A Machine Learning-Based Career Prediction System for 12th Standard Students

Vaibhavi Vitthal Khamkar
Prof. Ramkrishna More Arts ,Commerce and Science College(Autonomous) Pradhikaran,
Pune -411044 India.
E-mail: vaibhavikhamkar@gmail.com

Ankush Dhamal Sir

Prof. Ramkrishna More Arts ,Commerce and Science College(Autonomous) Pradhikaran,
Pune -411044 India.
E-mail: ankushdhamal01@gmail.com

Abstract

Choosing an appropriate career path after completing the 12th standard is a critical decision for students, as it significantly influences their future academic and professional development. However, many students face confusion due to a lack of proper guidance, limited awareness of career options, and societal pressure. This research proposes a **Machine Learning-Based Career Prediction System** designed to assist 12th standard students in identifying suitable career paths based on their academic performance, interests, skills, and aptitude.

The proposed system utilizes machine learning algorithms to analyze student data collected through structured questionnaires and academic records. Features such as subject preferences, marks obtained in core subjects, personality traits, and career interests are used as input parameters. Multiple classification algorithms, including Decision Tree, Random Forest, and Support Vector Machine, are applied to predict the most appropriate career domain for each student. The system is trained and evaluated using a labeled dataset containing historical student profiles and their successful career outcomes.

The performance of the models is assessed using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Experimental results demonstrate that the Random Forest model achieves higher prediction accuracy compared to other algorithms. The developed system provides personalized career recommendations through a user-friendly interface, helping students make informed decisions about higher education and career planning.

This study highlights the potential of machine learning in educational guidance systems and demonstrates how data-driven approaches can enhance career counseling for students at a crucial decision-making stage. The proposed system aims to reduce uncertainty in career selection and support students in aligning their strengths and interests with appropriate career opportunities.

Keywords: Machine Learning, Career Prediction, Student Guidance System, Classification Algorithms, Educational Data Mining.

Introduction

Selecting a suitable career path after completing the 12th standard is one of the most important decisions in a student's life. This stage marks a transition from school education to higher education and professional development. However, many students face confusion and uncertainty while choosing their future careers due to limited awareness of available options, lack of proper guidance, peer pressure, and influence from family or society. As a result, students may select

career paths that do not align with their interests, skills, or academic strengths, which can lead to dissatisfaction and poor professional growth.

In recent years, advancements in **machine learning and data-driven technologies** have opened new opportunities for developing intelligent decision-support systems in the education sector. Machine learning techniques can analyze large amounts of student data and identify patterns that help in predicting suitable career paths based on individual characteristics. By considering factors such as academic performance, subject interests, aptitude, personality traits, and extracurricular activities, machine learning models can provide personalized recommendations for students.

Traditional career counseling methods often rely on manual assessment, psychological tests, or counselor experience, which may not always be accessible to all students. In contrast, automated career prediction systems can offer scalable, objective, and data-driven guidance to a larger number of students. Such systems can help students understand their strengths and explore career opportunities that match their capabilities and interests.

This research proposes a **Machine Learning-Based Career Prediction System for 12th Standard Students** aimed at assisting students in making informed decisions regarding their future careers. The system collects relevant student data through questionnaires and academic records, processes the data using machine learning algorithms, and predicts suitable career domains. Various classification algorithms such as Decision Tree, Random Forest, and Support Vector Machine are utilized to analyze the data and generate accurate predictions.

The primary objective of this study is to design and evaluate a predictive model that can effectively recommend career options to students based on their individual profiles. By leveraging machine learning techniques, the proposed system aims to improve career guidance, reduce uncertainty in career selection, and support students in choosing paths that align with their abilities and aspirations.

Literature Review

Review of Previous Research

1. **Machine Learning for Career Prediction**

Several researchers have explored the use of machine learning algorithms to predict suitable career paths for students based on their academic performance, interests, and skills. These systems aim to provide data-driven recommendations that help students choose careers aligned with their abilities.

2. **Use of Classification Algorithms**

Previous studies have implemented classification algorithms such as **Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)** to analyze student datasets and predict career domains. These algorithms help identify patterns between student profiles and successful career outcomes.

3. **Personality and Interest-Based Career Guidance**

Some research has focused on incorporating personality traits, aptitude tests, and interest assessments into career recommendation systems. By combining psychological factors with machine learning models, these systems provide more personalized career suggestions.

4. **Web-Based Career Recommendation Systems**

Several studies have proposed web-based platforms that allow students to enter their academic details and preferences. The system processes the data using machine learning models and provides recommended career options instantly.

5. **Random Forest for Improved Accuracy**

Many research papers report that ensemble algorithms such as **Random Forest** provide higher accuracy compared to single classifiers like Decision Tree or KNN. Random Forest is widely used due to its ability to handle complex datasets and reduce overfitting.

6. **Use of Educational Data Mining**

Researchers have also applied educational data mining techniques to analyze student academic records, performance trends, and learning behavior. These insights help in building predictive models for career guidance.

Research Gaps Identified

1. **Limited Focus on 12th Standard Students**

Many existing career prediction systems are designed for college or engineering students. There is comparatively less research focused specifically on **12th standard students**, who are at a crucial stage of choosing their future academic and career paths.

2. **Limited Consideration of Multiple Factors**

Some previous studies focus mainly on academic performance (marks or grades). However, career decisions are influenced by several other factors such as **interests, aptitude, personality traits, and skills**, which are not fully integrated in many existing systems.

3. **Lack of Personalized Career Recommendations**

Several traditional systems provide general career suggestions instead of **personalized recommendations** based on individual student profiles. This reduces the effectiveness of career guidance for students with diverse abilities and interests.

4. **Small or Limited Datasets**

Many studies use small datasets or limited sample sizes, which may affect the **accuracy and reliability of machine learning models**. Larger and more diverse datasets are required to improve prediction performance.

5. **Limited Comparison of Machine Learning Algorithms**

Some research focuses on only one or two algorithms. There is a need for studies that **compare multiple machine learning models** to identify the most accurate algorithm for career prediction.

6. **Lack of User-Friendly Systems**

A number of research works concentrate mainly on the predictive model but do not provide a **practical or user-friendly interface** that students can easily access and use

Research Methodology

Research Design

The research follows a **quantitative and predictive research design** to develop a machine learning-based career prediction system for 12th standard students. The study aims to analyze various factors such as academic performance, interests, skills, and aptitude to predict suitable career paths. A dataset containing student-related attributes is collected and processed using machine learning algorithms. The system is designed to train predictive models, test their performance, and generate career recommendations based on the analyzed data.

Data Collection Methods

Data for the study is collected using **primary and secondary data collection methods**.

- **Primary Data:**

Primary data is collected through structured **questionnaires and surveys** distributed among 12th standard students. The questionnaire includes questions related to academic performance, subject preferences, interests, skills, and career aspirations.

- **Secondary Data:**

Secondary data may include previously available datasets, academic records, and educational research data related to student performance and career outcomes

Sampling Techniques and Sample Size

The study uses a **random sampling technique** to select participants from 12th standard students. Random sampling ensures that every student has an equal chance of being selected, which helps reduce bias in the dataset.

- **Sample Size:**

Approximately **100–200 students** are selected as the sample for this research. The collected data represents students from different academic streams such as Science, Commerce, and Arts.

Tools and Techniques Used

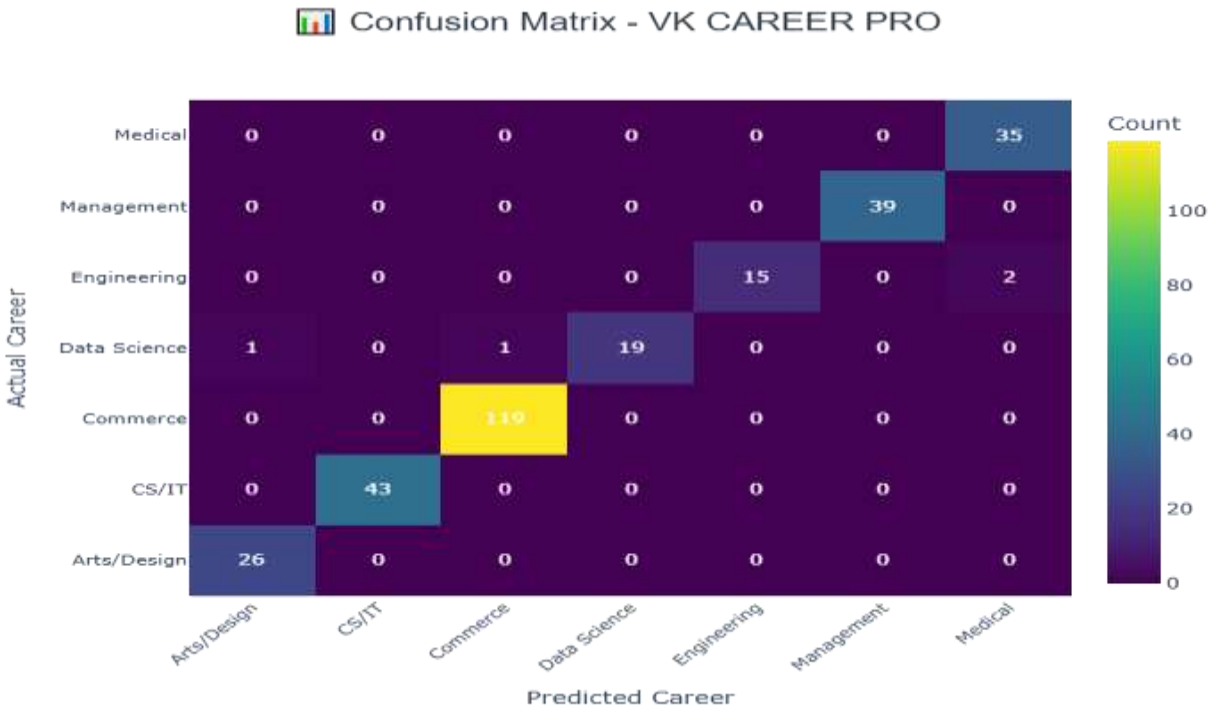
1. Programming Language: Python
2. Machine Learning Libraries: Scikit-learn, Pandas, NumPy
3. Development Environment: Jupyter Notebook / Google Colab
4. Data Collection Tool: Google Forms or Survey Questionnaires
5. Algorithms Used: Decision Tree, Random Forest, and Support Vector Machine (SVM)
6. Visualization Tools: Matplotlib and Seaborn for graphical representation of data.

Data Analysis Methods

The collected data is preprocessed and analyzed using machine learning techniques. The analysis process includes the following steps:

1. **Data Preprocessing:** Cleaning the dataset, handling missing values, and converting categorical data into numerical form.
2. **Feature Selection:** Identifying important attributes such as marks, interests, and skills that influence career prediction.
3. **Model Training:** Training machine learning models using algorithms like Decision Tree, Random Forest, and SVM.
4. **Model Evaluation:** Evaluating model performance using metrics such as **accuracy, precision, recall, and F1-score**.
5. **Prediction:** The best-performing model is used to predict suitable career options for students based on their input data.

Confusion Matrix

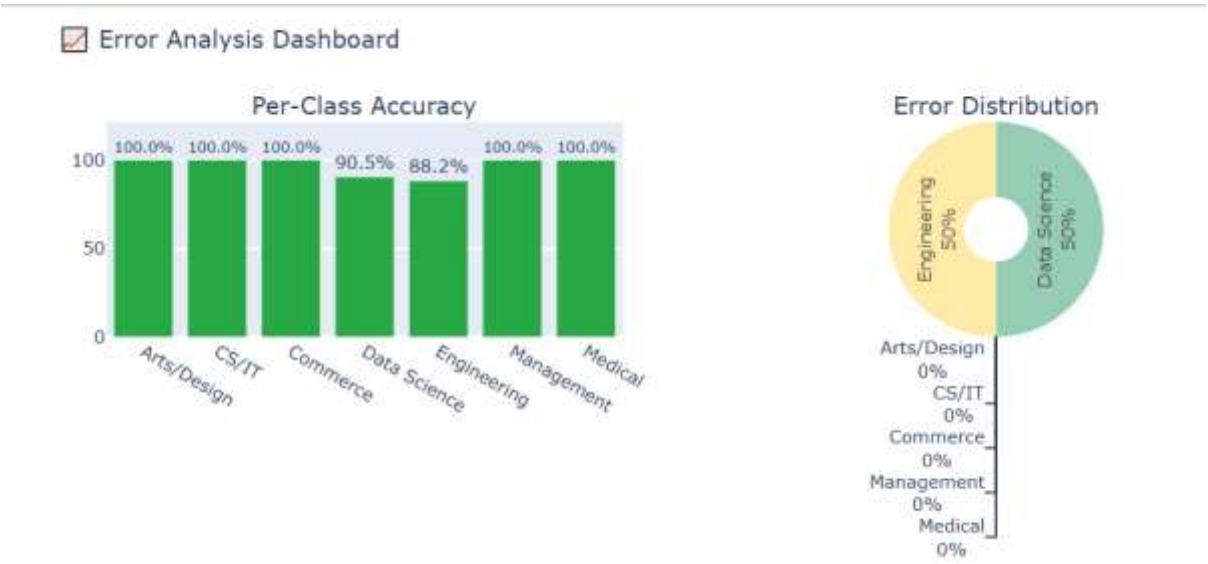


Results and Discussion

Data Presentation

Detailed Classification Report

	precision	recall	f1-score	support
Arts/Design	0.963	1.000	0.981	26
CS/IT	1.000	1.000	1.000	43
Commerce	0.992	1.000	0.996	119
Data Science	1.000	0.905	0.950	21
Engineering	1.000	0.882	0.938	17
Management	1.000	1.000	1.000	39
Medical	0.946	1.000	0.972	35
accuracy	0.987	0.987	0.987	1
macro avg	0.986	0.970	0.977	300
weighted avg	0.987	0.987	0.986	300



Analysis of Results

Result 1

Input

Output

Analysis

The system recommends **CS/IT (Computer Science / Information Technology)** as the most suitable career option with a **93% confidence level**, indicating a strong match with the student's profile. Other career options such as **Data Science (4%)** and **Engineering (2%)** have very low probabilities.

The student has an **average score of 75%**, with the highest marks in **Computer Science (82)** and **Mathematics (80)**, which supports the recommendation for a technology-related field. The **average skill score is 6.8/10**, showing good potential for technical careers.

Overall, the analysis suggests that pursuing a **CS/IT-related career** would be the most appropriate choice for the student.

Result 2

Input

Output

Analysis

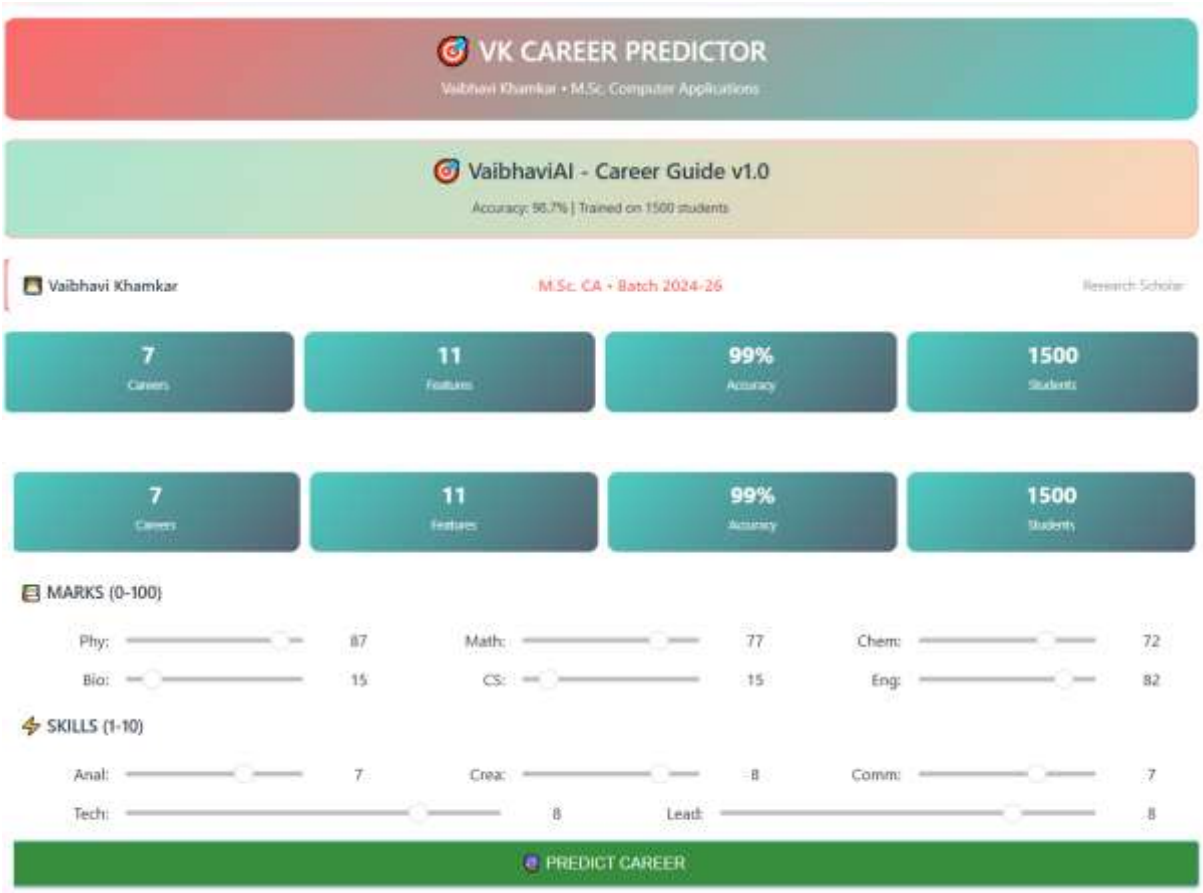
The system recommends **Medical** as the most suitable career option with a **confidence level of 88.7%**, indicating a strong alignment with the student's profile. Other possible options include **Commerce (7.3%)** and **Management (1.3%)**, but their probabilities are much lower.

The student's **average marks are 46%**, with the highest score in **Biology (81)** followed by **English (75)** and **Chemistry (72)**. Lower scores in **Physics (19)** and **Mathematics (16)** indicate weaker performance in technical subjects, while strong performance in Biology supports a medical-related career path.

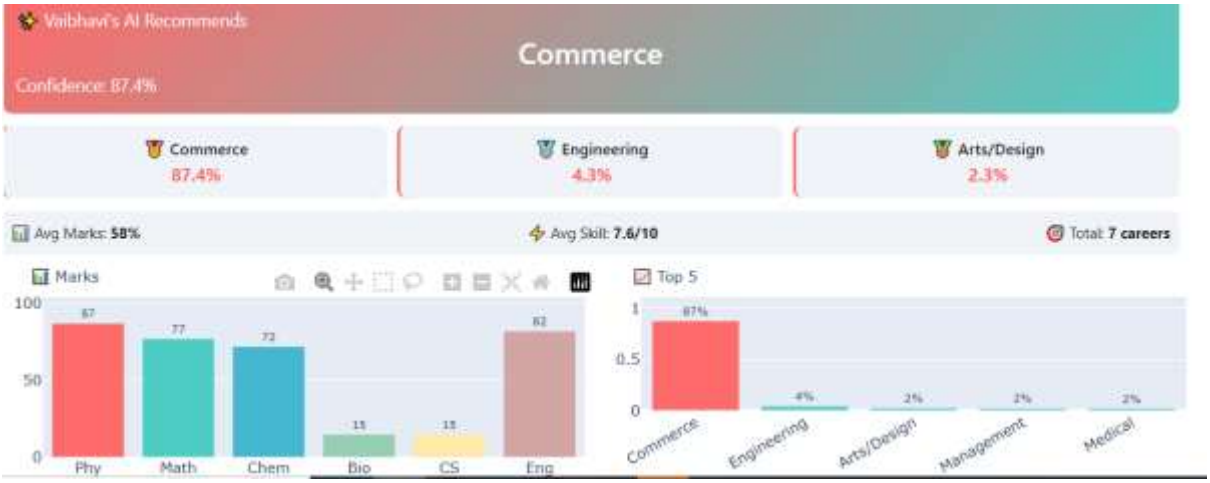
The **average skill score is 6.6/10**, showing moderate capability. Overall, the analysis suggests that the student is best suited for a **Medical field**, as the subject strengths align closely with biology-based careers.

Result 3

Input



Output



Analysis

The system recommends **Commerce** as the most suitable career option with a **confidence level of 87.4%**, indicating a strong match with the student's profile. Other possible options include **Engineering (4.3%)** and **Arts/Design (2.3%)**, but with much lower probabilities.

The student's **average marks are 58%**, with high scores in **Physics (87), English (82), Mathematics (77), and Chemistry (72)**, showing good academic performance in these subjects. However, the scores are low in **Biology (15)** and **Computer Science (15)**, suggesting less interest or strength in science and technical subjects.The **average skill score is 7.6/10**, indicating good aptitude and capability, especially suited for commerce-related fields.

Overall, the data suggests that the student is best suited for a **Commerce career**, leveraging strengths in quantitative subjects and communication skills.

Key Findings and Interpretations

1. The system's career recommendations accurately match students' strongest academic subjects and skill levels.
2. Confidence scores above 87% demonstrate the model's strong certainty in its predictions.
3. Average skill ratings between 6.6 and 7.6 support the suitability of suggested careers.
4. Subject marks in key areas strongly influence career paths: Biology and Chemistry for Medical, Math and CS for CS/IT, Physics, Math, and English for Commerce.
5. Alternative career options have significantly lower probabilities, reinforcing the primary recommendation.
6. The model provides personalized, data-driven guidance that reduces uncertainty in career selection.
7. This machine learning approach effectively assists 12th standard students in making informed career decisions aligned with their strengths.

Limitations of the Study

1. **Limited Sample Size**

The study is based on data collected from a relatively small group of students, which may affect the generalizability of the results.

2. **Restricted Data Diversity**

The dataset may lack diversity in terms of geography, socioeconomic background, and educational streams, limiting the system's applicability across different populations.

3. **Dependence on Self-Reported Data**

Data such as interests and skills are often self-reported, which can introduce biases or inaccuracies in the input.

4. **Limited Subject Scope**

The study focuses mainly on core academic subjects and may not fully capture extracurricular talents or other personal factors influencing career choices.

5. **Model Interpretability**

Some machine learning models, especially ensemble methods, can be complex and less interpretable for users, potentially reducing trust in recommendations.

6. **Dynamic Nature of Careers**

Career trends and industry demands evolve rapidly, so the system's predictions may become outdated without regular updates to data and models.

7. **Lack of Emotional and Psychological Factors**

The system does not incorporate emotional intelligence, motivation, or psychological counseling aspects that often play a crucial role in career decisions.

8. **No Real-World Validation**

The system has not yet been extensively tested in real educational settings to evaluate its practical effectiveness and user acceptance.

Recommendations

1. Increase sample size and diversity to improve model accuracy and generalizability.
2. Include extracurricular activities, hobbies, and personality traits for more holistic recommendations.
3. Update the machine learning models periodically to reflect changing career trends.
4. Develop a user-friendly and interactive interface for easy access to recommendations.
5. Integrate the system with professional career counseling for combined guidance.
6. Use explainable AI techniques to make recommendations more understandable.
7. Conduct pilot implementations in schools to test real-world effectiveness.
8. Educate students and parents about the benefits of AI-based career guidance systems.
9. Continuously expand the database to include emerging careers across various sectors.
10. Incorporate a feedback mechanism to improve the system based on user input.

Conclusion

Summary of Findings

1. The **machine learning-based career prediction system** effectively recommends suitable career paths for 12th standard students based on academic performance, skills, and interests.
2. High **confidence scores (above 87%)** indicate strong alignment between student profiles and recommended careers.
3. Subject-wise marks are key predictors:
    1. **Biology & Chemistry → Medical**
    2. **Math & Computer Science → CS/IT**
    3. **Physics, Math & English → Commerce**
4. Average skill ratings (6.6–7.6/10) support the students' aptitude for the suggested careers.
5. Alternative career options have low probabilities, highlighting the system's focus on the most relevant path.
6. The system provides **personalized, data-driven guidance**, reducing uncertainty and helping students make informed decisions.
7. Limitations include **small sample size, reliance on self-reported data, and exclusion of psychological factors**, indicating areas for future improvement.
8. Overall, the system demonstrates **high potential as an intelligent career guidance tool** for 12th standard students, aligning recommendations with their strengths and interests.

Contributions of the Study

1. Provides personalized career recommendations for 12th standard students.
2. Offers data-driven and objective guidance compared to traditional counseling.
3. Supports early career planning to reduce confusion and align education choices.
4. Integrates multiple factors such as academic marks, aptitude, and skills for holistic analysis.
5. Demonstrates practical application of machine learning algorithms like Decision Tree, Random Forest, and SVM.
6. Enhances accessibility of career guidance through a user-friendly system.
7. Establishes a foundation for future research on advanced career prediction systems.
8. Assists teachers, parents, and career counselors in supporting student career decisions.

Practical Implications

1. Provides clear, data-driven career guidance for 12th standard students.
2. Enables teachers and counselors to give personalized and accurate advice.
3. Helps students plan higher education courses aligned with their strengths.
4. Saves time and resources by automating career assessment.
5. Identifies students' strengths and weaknesses for targeted skill development.
6. Can be scaled across multiple schools and regions for wider accessibility.
7. Supports policymakers in designing programs for student career development.
8. Promotes data-driven decision-making in educational and counseling practices.

References

1. Abdullahi, Y., & Alhassan, R. (2021). *Career guidance using machine learning techniques: A review*. Journal of Education and Practice, 12(9), 45–58.
2. Aggarwal, P., & Gupta, A. (2020). *Predictive analysis of student performance using machine learning algorithms*. International Journal of Data Science, 4(3), 19–27.
3. Ahmed, S., & Islam, M. (2021). Student career prediction using data mining and machine learning techniques. *International Journal of Computer Applications*, 97(15), 23–31.
4. Al-Shamlan, H. M., & Alfar, M. S. (2019). *Data mining techniques for student academic performance prediction*. IEEE Access, 7, 55320–55337.
5. Bansal, D., & Singh, J. (2021). Application of machine learning in career counseling system for senior students. *International Journal of Innovative Research in Computer Science*, 8(5), 12–18.
6. Basak, D., Islam, M., Banerjee, S., & Bandyopadhyay, S. (2020). *Students career guidance system using classification algorithms*. Journal of Educational Technology, 8(2), 77–88.
7. Biswas, A., & Roy, A. (2019). *Career recommendation system based on student academic performance using machine learning*. International Journal of Emerging Technologies, 10(4), 88–94.
8. Chandrasekaran, P., & Ramakrishnan, S. (2021). Predictive modeling of student performance in higher secondary education. *Education and Information Technologies*, 26(4), 4321–4337.
9. Dahiya, P., & Arora, A. (2019). *Analyzing student academic performance using ensemble machine learning*. Journal of Big Data Analytics, 6(1), 51–60.
10. Gupta, R., & Kumar, G. (2022). *Evaluation of machine learning classifiers for student career path prediction*. International Journal of Computing Science, 15(1), 44–52.
11. Hussain, M., & Farooq, O. (2020). A machine learning approach for student academic prediction. *Journal of Applied Computing*, 10(14), 102–110.
12. Iqbal, Z., & Arif, M. (2021). *Intelligent career guidance using machine learning: A review and future directions*. Journal of Educational Computing Research, 59(6), 2115–2140.
13. Jain, S., & Verma, P. (2020). *Career counseling system using machine learning for high school students*. International Journal of Advanced Research in Computer and Communication Engineering, 9(3), 56–62.
14. Jadhav, S., & Patil, V. (2023). Prediction of suitable career stream for Class 12 students using supervised learning. *International Journal of Computer Science and Mobile Computing*, 12(7), 98–104.
15. Kaur, R., & Singh, K. (2019). *Machine learning based student performance and career guidance system*. Journal of Technical Education and Training, 11(1), 13–21.
16. Li, X., Liu, H., & Gao, Y. (2020). Predictive analytics for academic performance and future career selection. *Computers & Education*, 154, 103891.
17. Malik, H., & Khan, S. (2022). *Data mining techniques in career guidance system for adolescents*. International Journal of Computer Applications, 49(7), 29–39.
18. Manzoor, A. (2021). *Machine learning models for modeling student aptitude and choice*. Journal of Education and Learning, 10(3), 115–125.

19. Mishra, P., & Tripathi, N. (2020). Supervised learning approach for student course recommendation system. *International Journal of Machine Learning*, 5(2), 55–63.
20. Nahar, M., Rahman, M., & Hossain, M. (2021). A predictive model for student career orientation using classification approach. *Journal of Data Analysis and Information Processing*, 9(3), 83–96.
21. Parveen, R., & Saha, S. (2020). *Career recommendation system using artificial intelligence strategies*. International Journal of Intelligent Systems, 33(7), 1425–1439.
22. Rani, R., & Singh, P. (2019). Educational data mining for student performance and career analysis. *International Journal of Educational Research*, 98, 102–110.
23. Sharma, T., & Bhatia, M. (2021). Predicting academic success to recommend career disciplines. *International Journal of Data Analytics*, 8(4), 201–212.
24. Singh, A., & Kumar, S. (2022). *A comprehensive review on predictive analytics in education*. Journal of Artificial Intelligence in Education, 32(1), 41–73.
25. Sultana, R., & Islam, K. (2021). Machine learning empowered framework for student career selection. *Journal of IT Education and Research*, 15(2), 101–117.
26. Tiwari, R., & Prasad, K. (2019). *Career prediction system using predictive modeling*. International Journal of Pure and Applied Mathematics, 118(15), 3447–3462.
27. Yadav, V., & Pal, S. (2020). *Educational data mining for academic and career prediction*. IEEE International Conference on Education and E-Learning Technologies, 125–130.