artificial Intelligence (AI) shows that computers are now used to predict which career path a student may choose. These systems look at both exam marks and student behavior to make better predictions.

Pathak et al. [3] made a machine learning system to predict student placement results. They used different methods like Decision Tree, Random Forest, Logistic Regression, and Support Vector Machine (SVM). These methods study student data and help find patterns about job chances.

Their study found that school marks like GPA, subject marks, and exam scores are very important for prediction. These marks are strong signs of future career success because they are clear and easy to measure.

But new research says that marks alone are not enough. Student behavior is also important. This includes attendance in class, using online learning systems, completing assignments, and taking part in class activities. When both marks and behavior are used together, the predictions become more accurate.

Recent research in Educational Data Mining (EDM) and Artificial Intelligence (AI) shows that computers are now being used to study student data and predict their performance and career choices. Schools and colleges collect a lot of data about students. This includes marks, attendance, and learning activity. This data can be used to build smart systems that help predict how students will perform in the future.

Studies show that many AI methods like Decision Tree, Naïve Bayes, Support Vector Machine (SVM), and Neural Networks are used to find patterns in student data. These patterns help identify students who may do well or who may need support early.

New research also shows that using both exam marks (like GPA and scores) and student behavior (like attendance, participation, and online learning activity) gives better results than using only marks. This makes predictions more accurate [4], [5].

Researchers also found that combining different AI methods together (called ensemble learning) gives better results than using only one method [6].

However, there are still some problems. Some data is not balanced, meaning some groups have more data than others. Also, student behavior data is not always collected in the same way. Another problem is that some AI models are difficult to understand [7].

Because of this, researchers suggest using better systems that combine different types of data. They also recommend using explainable AI (XAI), which helps people understand how the AI makes decisions. This will make predictions more clear, fair, and useful for schools and colleges [8].

IV. METHODOLOGY

This study uses numbers and data to understand students' choices. Information is collected using a set of clear questions given to higher secondary students. The study looks at important factors like students' marks and their behaviour to see how they choose their career stream. The questions are checked for reliability using a method called Cronbach's Alpha to make sure they are consistent. The study also uses AI methods like Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression. These methods help to group data and predict which career stream a student may choose.

- *Likert Measurement Scale:*

Table 1 Likert Scale Distribution

Scale Value	Description
1	Strongly Disagree
2	Disagree
3	Neutral
4	Agree
5	Strongly Agree

- *Reliability Study*

Reliability analysis was done using a method called Cronbach's Alpha. This was used to check if the questions in the survey give consistent results. Cronbach's Alpha shows how closely related a group of questions are when measuring the same idea. Its value ranges from 0 to 1. A value of 0.70 or more is usually considered good in social science studies. In this study, Cronbach's Alpha was used for groups of questions about academic performance and behaviour. These included career awareness, family expectations, peer influence, and self-confidence. The results showed good consistency. This means the questions in each group measured the same idea properly. The dependent variable, which is career stream choice, was not tested for reliability because it is a category, not a set of questions. Overall, the results show that the data is reliable and suitable for further analysis and AI-based prediction.

Table 2 Reliability Test

Group	Questions	Cronbach's Alpha
Academic Performance	Q1-Q8	.91
Career Awareness	Q9-Q16	.85
Family Influence	Q17-Q23	.85
Peer influence	Q24-Q29	.85
Decision Certinity	Q30-Q36	.90

- *Descriptive Analysis*

The study looked at the data in a simple way to understand the main patterns. It used basic statistics to describe the information. The study included 160 people. Their data was studied using averages and how the values spread out. The results showed that academic performance was between 40% and 90%. The average score was 66.17%. This means most students had a moderate level of performance. Most questions that used a rating scale showed

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2010

high scores above 4.00. This means many students agreed strongly with the statements. The main factors found were academic influence, career awareness, and career confidence. These were very important in shaping career choices. Family and friends had a smaller effect compared to these factors. Overall, the study shows that students have good career awareness and confidence, while outside social influence is not very strong.

➢ *Hypothesis Testing*

First, the reliability of the questionnaire was checked using Cronbach's Alpha. After that, descriptive statistics were used to understand the basic details of the data. Then, hypothesis testing was done using inferential statistics.

The normality of the data was checked using the Shapiro-Wilk test. This test helped to find out whether the data followed a normal distribution or not. Based on the result, the correct statistical test was selected. All tests were conducted at a 5% level of significance.

- *Academic Performance and Career Stream Choice:*

To study the relationship between students' marks and their career stream choice, the following hypotheses were used:

✓ H0 (Null Hypothesis): There is no significant relationship between students' marks and the career stream they choose.
✓ H1 (Alternative Hypothesis): There is a significant relationship between students' marks and the career stream they choose.

The Shapiro-Wilk test showed that the academic performance data were not normally distributed because the p-value was less than 0.05. Therefore, the non-parametric Kruskal-Wallis test was used.

The Kruskal-Wallis test result showed no significant relationship between academic performance and career stream choice ($\chi^2 = 5.2313$, df = 7, p = 0.6318).

Since the p-value was greater than 0.05, the null hypothesis was accepted and the alternative hypothesis was rejected. This means that students' marks do not have a significant effect on their choice of career stream.

➢ *Hypothesis Testing: Behavioural Factors and Career Stream Choice*

This part of the study checked whether behavioural factors affect the career stream chosen by students.

✓ *The Following Hypotheses were Tested:*

- H02: Career awareness, family influence, friends' opinions, and self-confidence do not affect students' career stream choice.
- H12: Career awareness, family influence, friends' opinions, and self-confidence affect students' career stream choice.

The Shapiro-Wilk test showed that the data were not normally distributed. Therefore, the Kruskal-Wallis test, a non-parametric test, was used.

The results showed that career awareness did not have a significant effect on career stream choice (p = 0.1124). Family influence (p = 0.4175), peer influence (p = 0.4175), and self-confidence (p = 0.8142) also did not show a significant effect on career stream choice.

Since all p-values were greater than 0.05, the null hypothesis (H02) was accepted and the alternative hypothesis (H12) was rejected. This means that behavioural factors did not significantly influence students' choice of career stream in this study.

✓ *To Study the Effect of Career Awareness on Students' Career Stream Choice, the following Hypotheses Were Used:*

- H03: Knowing about different careers does not significantly affect the career stream students choose.
- H13: Knowing about different careers significantly affects the career stream students choose.

The Shapiro-Wilk test showed that the data were not normally distributed because the p-value was less than 0.05. Therefore, the non-parametric Kruskal-Wallis test was used.

The results of the Kruskal-Wallis test showed that career awareness did not have a significant effect on career stream choice ($\chi^2 = 11.657$, df = 7, p = 0.1124).

Since the p-value was greater than 0.05, the null hypothesis (H03) was accepted and the alternative hypothesis (H13) was rejected.

The findings show that career awareness did not significantly influence students' choice of career stream in this study.

This hypothesis was tested to understand whether family and friends affect the career stream chosen by students.

✓ *The Following Hypotheses were Used:*

- H04: Family and friends do not affect the career stream chosen by students.
- H14: Family and friends affect the career stream chosen by students.

The Shapiro-Wilk test showed that the data were not normally distributed. So, the Kruskal-Wallis test was used for the analysis.

The results showed that family influence did not have a significant effect on career stream choice (p = 0.4175). Peer influence also did not show a significant effect on career stream choice (p = 0.4175).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2010

Since both p-values were greater than 0.05, the null hypothesis (H04) was accepted and the alternative hypothesis (H14) was rejected. This means that family and friends did not strongly influence the career stream chosen by students in this study.

- *Model Performance*

To study the differences in the prediction performance of the AI models used in the study, the following hypotheses were used:

- ✓ H05 (Null Hypothesis): The AI models used in the study do not show a significant difference in prediction performance.
- ✓ H15 (Alternative Hypothesis): The AI models used in the study show a significant difference in prediction performance.

The prediction accuracy of the AI models was as follows: Decision Tree (47.91%), Naive Bayes (25.00%), and Support Vector Machine (58.33%).

The normality of the model performance data was checked using the Shapiro-Wilk test. The result showed that the data were normally distributed because the p-value was greater than 0.05 ($W = 0.95522$, $p = 0.5928$).

To compare the performance of the AI models, the Kruskal-Wallis test was used. The test results showed no significant difference in the prediction performance of the AI models ($\chi^2 = 2$, $df = 2$, $p = 0.3679$).

Since the p-value was greater than 0.05, the null hypothesis (H05) was accepted and the alternative hypothesis (H15) was rejected.

The findings show that the AI models used in the study did not have a significant difference in their prediction performance.

- *Interpretation and Findings:*

The study examined whether academic performance, career awareness, and family and peer influence affected students' career stream choice.

The results showed that academic performance did not have a significant effect on the career stream chosen by students. Career awareness also did not significantly influence students' career stream choice. Similarly, family and friends did not show a strong influence on the career stream selected by students.

The study also compared the prediction performance of different AI models, including Decision Tree, Naive Bayes, and Support Vector Machine. Although the accuracy values of the models were different, the statistical analysis showed that there was no significant difference in their prediction performance.

The findings of the study are based on the responses collected from 160 students. Since the sample size was limited, the results represent the views and patterns observed within this group only. If a larger sample size is considered in future studies, the findings and prediction results may vary. Therefore, future research with more participants may provide broader and more accurate insights.

V. DISCUSSION OF RESULTS

The study found that academic performance, career awareness, and family and peer influence did not have a strong effect on students' career stream choice. These findings are similar to some earlier studies which also showed that career decisions are not influenced by only one factor. Career choice is a complex process and can be affected by many personal, social, and environmental factors.

The study showed that academic marks did not significantly affect the career stream selected by students. This means that students may choose their career streams based on their interests, skills, goals, or future opportunities instead of only focusing on marks.

The findings related to career awareness also showed no significant influence on career stream choice. Some earlier studies found that career awareness helps students in making career decisions, while other studies reported that awareness alone may not be enough without proper career guidance and support.

The study also found that family and friends did not strongly influence students' career stream choice. This result is different from some previous studies which reported that parents and peers have a major role in career decisions. One possible reason for this difference may be that many students today prefer to make their own career choices based on their personal interests and future plans.

The study also compared the performance of AI models such as Decision Tree, Naive Bayes, and Support Vector Machine. Although the models showed different accuracy values, the statistical analysis showed that there was no significant difference in their prediction performance.

The findings are based on responses collected from 160 students. Since the sample size was limited, the results represent only the students included in the study. If more responses are collected in future studies, the findings and prediction results may be different. Future research with a larger sample size may provide more detailed and accurate results.

➢ *Major Finding of the Study*

- Academic performance did not strongly affect students' career stream choice.
- Career awareness did not have a significant influence on the career stream chosen by students.
- Family and friends did not show a strong effect on students' career stream choice.
- The AI models used in the study did not show a significant difference in prediction performance.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2010

- Among all the AI models, the Support Vector Machine (SVM) showed the highest prediction accuracy.
- The findings of the study were based on responses collected from 160 students.
- The results suggest that students' career stream choices may also be influenced by other personal, social, or environmental factors that were not included in this study.
- The study showed that different AI models can be used to predict career-related outcomes, but their performance may change depending on the dataset and the number of responses used in the study.

VI. CONCLUSION

The study examined the factors that affect students' career stream choice using statistical analysis and AI models. The results showed that academic performance, career awareness, and family and peer influence did not strongly affect students' career stream choice.

The study also compared the performance of different AI models, including Decision Tree, Naive Bayes, and Support Vector Machine (SVM). Among these models, SVM showed the highest prediction accuracy. However, the statistical analysis showed that there was no significant difference in the performance of the AI models.

The findings of the study were based on responses collected from 160 students. The results suggest that students' career choices may also be affected by other personal, social, or environmental factors that were not included in this study.

Overall, the study highlights the importance of exploring more factors and using larger datasets in future research to better understand students' career stream choices and improve prediction accuracy.

SUGGESTIONS / RECOMMENDATIONS

Schools should give better career guidance to students. They should explain different job options in a simple way.

Students should learn about many different careers. Schools can arrange talks, workshops, and visits to companies.

Parents and teachers should help students choose careers. They should talk with students and support their decisions.

Schools should have good career counselling services. Students should be able to ask questions and get help easily.

Students should try internships and skill activities. This will help them understand real jobs better.

Future studies should include more students from different places. This will give better and more correct results.

Future research should also study more things like students' interests, personality, and social media use.

➢ *Limitations of the Study:*
This study has some limits.

- It includes only selected students, not all students.
- The number of students taken for the study is small.
- Only a few things that affect career choice were studied.
- The AI model used gives only medium-level accuracy, not very high.

➢ *Scope for Future Research:*
Future studies can do more work in this area. They can include more districts, states, or even countries. This will help to get better and wider results.

Researchers can also study more factors. These can include social, psychological, and environmental factors that affect students' choices.

Future studies can use better AI and deep learning models. These models can help to get more accurate results.

Researchers should also use larger data sets. This will help to make the results stronger and more reliable.

Future research can also study students over a long time. This will help to see how their career choices change as they grow.

ACKNOWLEDGMENT

The author expresses sincere gratitude to all individuals and institutions who contributed to the successful completion of this research work. The support, guidance, and encouragement received throughout the study were invaluable.

The author is especially thankful to the faculty members of the Department of Computer Science for their continuous academic support, insightful suggestions, and motivation during the course of this research.

The author also extends heartfelt thanks to the participants who contributed data for this study, without whom this research would not have been possible.

Finally, the author acknowledges the support of family and friends for their constant encouragement and understanding throughout the research journey.

REFERNCES

[1]. F. Trujillo, M. Pozo, and G. Suntaxi, "Artificial intelligence in education: A systematic literature review of machine learning approaches in student career prediction," *Journal of Technology and Science Education (JOTSE)*, 2025. Available: https://jotse.org/index.php/jotse/article/view/3124

[2]. S. Jain, "AI-Driven Career Guidance System: A Predictive Model for Student Subject Recommendations Based on Academic Performance and Aspirations," *International Journal of Medical*

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26may2010

Informatics (IJMI), vol. 8, no. 1, 2024. [Online]. Available: https://healthinformaticsjournal.com/index.php/IJMI/article/view/781

[3]. A. Pathak, M. Matcha, M. Gopisetti, and S. Joshi, "A machine learning framework for predicting student placement outcomes," *International Information and Software Technology (ISI Journal)*, IIETA, 2025. doi: 10.18280/isi.300704.

[4]. Y. Zhang, Y. Yun, and R. An, "Educational data mining techniques for student performance prediction: Method review and comparison analysis," *Frontiers in Psychology*, vol. 12, 2021, doi: 10.3389/fpsyg.2021.698490.

[5]. M. Mubashar Hussain et al., "Prediction of student academic performance through data mining approach," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, 2024. doi: 10.33093/jiwe.2024.3.1.16.

[6]. N. Kansal and V. Kansal, "An efficient data mining approach to improve students' employability prediction," *International Journal of Computer Applications*, vol. 178, no. 47, pp. 29–35, 2019.

[7]. R. Sharma and S. S. Shrivastava, "A review on performance of student's prediction using data mining techniques," *International Journal of Advance Research and Innovative Ideas in Education*, vol. 7, no. 4, pp. 1367–1371, 2021.

[8]. S. Y. Kumar and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," *arXiv preprint*, arXiv: 1203.3832, 2012.