Formosa Journal of Science and Technology (FJST)
Vol. 4, No. 4, 2025: 1159-1176

A Predictive Model for Academic Major Selection Using AI and Labor Market Trends

Yosra Abdullah Salem Elewa[1], Mohammed Fadhl Abdullah[2*]
Faculty of Engineering & Computing, University of Science and Technology
Corresponding Author: Mohammed Fadhl Abdullah, m.albadwi@ust.edu

ARTICLEINFO

Keywords: Artificial Intelligence, Machine Learning, Recommendation System, Major Selection, Academic Guidance

Received : 29, March
Revised : 12, April
Accepted : 29, April

ABSTRACT

Selecting the right university major is a crucial decision that influences students' academic success and future careers. This study presents an AI-driven recommendation system using supervised machine learning, focusing on the Random Forest classifier. The system is trained on the "Arab University Graduate Data Set" (1,000 records), including features such as high school GPA, entrance exam scores, and employment rates. Hyperparameter tuning improved model performance, achieving 97.5% accuracy. Feature importance analysis highlighted GPA and employment rate as key factors. Unlike prior work focused on developed regions, this study explores AI's potential in Yemeni universities with limited guidance resources. Findings support the use of AI tools to align educational decisions with labor market needs, improving student outcomes.

ISSN-E: 2964-6804
DOI: https://doi.org/10.55927/fjst.v4i4.55
https://traformosapublisher.org/index.php/fjst

INTRODUCTION

Artificial intelligence (AI) has increased rapidly in recent decades, changing different aspects of society, including medicine, robotics, education, and autonomous systems. The AI field is experiencing a revival from its emergence during the 1950s, which is inspired by progress that enables computer systems to learn and data-driven decisions autonomously(Elewa et al., 2025). The rapid development of AI has created considerable interest in potential applications and various domains (Fadhl & Elewah, 2024; Naher et al., 2023).

In education, AI technologies change teaching and learning methods. The AI-driven recommendation system draws attention to their ability to help with informed educational alternatives by analyzing historical data, student performance matrixes, and employment trends. The use of the recommended system (RS) in previous years has become more widespread (Dascalu et al., 2016). These systems benefit from advanced machine learning algorithms to generate individual recommendations, and reduce the dependence on arbitrary or biased guidance from traditional sources (Alsayed et al., 2021; Tarus et al., 2018). AI-based technologies facilitate student interactions, enrich learning experiences, and affect students' success in higher education (Abdullah et al., 2025; Iqbal et al., 2022).

AI can be classified into different types: machine learning, natural language processing, robotics, and expert systems. Of these, machine learning techniques, which Random Forest (RF) classify, gained popularity because of their accuracy and ability to treat complex computer structures (Zayed et al., 2022). RF has been used effectively in a future-state model that helps students' alternatives adapt to market trends, ensuring better career prospects after graduation (Alnomay et al., 2024).

Despite the progress of AI-based education councils, most research focuses on developed countries where students have access to structured career counseling resources. However, career counseling services are limited or disabled in Yemen and other development sectors, making the most important selection even more challenging. Existing studies have not conducted a comprehensive study of AI's ability to support students at Yemeni universities. This research corresponds to the educational landscape of Yemen, especially the educational landscape by developing an AI manual recommendation system.

Choosing a university major is one of the most important decisions of students, which affects their academic and professional future. Many high school graduates struggle with selecting the right field of study due to factors such as lack of experience, inadequate knowledge about different disciplines, and external influences like family and societal expectations (Alghamdi et al., 2019). In addition, students are often dependent on subjective recommendations from colleagues and internet findings, which do not always match their abilities and career ambitions (Zayed et al., 2022).

The main objective of this study is to develop an AI-based recommendation system that helps students choose the most suitable university major based on academic performance, career prospects, and market requirements. The research also aims to assess the accuracy and reliability of machine learning models in academic advising. The research questions are as follows:

- What are the key factors influencing students’ selection of a university major?
- How effective is the Random Forest algorithm in predicting suitable majors for students?

The findings of this study hold significant implications for students, educators, and higher education policymakers. By implementing AI-driven recommendation systems, universities can increase their academic counseling services, reduce student dropout rates, and improve overall satisfaction in academic decision-making (Alsayed et al., 2021; Hammoudi Halat et al., 2023). This study is particularly relevant to Yemeni universities, where limited career guidance resources often result in mismatched student specializations, leading to inefficiencies in higher education (Mengash, 2020). A well-structured AI-based system can provide valuable insight into coordinating students' abilities with labor market needs and eventually contributing to national economic growth (Tang et al., 2024).

The remaining part of this paper is structured as follows: Section II provides literature reviews, AI-based Education Council discusses relevant studies and existing methods. Section III emphasizes the research method, and describes data preparation, feature choices, and development of machine learning models. Section IV presents experimental results and evaluates the performance of the proposed system. Section V AI-inaccurate discusses findings for academic advice and their implications. Finally, section VI concludes the study with key findings and recommendations for future research.

THEORETICAL REVIEW

Choosing a suitable university major is an important decision that significantly affects education success and career path. Various factors, including academic performance, career ambitions, and labor market trends, affect this option. Traditional counseling methods often rely on subjective assessment and may lack scalability. To address these limitations, researchers have discovered Artificial Intelligence (AI) and Machine Learning (ML) as promising solutions for academic advice. AI-driven recommendation systems utilize historical educational data, student characteristics, and labor market insights to generate personalized academic guidance, reducing reliance on conventional counseling approaches (Tarus et al., 2018; Alsayed et al., 2021).

The recommended system has been widely explored in education to assist students in selecting suitable majors. Studies have shown that machine learning models, such as decision trees and collaborative filtering, effectively analyze student profiles and academic history to provide well- informed suggestions (Alshaikh et al., 2021). Hybrid approaches, which combine many techniques, have been found to increase the accuracy of predictions (Batmaz et al., 2019).

Several studies have highlighted the effectiveness of machine learning algorithms in predicting suitable university majors for students. Random Forest (RF) is especially known for its better classification performance. (Alsayed et al.,

2021) found that RF improved accuracy and other classification models in accuracy and generality when predicting optimal educational routes based on previous performance and career opportunities. Similarly, (Zayed et al., 2022) RF achieved the highest accuracy (97.70%) compared to the decision trees (DT) and Support Vector machines (SVM) while predicting the most important alternatives to students based on educational registers and degree of employment. These findings support the use of an ML-controlled decision structure in the Academic Council.

Deep learning models have also been detected in career recommendations.(Bahalkar et al., 2024) introduced an AI-driven career guidance system using an Editorial Performance and aspirations to predict the recommendations of the student subject. His study highlighted the importance of personalization in career consultation, showing that AI-powered recommendations have been closely aligned with traditional methods, while increased adaptability is offered. Data was also addressed to ensure fairness in educational decision-making, including data privacy and prejudice mitigation.

Machine learning techniques have been extensively studied for academic and career advising. (Siswipraptini et al., 2024) developed a personal career path Recommended model for IT students using the Naive Bayes algorithm, which integrates job market data with students' interests to improve the recommended accuracy. The study found that over 83% of users were satisfied with the model’s recommendations. Similarly,(Kim & Lim, 2023) examined intelligent recommendation systems for non-ICT students, and discussed content-based filtration, collaboration filtration, and hybrid models. Their findings emphasize AI's role in improving future accuracy and addressing the problems of colds in recommendations.

Beyond academic advising, AI has also been applied to enhance student well-being. (Demong et al., 2023; DS & K, 2015) detected the impact of personality symptoms on the social well-being of students, using machine learning classification techniques to identify factors affecting welfare levels. Their findings outline the importance of integrating personality-based recommendations Educational decision-making system.

Educational data mining (EDM) techniques have been employed to predict student performance and educational success. (Tripathi et al., 2024) used machine learning models to analyze academic achievement tests, admission scores, and early-stage course performance to identify students at risk of underperformance. Their findings indicate that early intervention strategies based on AI-driven insights can significantly reduce failure rates. (Cheng, 2024) introduced a Bayesian classifier-based personalized recommendation system for university entrepreneurship education, integrating textual, visual, and behavioral data to enhance learning experiences.

The use of logistic regression in academic advice was valid by previous studies (Chai & Gibson, 2015; Mason et al., 2018). However, more recent studies advocate the superiority of the Random Forest algorithm in prediction accuracy (Falát & Piscová, 2022; Hung et al., 2019; Lottering et al., 2020). Furthermore, (Alyahyan & Düştegör, 2020; Ojajuni et al., 2021) demonstrated that machine-learning models consistently outperform traditional statistical methods in predicting student success.

Despite advancements in AI-driven academic advising, most research has focused on developed countries where structured career counseling services are available. Studies have acknowledged that existing AI-based models often lack adaptability to diverse educational environments, particularly in regions with limited career guidance resources (Hammoudi Halat et al., 2023; Kamal et al., 2024). Additionally, many existing models may exhibit biases due to the underrepresentation of key factors such as socioeconomic background, student interests, and regional job market variations.

While AI-based recommendation systems have been successfully implemented in higher education, research on their application in developing countries, particularly in Yemen, remains scarce. Current studies do not comprehensively integrate labor market trends with academic performance to produce holistic and well-rounded recommendations. Moreover, potential biases in AI-driven academic advising systems require further investigation to ensure fairness and inclusivity.

This study addresses these gaps by developing an AI-driven recommendation system tailored to the Yemeni educational landscape. Unlike prior research, this work integrates both academic indicators and real-world employment trends to enhance recommendation accuracy. In addition, it investigates potential biases in AI-based advising and proposes improvements to ensure fairness and efficiency in academic decision-making. By leveraging machine learning techniques, particularly Random Forest, this research contributes to the development of a data-driven framework that can support students in making well-informed and strategic decisions about their university majors.

METHODOLOGY

This study employs a quantitative and experimental research approach using machine learning techniques to analyze historical student data and labor market trends. The primary goal is to develop an AI-driven recommendation system that predicts the most appropriate capital for high school students based on academic results and employment results from graduates. The methodology includes several stages, including data collection, preprocessing, feature engineering, model training, and evaluation, to ensure the accuracy and efficiency of the system.

Data Collection and Preparation

The dataset used, "Arab Universities Graduate Dataset," includes 1,000 student records, including large educational and employment functions such as GPA in high school, entrance exams, GPA, employment status, and main-specific employment rate. These features were selected based on their possible impact on larger selections and career success. Given the study’s quantitative nature, all data were numerically structured to enable efficient machine-learning analysis.

To increase the relevance of the recommendations, the dataset was refined by emphasizing academic performance indicators while ensuring that labor market trends were also taken into account. This approach allows the system to offer data-driven guidance to high school students, aligning their academic strengths with career prospects. Table 1 emphasizes the main features included in the dataset, highlighting the factors considered in building the recommendation model.

Table 1. Dataset attributes description

No	Feature Name	Description	Data Type
1	Gender	Student gender (Male = 1, Female = 0)	Categorical
2	HSGPA	High school GPA	Numerical
3	HSType	High school type (Literary = 0, Scientific = 1)	Categorical
4	EEScore	Entrance exam score	Numerical
5	UGPA	Undergraduate GPA (average GPA in the undergraduate studies)	Numerical
6	EmpStat	Employment status after graduation (1 = employed, 0 = not employed)	Categorical
7	UMajor	Undergraduate major (Medicine = 0, Engineering = 1, Computer Science = 2, Business = 3)	Categorical
8	AvgUGPA_Major	Average undergraduate GPA for the major (calculated from UGP data)	Numerical
9	EmpRate_Major	Employment rate for the major after graduation (calculated from EmpStat data)	Numerical

This dataset was chosen for this research, due to its public access and unique inclusion of both students' academic performance and employment rates for each chosen Major. By incorporating these two critical factors, the dataset provides valuable insights into relation to academic performance and labor market results, which improves the recommendation system's ability to coordinate the students' educational alternatives with future career opportunities.

Data Processing and Visualization

At this stage, the dataset was cleaned and processed to prepare it for analysis and machine learning functions. Several preaching techniques were implemented using special Python libraries. The following steps were performed:

Handling Missing Values

In some fields, Missing values were found particularly in employment status. Since some graduates were not employed, missing values in employment-related fields were properly controlled to maintain data integrity.

Encoding Categorical Features

To ensure compatibility with machine learning models, categorical features were transformed into numerical values to predict the most suitable undergraduate major for students. The encoding scheme is detailed in Table 2:

Table 2. Label Encoding Conversions

Feature Name	Feature Labels
Gender	Female (0), Male (1)
HSType	Literary (0), Scientific (1)
UMajor	Medicine (0), Engineering (1), Business (2), Computer Science (3)
EmpStat	Unemployed (0), Employed (1)

Feature Engineering

Additional features were derived to enhance predictive accuracy including AvgUGPA_Major, which represents the average undergraduate GPA per major, and EmpRate_Major, which indicates the employment rate per major. These features provide deep insights into relationships between educational performance and job market outcomes, which improve the ability of the recommended system's ability to predict the most suitable undergraduate major for students.

Data Standardization

To enhance model performance and ensure all numerical features have a mean of zero and a significance of standard deviations of one, the dataset was standardized using standards according to the following equation (1):

$$\frac{x_i - \mu}{\sigma} = {}'x \qquad \textit{Equestion (1)}$$

Where 'x is the standardized value, xi is the original feature value, μ is the mean of the feature, and σ is the standard deviation.

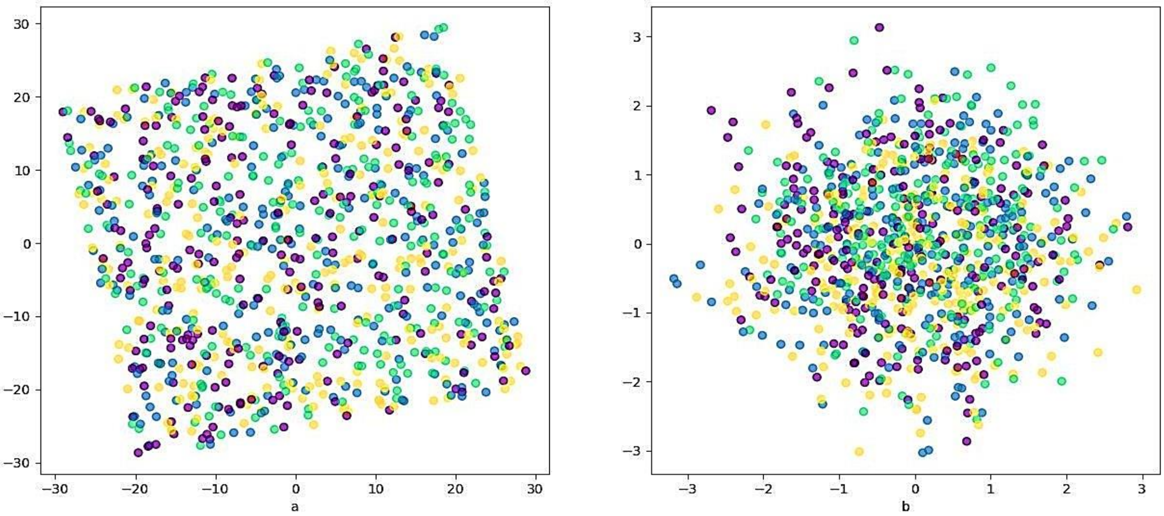

Figure 1: Effect of Standardization on the Dataset: (a) Before standardization; (b) After standardization.

The impact of standardization on the dataset is depicted in Figure 1. The left conspiracy in Figure 1a represents the raw Dataset before standardization, where features display different scales and variances. The data points are unevenly distributed, making comparing the distribution of features difficult. After implementing the standards, the right plot in Figure 1b illustrates the dataset, which converts the features of a mean of 0 and a standard deviation of 1. This change ensures that all features are on a comparative scale, preventing some individual variables from affecting models and increasing the stability of machine learning algorithms.

Data Splitting

The dataset containing 1000 samples was divided into training and test sets using the 80:20 ratio. In particular, 800 samples were trained, while the remaining 200 samples were used for testing. This ensures a sufficient amount of data for model training while burning a part to evaluate normalization performance.

Data Visualization

Data provides valuable insights on quality, trends, and functions, and helps choose the right model. Figure 2 presents a correlation matrix for all the features, which indicates that there is no strong coplanarity between the input variables. This scene highlights the relationships between different features, which helps identify the most impressive variables for making recommended models. Understanding these correlations improves the model's ability to coordinate education with the requirements of the labor market, and ensures a diverse and reliable data set.

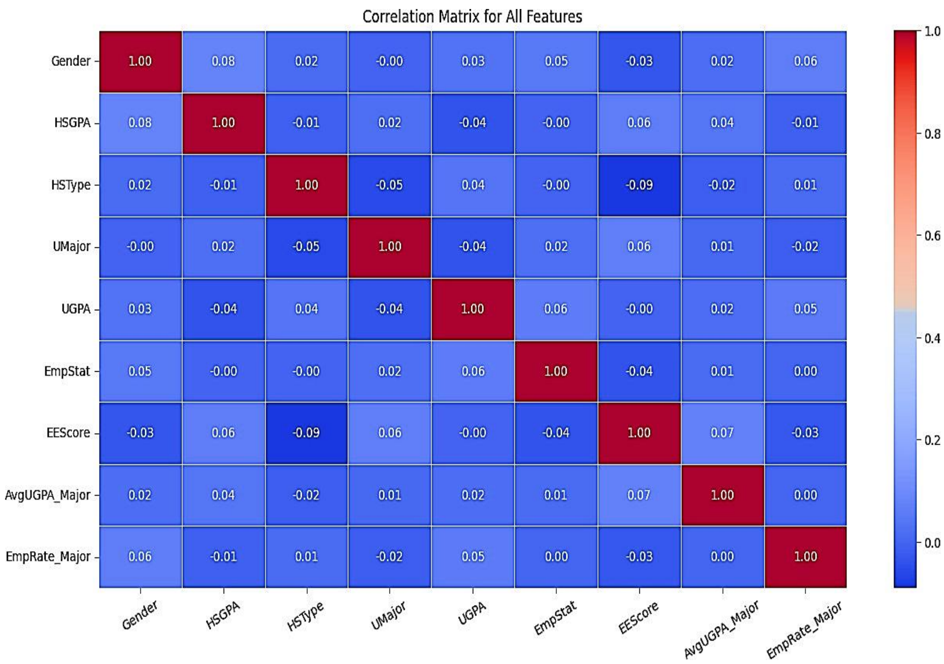

Figure 2. Correlation Matrix for all features

Model Classification and Tuning

This study employed the Random Forest (RF) algorithm to classify students' majors based on various features. The model's hyperparameters were perfect for increasing performance. Unlike GridSearchCV, which is often used to customize hypermeters, this study classification depends on the manual setting to adjust the classification accuracy and main parameters such as n-estimators and max-depth.

Random Forest is a supervised machine learning classifier that produces several decisions during training and makes predictions based on majority votes from all trees. It is recognized for its efficiency in classification and regression tasks, strength towards the outside, the ability to handle high-dimensional data sets, and a low trend to overfit compared to other algorithms (Zayed et al., 2022). The most important hypermeters used in random forest algorithm for this study include criteria, measuring the quality of division of each node; Max-depth, which controls the complexity of the model and reduces overfitting; Min-samples-leaf, which determines the minimum number of required samples in the leaf node; Min-samples-Split, which determines the minimum number of minimum samples required to share the internal node; And n-estimators, which define the number of trees in the forest, where this price usually increases performance, but at a high computational cost.

To evaluate the efficiency of the model, various performance metrics were used, including accuracy, precision, F1-score, and confusion matrix. These metrics provide a comprehensive understanding of the model's ability to classify respective majors properly. To increase the model's normality and reduce the risk of overfitting, cross-validation was used during the training process. This technique involves dividing training data into several subsets, where each subset is used for verification in various repetitions. This emphasizes that the model is tested on diverse data samples, resulting in more reliable and strong performance.

By choosing hyperparameters carefully and incorporating cross-validation, random classification demonstrated the strong ability of the student to classify student majors based on the features provided by random classification. These stages helped to ensure the reliability and purpose of the model in the real world's educational and career guidance scenarios to classify their respective majors correctly.

RESEARCH RESULTS

A Random Forest classifier was used in the first experiment of this study without any hyperparameter adjustments. Training-to-test ratios of 80:20 and 70:30 were used in this experiment twice.

As shown in Table 3, the Random Forest classifier has 97.5% accuracy with an 80:20 Training testing ratio and 95.6% accuracy with a 70:30 ratio. Although these high accuracy points suggest strong model performance, they may also indicate possible overfitting for training data. To improve generalization and ensure the strength of the model, techniques such as cross-validation and feature importance analysis should be considered.

Table 3. Comparison of accuracy results between different Training Ratio

ML Model	30:70 Testing: Training Ratio	20:80 Testing: Training Ratio
Random Forest Classifier	95.6%	97.5%

The confusion matrix in Figure 3 reflects the classification performance of our recommendation model across four fields: Medicine, Engineering, Computer Science, and Business. The results indicate high classification accuracy with minimal abortion. In particular, the model correctly classified 39 Medicine students, 50 Engineering students, 54 Computer Science students, and 52 Business students.

A smaller abortion is seen in the computer science category, where 5 students were incorrectly classified as medical students. Despite this, the model still shows strong predictive capabilities, achieving high accuracy in classifying student majors. This assessment illuminates the reliability of the model, although further processing, such as feature selection or hyperparameter tuning, can increase performance.

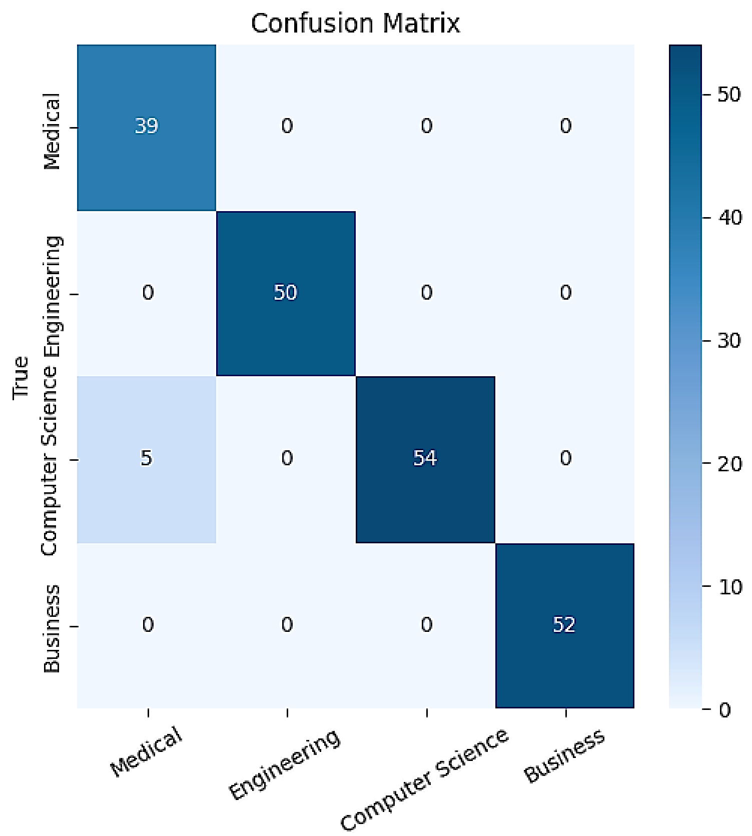

Figure 3. confusion matrix

To assess the classification performance of the model, Table 4 presents precision, recall, and F1- score metrics for each major. The findings imply that the model performs strong classification performance in different fields, with an F1 score between 0.94 and 1.00. The model exhibits near- perfect classification ability and emphasizes the accuracy of distinguishing students between different academic disciplines. In particular, engineering and computer science areas achieved a correct score of 1.00 in all metrics, while the business was recalled 0.92, suggesting a minor trend for error classification in this category.

In addition, Table 5 model performance offers a summary of the calculations, where the model's overall accuracy was 0.97, demonstrating strong and reliable future capabilities. The exact, recall and F1 scores macro and weighted averages were extraordinarily high (0.97 and 0.98), above, and showed the models' ability to effectively generalize in different student groups. This high level of performance confirms that the model is effectively trained and correct, making it a reliable tool to guide high school students in choosing their academic majors based on available data.

Table 4. Classifications Performance by Major

Class	Precision	Recall	F1-Score	Support
Medicine	0.89	1.00	0.94	39
Engineering	1.00	1.00	1.00	50
Business	1.00	0.92	0.96	59
Computer Science	1.00	1.00	1.00	52

Table 5. Model performance Metrics

Metric	Precision	Recall	F1-Score	Support
Accuracy	-	-	0.97	200
Macro Avg	0.97	0.98	0.97	200
Weighted Avg	0.98	0.97	0.98	200

As shown in Table 6, the consequences of the hyperparameter tuning process to customize the random forest model were the optimal criteria for the division (criteria) "entropy", as it took a better decision in dividing the data compared to "Gini". When it comes to the depth of the tree, the best- performing value was found to be 5, which effectively captures the required pattern in the data while preventing overfitting. A deep tree can necessarily increase complexity without improving classification performance.

For min-samples-leaf parameters, the optimal value was determined to be 10, ensuring that each leaf node consists of a sufficient number of samples to maintain generalization and avoid highly specific patterns. Similarly, the best value for min-samples-split was 10, which shows that a division is most effective when at least ten samples are available. This setting prevents unnecessary division and increases model stability. Finally, the optimal number of trees (N-estimators) was found 50, which balances computational efficiency with classification accuracy. The small number of trees reduces training time, while still providing a strong future performance. These customized hyperparameter values suggest that the model has been that the model has been fine-tuned to achieve high accuracy while maintaining the ability to normalize ignorant data effectively.

Table 6. Optimized Hyperparameters for Random Forest

Hyperparameter Name	Hyperparameter Values	Hyperparameter Optimal Value
criterion	['gini', 'entropy']	entropy
max_depth	[None, 5, 10, 15]	5
min_samples_leaf	[1, 5, 10]	10
min_samples_split	[2, 5, 10]	10
n_estimators	[50, 100, 200]	50

A necessary method of evaluating the effectiveness of machine learning model is the receiver operating characteristic (ROC) curve. This 2D ROC curve represents the relationship between false positive rate (FPR) and true positive rate (TPR), which provides insight into classification ability for model.

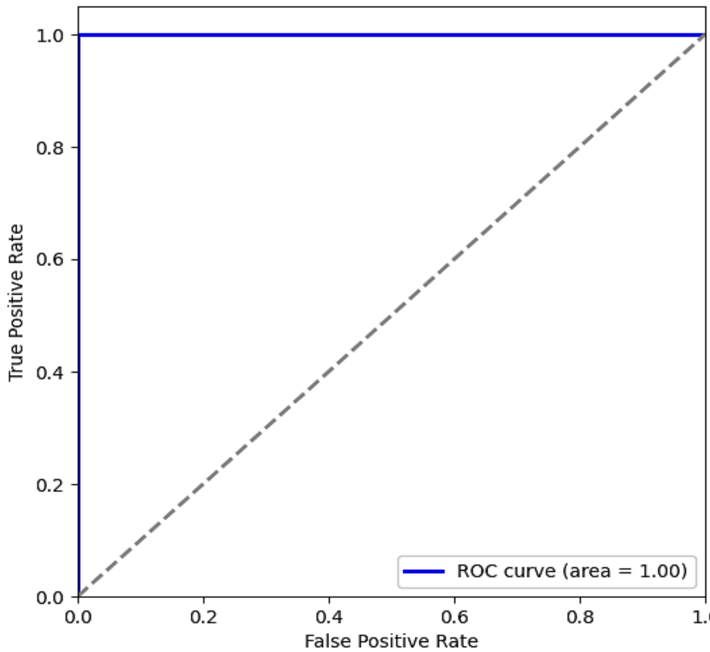

Figure 5. Receiver Operating Characteristic (ROC) curve

As depicted in Figure 5, when the ROC curve upper left corner, indicating an ideal classification with an Area Under the Curve (AUC) of 1.00. This result suggests that the model achieves 100% accuracy in distinguishing between different classes, as clarified by a straight horizontal line at TPR = 1.0. The diagonal dashed line represents a random classification, and the fact that the model's curve lies entirely above it confirms its superior predictive performance.

Although these extraordinary results highlight the model's strong discriminatory ability, it also increases concern about the presence of potential overfit or very different datasets. Further recognition is recommended by using the external dataset, to ensure the strength of the model and the real world.

To investigate the importance of the input variables for the model, the feature importance plot, as illustrated in Figure 6, was utilized. This plot ranking has decreasing order according to their contribution to the model's decision-making process. The results suggest that the top-ranked variables have a significant impact on the predictions of the model compared to lower-ranked ones.

The plot shows that the employment rate in the undergraduate major is the most effective factor ever, and plays an important role in predicting students who are recommended. This suggests that the results of the careers associated with different undergraduate programs heavily influence the recommendation process. Other features, such as high school type, employment status, and average undergraduate GPA, are relatively small contributions but still play a role in shaping predictions. Contrary to this, factors such as high school GPA, gender, and undergraduate GPA indicate minimal impact, suggesting that removal of them cannot affect the model's future accuracy. This observation emphasizes the ability of convenience choices to improve the efficiency of the model and at the same time maintain the strength.

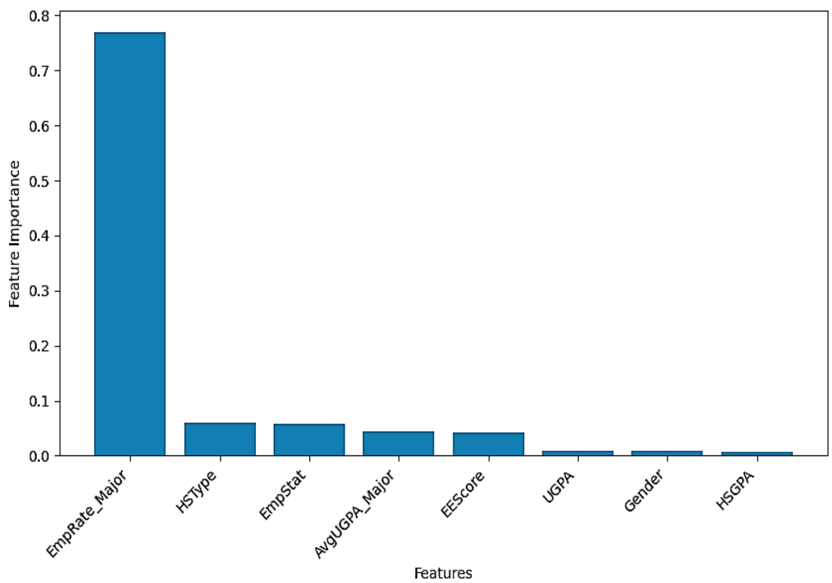

Figure 6. Importance Chart for Input Features

DISCUSSION

The findings from this study show that the developed AI-driven recommendation system for undergraduate major selection exhibits high classification accuracy in all tested scenarios. Random Forest Classifier achieved a total accuracy of 97%, with high precision, recall, and F1 scores, demonstrating its effectiveness in predicting students' recommended majors. The confusion matrix further supports these findings and shows minimum misclassifications across the four categories: Medicine, Engineering, Computer Science, and Business. Although this high performance suggests that the model is effective, it also raises concerns regarding potential overfitting. The strong classification ability of the model can be attributed to an overly separable dataset or excessive reliance on specific features, such as employment rate and academic performance, which strongly correlate with major selection. This assumption is further reinforced by the feature importance analysis, which identified employment rate in a major (EmpRate_Major) as the most influential factor, followed by high school type (HSType) and employment status (EmpStat). Conversely, attributes like gender and high school GPA (HSGPA) had minimal impact on the predictions of the model, suggesting that the model primarily relies on career-related and academic factors.

In addition, the ROC curve analysis confirmed the ability of the model to distinguish between different majors with an AUC of 1.00 for all categories. Although this result indicates strong classification performance, it also suggests the possibility of overfitting. To enhance the model's generalization ability, future work must consider starting more different datasets to capture a wider range of student profiles, applying regularization techniques to prevent overfitting, and incorporating external validation to assess its performance on unseen data. These improvements will guarantee that models are strong and useful in real-world scenarios.

CONCLUSION AND RECOMMENDATION

This study developed an AI-driven recommendation system to support high school graduates in choosing the most suitable undergraduate major, leveraging machine learning techniques with a primary focus on the Random Forest classifier. The proposed model achieved a classification accuracy of 97.5%, outperforming traditional classification techniques and demonstrating its robustness in predicting academic pathways. Compared to existing approaches, this study integrated both academic performance indicators and labor market trends, ensuring a data-driven and employment-oriented recommendation system. Hyperparameter tuning played an important role in enhancing the model's performance, optimizing classification accuracy, and reducing predicting errors. Feature importance analysis has shown that the employment rate within a major and high school GPA was the most effective, which guides the recommendation process. In addition, data preprocessing techniques, including normalization and encoding, significantly contributed to the improvement of the model's efficiency and reliability.

The Findings emphasize the ability of AI-based education councils, especially in areas such as limited career advisory resources such as Yemen. However, there are several challenges, including dataset boundaries and capacity for model overfitting. Future research should detect ways to learn further machines, such as deep learning techniques so that classification accuracy can be further improved. In addition, expanding the dataset, including real-time labor market trends, and incorporating students' interests and socioeconomic factors will increase the adaptability and fairness of the system. Ultimately, this study highlights the importance of AI-driven decision support systems in bridging the gap between education and employment, providing students to adapt to the success of their education and career.

FURTHER STUDY

Future research could examine the use of deep learning techniques to improve prediction accuracy and reduce overfitting. In addition, the development of a broader dataset with the integration of real-time labor market data will increase the relevance of recommendations. Customization of the system by considering students' interests and socioeconomic factors is also important to create more personalized and fair recommendations. Validation of the model across different geographical contexts is necessary to improve generalizability. The development of web-based or mobile applications can expand user access, and longitudinal evaluation is needed to measure the long-term effectiveness of this recommendation system.

REFERENCES

Abdullah, Y., Elewa, S., & Abdullah, M. F. (2025). *Assessing Awareness and Usage of ChatGPT among Lecturers and Students : A Case Study of the University of Science & Technology , Aden , Yemen Assessing Awareness and Usage of ChatGPT among Lecturers and Students : A Case Study of the University of Science*. *31*(1), 425–439.

Alghamdi, S., Alzhrani, N., & Algethami, H. (2019). Fuzzy-Based Recommendation System for University Major Selection. *International Joint Conference on Computational Intelligence*, *1*(Ijcci), 317–324. https://doi.org/10.5220/0008071803170324

Alnomay, I., Alfadhly, A., & Alqarni, A. (2024). A Comparative Analysis for GPA Prediction of Undergraduate Students Using Machine and Deep Learning. *International Journal of Information and Education Technology*, *14*(2), 287–292. https://doi.org/10.18178/ijiet.2024.14.2.2050

Alsayed, A. O., Rahim, M. S. M., Albidewi, I., Hussain, M., Jabeen, S. H., Alromema, N., Hussain, S., & Jibril, M. L. (2021). Selection of the right undergraduate major by students using supervised learning techniques. *Applied Sciences (Switzerland)*, *11*(22). https://doi.org/10.3390/app112210639

Alshaikh, K., Bahurmuz, N., Torabah, O., Alzahrani, S., Alshingiti, Z., & Meccawy, M. (2021). Using Recommender Systems for Matching Students with Suitable Specialization: An Exploratory Study at King Abdulaziz University. *International Journal of Emerging Technologies in Learning*, *16*(3), 316–324. https://doi.org/10.3991/ijet.v16i03.17829

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-0177-7

Bahalkar, P., Peddi, P., & Jain, S. (2024). AI-Driven Career Guidance System: A Predictive Model for Student Subject Recommendations Based on Academic Performance and Aspirations. *Frontiers in Health Informatics*, *13*(3), 8216–8230.

Batmaz, Z., Yurekli, A., Bilge, A., & Kaleli, C. (2019). A review on deep learning for recommender systems: challenges and remedies. *Artificial Intelligence Review*, *52*(1), 1–37. https://doi.org/10.1007/s10462-018-9654-y

Chai, K. E. K., & Gibson, D. (2015). Predicting the risk of attrition for undergraduate students with time based modelling. *Proceedings of the 12th Int. Conference on Cognition and Exploratory Learning in the Digital Age, CELDA 2015*, *Celda*, 109–116.

Cheng, P. (2024). Personalized Recommendation of College Students' Innovative Entrepreneurship Education Course Content Based on Bayesian Classifier. *Journal of Electrical Systems*, *20*(6s), 1493–1504. https://doi.org/10.52783/jes.3068

Dascalu, M. I., Bodea, C. N., Mihailescu, M. N., Tanase, E. A., & Ordoñez De Pablos, P. (2016). Educational recommender systems and their application in lifelong learning. *Behaviour and Information Technology*, *35*(4), 290–297. https://doi.org/10.1080/0144929X.2015.1128977

Demong, N. A. R., Shahrom, M., Rahim, R. A., Omar, E. N., & Yahya, M. (2023). Personalized Recommendation Classification Model of Students' Social Well-being Based on Personality Trait Determinants Using Machine Learning Algorithms. *Journal of Information and Communication Technology*, *22*(4), 545–585. https://doi.org/10.32890/jict2023.22.4.2

DS, G., & K, K. (2015). Developing an Intelligent Recommendation System for Course Selection by Students for Graduate Courses. *Business and Economics Journal*, *7*(2), 109–125. https://doi.org/10.4172/2151-6219.1000209

Elewa, Y. A., Munassar, N. M., & Abdullah, M. F. (2025). *Optimizing Intelligent Marketing with AI-Based DSS (Case Study : Sprout Social at University of Science and Technology , Aden , Yemen)*. *1*(1), 1–10.

Fadhl, M., & Elewah, Y. (2024). AI in Education: Examining ChatGPT Awareness and Usage among University of Science & Technology Students, Aden. *International Research Journal of Innovations in Engineering and Technology (IRJIET)*, *8*(8), 26–36. https://doi.org/10.47001/IRJIET/2024.808004

Falát, L., & Piscová, T. (2022). Predicting GPA of University Students with Supervised Regression Machine Learning Models. *Applied Sciences (Switzerland)*, *12*(17). https://doi.org/10.3390/app12178403

Hammoudi Halat, D., Abdel-Salam, A. S. G., Bensaid, A., Soltani, A., Alsarraj, L., Dalli, R., & Malki, A. (2023). Use of machine learning to assess factors affecting progression, retention, and graduation in first-year health professions students in Qatar: a longitudinal study. *BMC Medical Education*, *23*(1), 1–11. https://doi.org/10.1186/s12909-023-04887-w

Hung, J. L., Shelton, B. E., Yang, J., & Du, X. (2019). Improving Predictive Modeling for At-Risk Student Identification: A Multistage Approach. *IEEE Transactions on Learning Technologies*, *12*(2), 148–157. https://doi.org/10.1109/TLT.2019.2911072

Iqbal, N., Ahmed, H., & Azhar, K. A. (2022). EXPLORING TEACHERS' ATTITUDES TOWARDS USING CHATGPT. *Global Journal for Management and Administrative Sciences*, *3*(4), 97–111. https://doi.org/10.46568/gjmas.v3i4.163

Kamal, N., Sarker, F., Rahman, A., Hossain, S., & Mamun, K. A. (2024). Recommender System in Academic Choices of Higher Education: A Systematic Review. *IEEE Access*, *12*(January), 35475–35501. https://doi.org/10.1109/ACCESS.2024.3368058

Kim, T. Y., & Lim, J. B. (2023). Developing an Intelligent Recommendation System for Non-Information and Communications Technology Major University Students. *Applied Sciences (Switzerland)*, *13*(23). https://doi.org/10.3390/app132312774

Lottering, R., Hans, R., & Lall, M. (2020). A machine learning approach to identifying students at risk of dropout: a case study. *International Journal of Advanced Computer Science and Applications*, *11*(10), 417–422. https://doi.org/10.14569/IJACSA.2020.0111052

Mason, C., Twomey, J., Wright, D., & Whitman, L. (2018). Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression. *Research in Higher Education*, *59*(3), 382–400. https://doi.org/10.1007/s11162-017-9473-z

Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access*, *8*, 55462–55470. https://doi.org/10.1109/ACCESS.2020.2981905

Naher, J., Haque Maruf, M., Tamzid Bakht, S., & Samin Sadaf, S. (2023). *A Survey to Understand the experience of ChatGPT Usage among Engineering University Students in Bangladesh*. www.techniumscience.com

Ojajuni, O., Ayeni, F., Akodu, O., Ekanoye, F., Adewole, S., Ayo, T., Misra, S., & Mbarika, V. (2021). Predicting Student Academic Performance Using Machine Learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 12957 LNCS* (Issue September). Springer International Publishing. https://doi.org/10.1007/978-3-030-87013-3_36

Siswipraptini, P. C., Warnars, H. L. H. S., Ramadhan, A., & Budiharto, W. (2024). Personalized Career-Path Recommendation Model for Information Technology Students in Indonesia. *IEEE Access*, *12*, 49092–49105. https://doi.org/10.1109/ACCESS.2024.3381032

Tang, Z., Jain, A., & Colina, F. E. (2024). A Comparative Study of Machine Learning Techniques for College Student Success Prediction. *Journal of Higher Education Theory and Practice*, *24*(1), 101–116. https://doi.org/10.33423/jhetp.v24i1.6764

Tarus, J. K., Niu, Z., & Mustafa, G. (2018). Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, *50*(1), 21–48. https://doi.org/10.1007/s10462- 017-9539-5

Tripathi, G., Reddy, K. N., Sahithi, K., Maddineni, A., & Kumar, #. (2024). Issue 3 Dr. Geeta Tripathi et al. *International Journal of Scientific Research in Science and Technology (Www.Ijsrst.Com)*, *11*(3), 1–10. https://doi.org/10.32628/IJSRST