Career Prediction Website Using Machine Learning

B Anitha, Adiba Fatima, Anshika Awasthi, Atifa Batool

[1]Assistant Professor, Department of IT Bhoj Reddy Engineering College for Women, India.

[2,3,4]B. Tech Students, Department Of IT, Bhoj Reddy Engineering College For Women, India.

ABSTRACT

In today's rapidly evolving educational landscape, students often face significant uncertainty when choosing suitable career paths after completing their higher secondary education. Traditional methods of career counseling rely heavily on subjective judgment, one-time assessments, or generic suggestions that fail to consider the unique capabilities, interests, and learning trajectories of each student. This project proposes the development of an intelligent, machine learning-based Career Guidance System designed to offer personalized and data-driven career recommendations. By leveraging supervised and unsupervised learning algorithms including Decision Trees, Support Vector Machines, and XG Boost the system evaluates academic records, skill assessments, and psychometric inputs to predict the most suitable academic streams and career domains. The system also incorporates a conversational chatbot to improve user interaction and adaptability, ensuring continuous learning and feedback. The proposed model aims to bridge the gap between student potential and career decision-making, offering scalable, real-time, and insightful guidance to empower students in making informed academic and professional choices.

Keywords*: Career Counseling, Machine Learning, Data-Driven, Decision Trees, Support Vector Machines, XGBoost, Academic Records, Skill Assessments, Job Prediction, Conversational Chatbot, Career Decision-Making.*

1. INTRODUCTION

Choosing a career goes beyond selecting a course , it involves understanding one's aspirations, strengths, and long-term goals. Career counseling plays a critical role in helping students recognize their potential and make informed academic choices. However, many students often make decisions based on external guidance from parents or teachers, which may not align with their true abilities. Career decision-making is a critical process for students, as it directly influences their academic growth, professional development, and long-term opportunities. Traditional career counseling methods often rely on subjective advice from parents or teachers, or on one-time assessments that fail to capture the dynamic nature of student learning. These methods are typically static, lack adaptability, and may not align with an individual's actual skills and interests.

Prior studies highlight that students encounter multiple barriers in career decision-making, including lack of motivation, insufficient occupational knowledge, and difficulty aligning academic performance with personal aspirations [2], [3]. To address these challenges, researchers have proposed technology-driven solutions, such as machine learning–based career guidance frameworks [1], [4], [5].

Machine learning provides significant advantages by enabling personalized, data-driven career recommendations. By analyzing academic records, psychometric test scores, and skill assessments, ML models can identify suitable

career paths more accurately than traditional approaches. This study introduces a Career Prediction Website that integrates multiple machine learning algorithms—Decision Tree, Support Vector Machine, and XGBoost—alongside a chatbot for interactive guidance. The system adapts dynamically to student profiles, ensuring scalable, reliable, and real-time career recommendations that bridge the gap between student potential and informed decision-making.

2-LITERATURE SURVEY

Rane et al. [1] developed a modular career prediction system using machine learning and web technologies such as Django and PostgreSQL. Their framework employed Decision Trees, K-Nearest Neighbors (KNN), and K-Means clustering, focusing on automation but lacking adaptability for dynamic student profiles.

Kulcsár et al. [2] analyzed the psychological barriers faced by students in making career decisions, emphasizing challenges such as lack of motivation and occupational knowledge. Similarly, Gati and Saka [3] designed the widely used Career Decision-Making Difficulties Questionnaire (CDDQ), which highlights the importance of psychological and personal factors in career guidance.

More recently, Majjate et al. [4] proposed an AI-powered academic guidance and counseling system that leverages machine learning techniques to deliver personalized career advice. Their system demonstrates how artificial intelligence can enhance scalability and interactivity in career counseling. Tahir et al. [5] developed a career counseling system using personality and interest assessments, showing how psychometric inputs combined with machine learning improve recommendation accuracy.

Our work builds upon these studies by integrating multiple machine learning algorithms with real-time adaptability through chatbot interaction, thereby addressing both the technical and psychological aspects of career decision-making.

3-METHODOLOGY

3.1 Dataset:

The dataset consists of student information, including academic performance, psychometric test scores, and skill-based assessments. Data preprocessing steps were applied to ensure quality and consistency. This involved:

- Removing incomplete or noisy records.
- Normalizing academic and test scores to a uniform scale.
- Feature selection to retain only variables most relevant to career prediction.

3.2 Algorithms:

To perform predictive modeling, three supervised learning algorithms were implemented. Each algorithm was selected for its complementary strengths in handling classification problems related to career prediction.

- ***Decision Tree (DT):***

Decision Trees are non-parametric supervised learning models used for classification and regression tasks. They

work by recursively splitting the dataset into subsets based on feature values, forming a tree-like structure. The decision rules generated at each node are easy to interpret, making DTs suitable when transparency is required in model decisions. However, they can be prone to overfitting, especially with small or noisy datasets, which limits their generalizability. Despite this, they provide a valuable baseline model due to their simplicity and interpretability.

- ***Support Vector Machine (SVM):***

  SVMs are powerful classifiers that aim to find an optimal hyperplane that maximally separates data points of different classes. They are particularly effective in high-dimensional feature spaces, where other algorithms may struggle. By using kernel functions, such as linear, polynomial, or radial basis function (RBF), SVMs can model complex non-linear decision boundaries. SVMs also generalize well to unseen data, which makes them suitable for predicting careers from diverse and multi-dimensional student datasets. The main challenge lies in their computational cost for large datasets and in selecting an appropriate kernel.

- ***XGBoost (Extreme Gradient Boosting):***

  XGBoost is an ensemble learning algorithm based on the gradient boosting framework. It builds multiple decision trees sequentially, with each tree correcting the errors of its predecessors. XGBoost introduces regularization techniques, such as L1 and L2 penalties, to prevent overfitting while maintaining high accuracy. It is highly efficient, scalable, and capable of handling missing values. Due to its robustness and superior performance in many machine learning competitions, XGBoost was chosen as the primary algorithm for this study. Its ability to capture complex relationships and interactions among features makes it particularly well-suited for adaptive career prediction tasks.

3.2 Tools and Technologies:

The system was implemented using:

- Python with libraries such as NumPy, Pandas, and Scikit-learn for data handling and modeling.
- SQLite for lightweight database management.
- ConferBot for chatbot integration to enhance user interaction.

4-DESIGN

System Architecture:

The purpose of the system architecture is to provide a structured framework that satisfies both the technical and operational requirements of the proposed career prediction website. It defines the overall organization of the system and ensures that each module functions cohesively. The architecture emphasizes the high-level structure of the system, describing how its components interact and exchange information.

The design follows a modular approach, consisting of data input, preprocessing, machine learning models, prediction generation, and chatbot integration. Each module is independent but connected, which allows for scalability and easier future enhancements. A flow diagram (Fig. 4.1) illustrates the architecture, showing how student data moves from collection and preprocessing to prediction and interactive guidance.

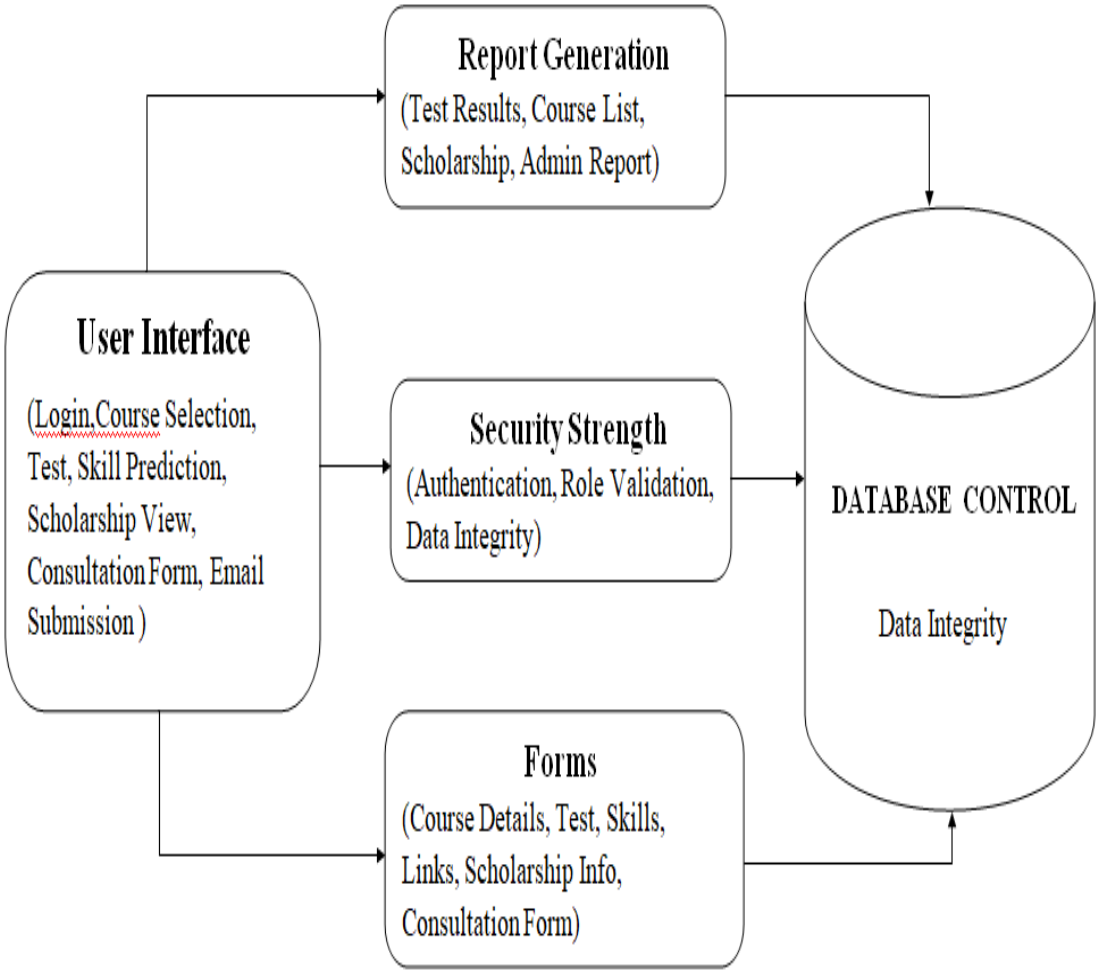


Fig. 4.1 System Architecture

Technical Architecture:

The technical architecture specifies the technologies and tools used to implement the system. The frontend of the website is built using HTML, CSS, and JavaScript to provide an interactive and user-friendly interface. The backend is developed in Python, integrating machine learning models through libraries such as NumPy, Pandas, Scikit-learn, and XGBoost. Data is stored and managed in an SQLite database, which ensures lightweight yet reliable performance. The machine learning pipeline includes data cleaning, normalization, and feature selection before model training and testing. The chatbot is integrated using ConferBot, enabling interactive communication and adaptive guidance for students.

This layered architecture separates concerns between frontend, backend, and database, while maintaining flexibility for integrating additional algorithms, larger datasets, and advanced chatbot features in the future.

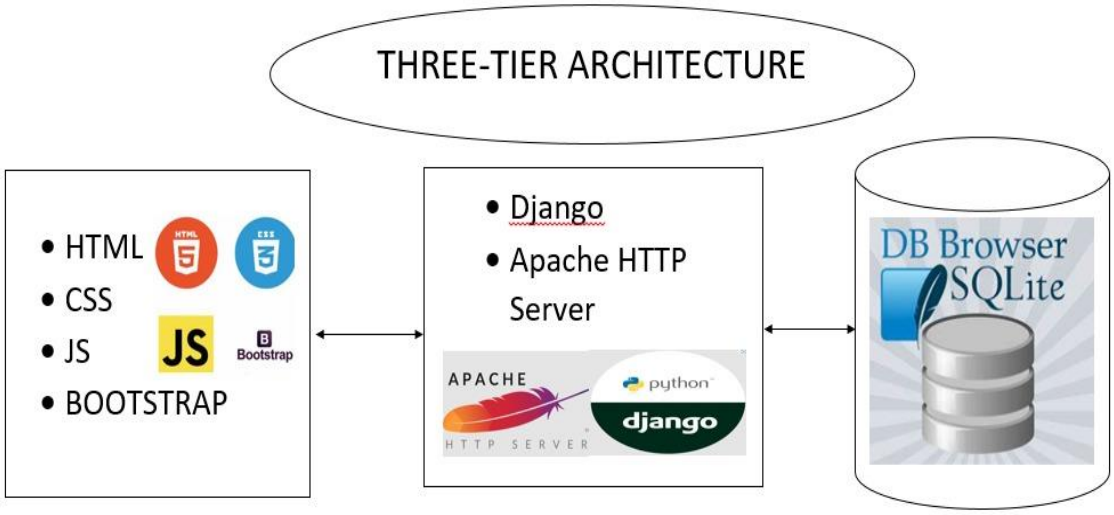


Fig. 4.2 Technical Architecture

5-RESULTS AND EVALUATION

5.1 Feature Selection

To enhance model performance and reduce dimensionality, two statistical methods were applied for feature selection:

Chi-Squared Statistic:

Evaluates the independence between features and the target variable.

Features with higher chi-squared values were considered more relevant.

This method was particularly effective for categorical variables, ensuring that only features with strong predictive relationships were retained.

Mutual Information Statistic:

Measures the dependency between features and the target variable.

Unlike chi-squared, it can capture non-linear relationships, making it useful in identifying features with subtle but meaningful contributions to prediction accuracy.

The combination of these two methods allowed us to retain the most informative features while discarding redundant or irrelevant ones, ultimately improving model training efficiency and performance.

5.2 Performance of Algorithms

After applying feature selection, the three supervised learning algorithms—Decision Tree (DT), Support Vector Machine (SVM), and XGBoost—were implemented and evaluated. Their performance was measured using accuracy, precision, recall, and F1-score.

Decision Tree (DT): Achieved an accuracy of 50%. While highly interpretable, the model showed overfitting tendencies and lower generalization capability.

Support Vector Machine (SVM): Reached an accuracy of 70%, outperforming DT by achieving better classification margins and fewer misclassifications.

XGBoost: Delivered the highest accuracy of 93.33%, demonstrating robustness through boosting and regularization, making it the most reliable of the three models.

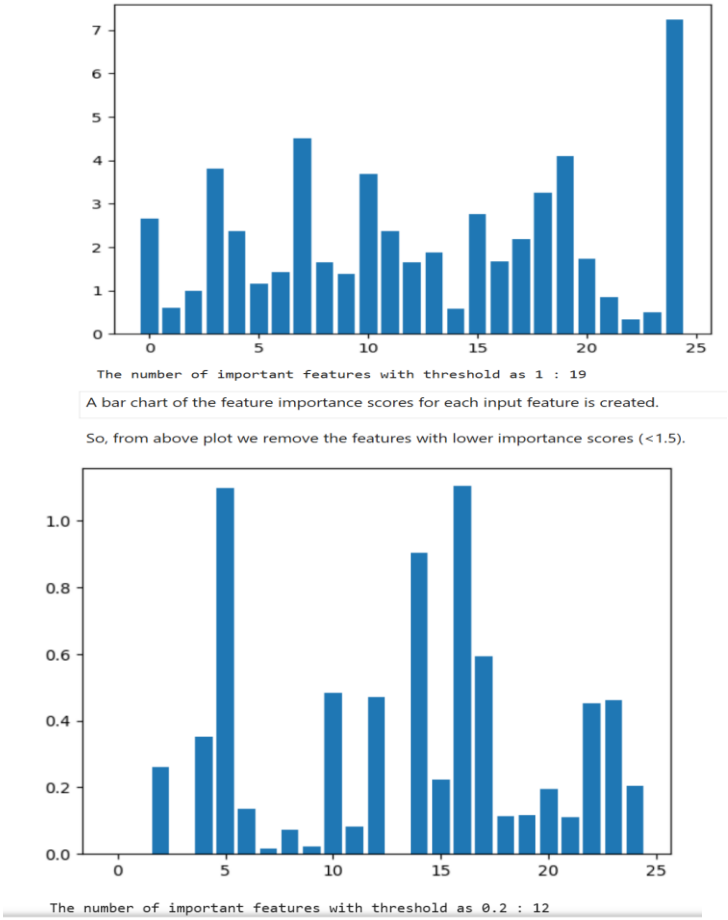
The comparative study revealed that:

- XGBoost consistently outperformed both DT and SVM across all metrics.
- SVM achieved balanced performance with better generalization than DT.
- Decision Tree remained the most interpretable model but was less accurate.

Performance ranking: **XGBoost > SVM > Decision Tree.**

5.3 Visualization of Results

Figure 5.3.1 : Feature importance plot highlighting which selected features contributed most to predictions.



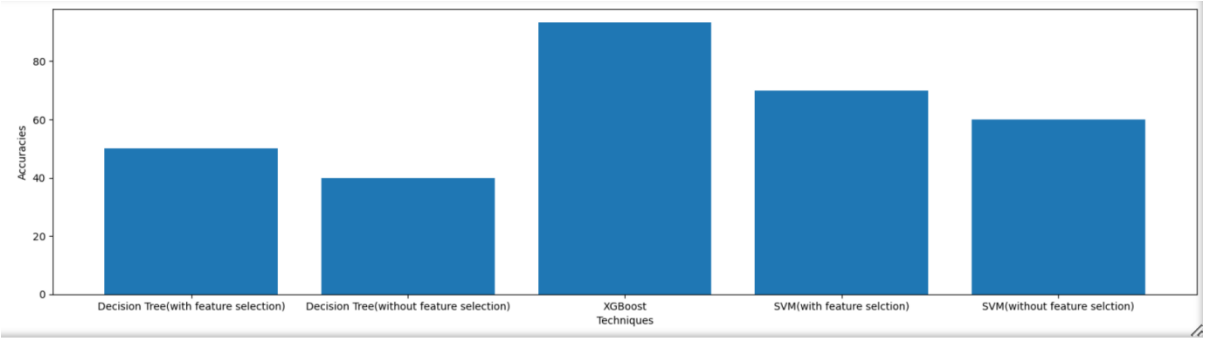


Figure 5.3.2: Bar chart of accuracy scores for DT, SVM, and XGBoost.

5.4 Evaluation Summary

The evaluation highlights that:

- Feature selection using Chi-Squared and Mutual Information successfully reduced irrelevant features while retaining the most predictive ones.
- Among the algorithms, XGBoost demonstrated superior performance with strong robustness and accuracy.
- SVM offered a good trade-off between performance and generalization.
- Decision Tree was interpretable but less effective in predictive accuracy.

6. DISCUSSION

The results show that ensemble methods such as XGBoost are more effective for career prediction tasks compared to traditional classifiers. Unlike prior works [1], which lacked adaptability, our system introduces continuous learning and chatbot-driven interaction. The inclusion of feature selecting methods also improved accuracy. Limitations include dataset size and generalizability across diverse student population.

7-CONCLUSION

In a rapidly evolving educational landscape where students are often overwhelmed with choices, our proposed machine learning-based career guidance system serves as a timely and transformative solution. Traditional career counseling methods, though well-intentioned, lack the depth, personalization, and adaptability needed to align academic choices with individual aspirations and abilities. Our project addresses this critical gap by employing advanced predictive modeling techniques, including Decision Tree Classifiers, XGBoost, and Support Vector Machines, to deliver data-driven, reliable, and dynamic career recommendations. By offering real-time adaptability, the system empowers students to make better-informed academic and career decisions. Future work will involve expanding the dataset, integrating more psychometric features, and deploying the system on a larger scale for broader applicability.

Ultimately, this intelligent career guidance platform empowers students to discover their true potential, confidently pursue appropriate engineering streams, and prepare for both national and global career opportunities. Through its innovative use of AI and user-centric design, this project marks a significant step forward in redefining how students plan their futures.

REFERENCES

[1] M. Rane, T. Kakkad, S. Kalal, V. Jain, J. Chandegara, and S. Jagtap,
"Career Prediction Website using Machine Learning," in *Proc. 2023 3rd Int. Conf. Intelligent Technologies (CONIT)*, 2023, pp. 1–6. IEEE.

[2] V. Kulcsár, A. Dobrean, and I. Gati,
"Challenges and difficulties in career decision making: Their causes, and their effects on the process and the decision," *Journal of Vocational Behavior*, vol. 116, 2020, Art. no. 103346.

[3] I. Gati and S. Saka,
"High school students' career-related decision-making difficulties," *Journal of Counseling & Development*, vol. 79, no. 3, pp. 331–340, 2001.

[4] H. Majjate, Y. El Younoussi, K. Berrada, and O. Moussaoui,
"AI-Powered Academic Guidance and Counseling System," *AI*, vol. 4, no. 1, pp. 81–94, 2023.

[5] M. Tahir, A. Mustafa, and H. A. Rauf,
"Career Counseling System for Students Based on Personality and Interest Assessment Using Machine Learning," *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 12, no. 7, pp. 143–150, 2021.