International Journal of Science, Technology and Society
2025, Vol. 13, No. 2, pp. 35-42
https://doi.org/10.11648/j.ijsts.20251302.11

Research Article

Career Guidance System Using Decision Tree, Random Forest, and Naïve Bayes Algorithm

Chidi Ukamaka Betrand[1, *], Obinna Banner Aliche[2], Chinwe Gilean Onukwugha[1], Christopher Ifeanyi Ofoegbu[1], Douglas Allswell Kelechi[1], Ihechiluru Chinwe Ugbor[1], Nneka Martina Oragba[1]

[1]Department of Computer Science, School of Information and Communication Technology, Federal University of Technology, Owerri, Nigeria

[2]Department of Air Traffic Control Services, Directorate of Air Traffic Services, Nigerian Airspace Management Agency, Federal Ministry of Aviation and Aerospace Development, Port Harcourt, Nigeria

Abstract

Students often struggle with identifying the right options that align with their interests, abilities, and aspirations. Most students lack the required knowledge to make the right decisions. After receiving a degree, the path to career specialization always seems unclear for most students. But, if a student can manage to get it right by choosing the right path for their career, they will experience significant economic and psychological benefits. Choosing the right career path is a critical decision that can significantly impact an individual's future. Providing effective career guidance is therefore essential, especially for students who often face challenges in aligning their interests, skills, and aspirations with suitable career options. This study addresses this need by developing and evaluating a comprehensive Career Guidance System utilizing three machine learning algorithms: Decision Tree, Random Forest, and Naive Bayes. The system was built using an iterative approach, incorporating a user-friendly web page and an interactive chatbot to enhance the career guidance experience. Developed and deployed using Python and the powerful Django framework, the system leverages cutting-edge technologies to deliver personalized recommendations tailored to each student's unique profile. To evaluate the system's performance, key metrics such as accuracy, precision, recall, and F1 score were employed. Notably, the Random Forest classifier outperformed the other algorithms, achieving the highest accuracy. This superior performance highlights the algorithm's ability to capture complex relationships between student interests, passions, and career choices, making it an ideal choice for career guidance applications. The Career Guidance System developed in this study holds significant potential for revolutionizing the career counseling process. The choice of algorithms used in this study was chosen given the specific needs of the project, especially considering specific concerns of scalability and accuracy. in the advancement of computer science and its applications in career counseling. The findings demonstrate the system's overall efficiency and effectiveness, paving the way for its wider adoption and further refinement to support students in making informed and fulfilling career choices.

Keywords

Recommder System, Real Time, Career, Guidance, Na ïve Bayes, Optimization, Decision Tree

[*]Corresponding author: chidi.betrand@futo.edu.ng (Chidi Ukamaka Betrand)

Received: 8 October 2024; **Accepted:** 30 October 2024; **Published:** 18 March 2025

1. Introduction

Career development and its pursuit is an important aspect of a student's life. This is because it shapes their future and influences their personal and professional growth. Selecting an appropriate career path is one of the most significant decisions an individual can make in life. The chosen career has major implications for a person's future employment prospects, earnings potential, job satisfaction, and more. However, picking a career is complex and students often struggle with identifying the right options that align with their interests, abilities, and aspirations. Most students lack the required knowledge to make the right decisions [1]. After receiving a degree, the path to career specialization always seems unclear for most students. But, if a student can manage to get it right by choosing the right path for their career, they will experience significant economic and psychological benefits. Recent advances in technology have led to an exponential growth in computing-related career opportunities. This has led to a high demand for computing professionals. This demand has boosted the number of computing job openings [2]. There is now high market demand for qualified computing professionals to fill critical roles in fields like software engineering, computer programming, data science, and information technology. One of the most important decisions a student or graduate must make is to find their first suitable job in any field related to computing. This sounds profitable but the ambiguity that lies in this career decision stage leads some to prolong graduation or switch majors after making unviable choices [3].

This system will employ three Machine Learning algorithms: Random Forest, Decision Tree, and Naïve Bayes. Advanced ML algorithms like random forests or decision trees can detect key patterns accounting for multiple strengths, weaknesses, interests, and external factors human advisors cannot match. The algorithms can ingest diverse student data points to determine optimal career and development suggestions tailored to the individual.

2. Review of Related Works

As the job market grows increasingly complex and competitive, effective career guidance has become crucial for young people and career-changers to identify well-matched vocations aligned with their talents and interests [4]. However, traditional methods of career counseling conducted by human experts struggle to meet this growing need, often providing generalized advice that fails to take into account individuals' unique attributes. An automated system using ensemble learning for the prediction of a suitable job career in the software industry for students seeking opportunities took into consideration the personality traits and interests of each student in relation to what they can offer [5]. A Smart Career Guidance System that help students who are about to begin their higher education analyze their skills, abilities, and interests and recommends five fields that are most suitable for them was presented. The various machine Learning techniques applied mimiked a one-on-one meeting with an experienced career counselor. A total of 392 graduates completed this online survey as SMOTE oversampling was used to evaluate the machine learning classifiesrs. The XGBoost and Random Forest classifiers recommended the best-suited career options [6]. Repaso & Caparino (2020) in their study explored classification techniques to predict graduates' career specialization in Information Technology. The data were analyzed using Machine learning algorithms, with 18 attributes. Naïve Bayes and Random Forest models demonstrated higher accuracy, with satisfactory ROC and RMSE values. However, the study's findings may not generalize beyond Bulacan State University Sarmiento Campus' Information Technology program where the datasets for the development of the model were gotten. Further research is needed to explore interventions for improving student outcomes in some other courses [7].

Various academic and personal factors influencing career selection were examined as data mining techniques were applied to analyze factors and predict career paths, employing multiple performance evaluation metrics. The study identified the Part classifier as the most effective technique for career prediction, surpassing other methods. The research sheds light on the complex dynamics influencing career decisions in Bangladesh's evolving job market as a casestudy. By leveraging data mining techniques, insights into academic and personal factors affecting career choices are gained, with the Part classifier emerging as a superior predictive tool. However, the study's scope did not fully capture all factors influencing career decisions, and its findings specific to the context of Bangladesh. Further research is necessary to explore additional variables and validate predictive models in diverse settings [8].

A Career Guidance System Using Machine Learning was proposed by Panthee *et al.*, (2023), [9] focused on providing career guidance to students by predicting their personality traits and learning styles to assist in job role selection. The study utilizes web-based questionnaires to collect data on users' Big Five personality traits and VAK learning types. The system then recommends career choices tailored to individual profiles. Machine learning algorithms are used to evaluate the performance of the system in predicting personality traits and learning styles, validating the accuracy of the results. The research underscores the importance of career guidance systems in helping students leverage their capabilities and interests to make informed career decisions. Further research is needed to assess the system's effectiveness across diverse user populations and validate its recommendations in real-world career settings. Addressing the common dilemma faced by

students worldwide upon completing higher secondary education and focusing on guiding them towards suitable career paths, the computerized career counseling system offers a systematic approach to help students make informed decisions about their future career paths [10]. Despite its potential benefits, the system's reliance on objective assessments may not fully capture the nuanced preferences and aptitudes of individual students. Further research is needed to assess the system's effectiveness in diverse cultural and educational contexts and to explore additional factors influencing career decision-making processes.

On the challenges of providing career guidance in the era of life-long learning, particularly in bridging education and employment services, exploring the potential of artificial intelligence (AI) to enhance career guidance across higher education and working life [11]. The research focuses on the development of AI-driven systems to support career guidance in higher education institutions. Results are derived from focus groups, scenario work, and practical trials, mapping the requirements and possibilities of AI in career guidance from the perspectives of students, guidance staff, and institutions. The findings highlight the potential value, functions, drivers, and barriers of adopting AI in career guidance to support higher education and life-long learning. A paper that addresses the importance of question classification in automatic question answering systems, particularly focusing on questions related to higher studies and career choices. The aim is to classify questions based on students' abilities and skills using natural language processing and machine learning techniques. The research involves collecting and labeling a dataset of student questions, preprocessing textual content, and extracting features using the TF-IDF statistical representation. Four supervised machine learning algorithms—logistic regression (LR), decision tree (DT), K-nearest neighbors (KNN), and support vector machine (SVM)—are implemented for classification. Additionally, a feature selection technique is applied to enhance system performance [12].

The critical need for effective career selection in today's world, emphasizing the complexity and challenges students face in making informed decisions. Aiming to provide a solution through an application that offers career guidance based on students' skills, interests, abilities, and capabilities, particularly after completing 10th and 12th grades. Lohar et al., 2023) evaluates students through aptitude tests and provides tailored career guidance to help them achieve success [13]. Additionally, a Recommendation Engine, constructed using Machine Learning techniques implemented in Python, recommends career options to students. Three Machine Learning Algorithms—Naive Bayes, K-Nearest Neighbor, and Random Forest Classifier—are utilized for recommendation. Upon accessing the portal, students undergo a series of tests to assess their IQ, EQ, and Personality traits [14]. Another project that utilizes machine learning algorithms to analyze personal and academic information provided by students, ultimately recommending the best career options and domains based on their skills and attributes [15]. An AI-based career counseling framework that integrates machine learning techniques to match students with suitable courses and careers based on their interests and aptitudes. The system architecture follows a multi-agent expert system approach with separate components for a knowledge base, inference engine, explaining predictions, etc. The knowledge base stores information collected from students and parents through interviews/surveys about factors like subjects of interest, skills, and career aspirations. The inference engine applies rules and logic to this knowledge to deduce recommendations. Explainable AI is incorporated to provide reasoning behind suggested courses of action. The framework trains machine learning models like decision trees, random forests, Naive Bayes on historical educational datasets. The algorithms analyze attributes like grades, test scores, extracurricular activities to identify patterns linking student profiles with optimal career trajectories. Comparative evaluation of the ML models on classification metrics like accuracy, precision reveals top performers [16]. Addressing the challenge of higher education's lack of knowledge regarding student talents, which often results in decreased chances of success and potential job difficulties due to choosing the wrong course [17]. This study seeks to predict student skills early to enable mentors to provide timely advice and improve student success rates. The study employs a model created using Naïve Bayes, J48, Random Forest, and Support Vector Machine (SVM) classification algorithms, utilizing 100 attributes. The research explores a dynamic dataset by applying data mining methods to analyze student insights based on characteristics related to academic, technical, environmental, and interpersonal factors. Among the models built, the Naïve Bayes and Random Forest algorithms demonstrated superior accuracy ratings. The model's performance was tested in a constraint-based learning environment and found to be effective in predicting student skills and offering insights into their academic and personal attributes. The study highlights the importance of early prediction of student skills to enhance academic advising and improve student success rates.

On enhancing the efficacy of a model for student career recommendations by employing feature selection, data sampling, and machine learning techniques. In their study. Nuanmeesri & Poomhiran, 2023) aims to investigate the impact of various methods on model performance and identify the most effective approach for predicting student careers [18]. The research employs a wrapper approach for feature selection and evaluates its effectiveness in combination with oversampling techniques. It explores the performance of a model combining k-mean, Synthetic Minority Over-sampling Technique (SMOTE), and Edited Nearest Neighbor (ENN) methods. The study develops models using machine learning techniques, with a focus on the Multi-Layer Perceptron Neural Network (MLPNN) classifier. The performance of various classifiers, including Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Random Forest, Decision Tree,

Na ïve Bayes, and Logistic Regression, is assessed. The results indicate that the wrapper approach improves the model's performance, particularly when combined with oversampling techniques such as SMOTE. The model incorporating k-mean, SMOTE, and ENN methods demonstrates significant improvements in performance. Furthermore, the model utilizing MLPNN outperforms other classifiers in terms of efficiency. The developed model achieves a prediction accuracy of 95.88% and a root mean squared error of 2.0968. The study concludes that the combination of feature selection, data sampling, and machine learning techniques enhances the efficacy of the model for student career recommendations.

3. Materials and Methods

Data Collection: The career guidance system dataset from mendeley was used to train and evaluate these models. This dataset can be accessed via https://data.mendeley.com/datasets/kzt6h7pz97/1 With the dataset, the career recommendation system was developed. This training was done in Python on a Jupyter Notebook. The model that has the highest precision, recall rate, F1 score, and accuracy will be deployed and used for real-time prediction. The preferred model will be used to power a web application built with the Django Python framework.

3.1. Feature Extraction

Feature selection is a crucial step before training a prediction model. High dimensionality in a dataset can cause a reduction in performance. High complexity takes a long time to train the model. This necessitates removing any unnecessary features from the list of variables. Choosing the right algorithm to perform feature selection is necessary. The algorithm that needs to be used is dependent on the dataset's independent variables and the target variable. For the career prediction system dataset, the SelectKBest extraction method was used. This feature extraction approach ensures that relevant features are retained and the unwanted features in the given data are dropped. Relevant features influencing career choices were selected from the dataset. Feature engineering techniques were applied using Jupyter Notebook to derive new features or enhance existing ones.

Machine learning algorithms including Random Forest, Decision Tree, and Naive Bayes were chosen for mapping student attributes to career trajectories. Jupyter Notebook was used for model training and evaluation. The dataset was split into training and testing sets, and cross-validation techniques were applied to assess model performance.

3.2. Model Evaluation

Three algorithms were for the prediction process –Decision Tree, Random Forest and Na ïve Bayes. A performance comparison of the prediction models is presented. The aim is to find the most suitable model to be used for this research problem. The models are: Decision Tree, Na ïve Bayes and Random Forest Algorithm.

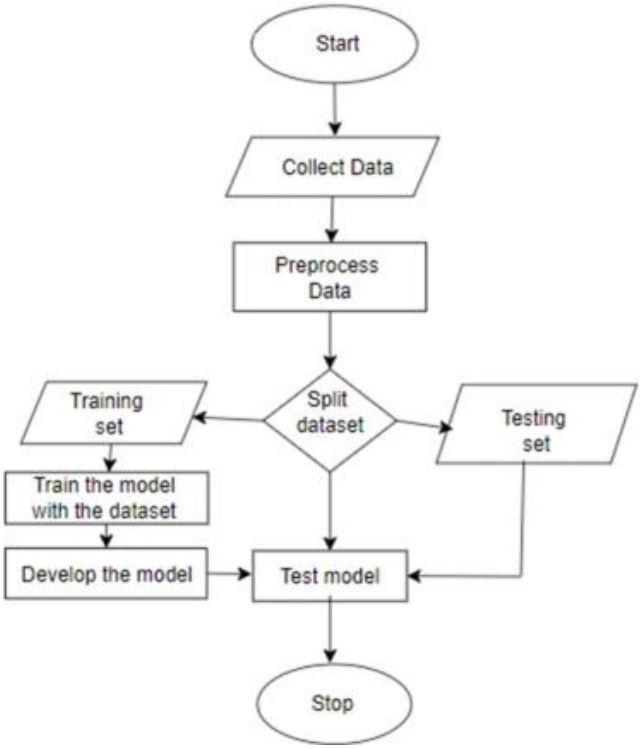

Figure 1. *Flowchart of how the model is trained.*

Metrics such as accuracy, precision, recall, and F1-score were used to evaluate the performance of the trained models. Comparative analysis of different algorithms was conducted using Jupyter Notebook to assess their strengths and limitations. The classification report is shown for each of the model is shown below.

To assess the accuracy of the machine learning models, the following standard performance metrics were used: Accuracy, Precision and Recall, F1 Score. The chatbot was developed using pure python code with a simple logic that responds to users query and give a suggestion on their interested field and the possible prospects of that field.

4. Results and Discussion

Employing three Machine Learning algorithms—Random Forest, Decision Tree, and Na ïve Bayes—the system went through training and evaluation using the career guidance system dataset. The training process done using Python, facilitated by a Jupyter notebook. The model exhibiting superior precision, recall rate, F1 score, and accuracy was selected for real-time predictions. A chatbot w integrated into the system to enable the student interact with the system. This chatbot developed with ChatterBot. It is a Python library for developing chatbots using Natural Language processing techniques.

The Natural Language Processing technique to used in this study is known as TF-IDF (Term Frequency-Inverse Document Frequency) method for similarity calculation. These chatbots can engage in conversations with students, respond to queries, and simulate human-like interactions.

Deployment is slated for a Django-built web application, fostering user-friendly access to personalized career guidance and contributing to a more enlightened and content future workforce. To develop this system the following hardware and software will be used: RAM (8.00GB), Intel (R) Core (TM) i5-4200M CPU @ 2.50GHz. The software requirements are Python, machine learning algorithms (Random Forest, Decision Tree, Support Vector Machine, and Naïve Bayes), Django, Web browser, and Front-end Technologies (HTML, CSS, JavaScript, for building the user interface of the web application). The new system will be implemented in real time which can be used for predicting suggested careers.

Visual Studio Code was used as the IDE for developing the interactive web application interface using Django. Design and development of the web application interface were carried out, incorporating user experience considerations and accessibility features. The figure below shows the setup for developing the web app.

The trained machine learning models were integrated with the web application interface developed in Django. Deployment of the platform was achieved by hosting the web application on a server, allowing users to access personalized career recommendations instantly through their browsers.

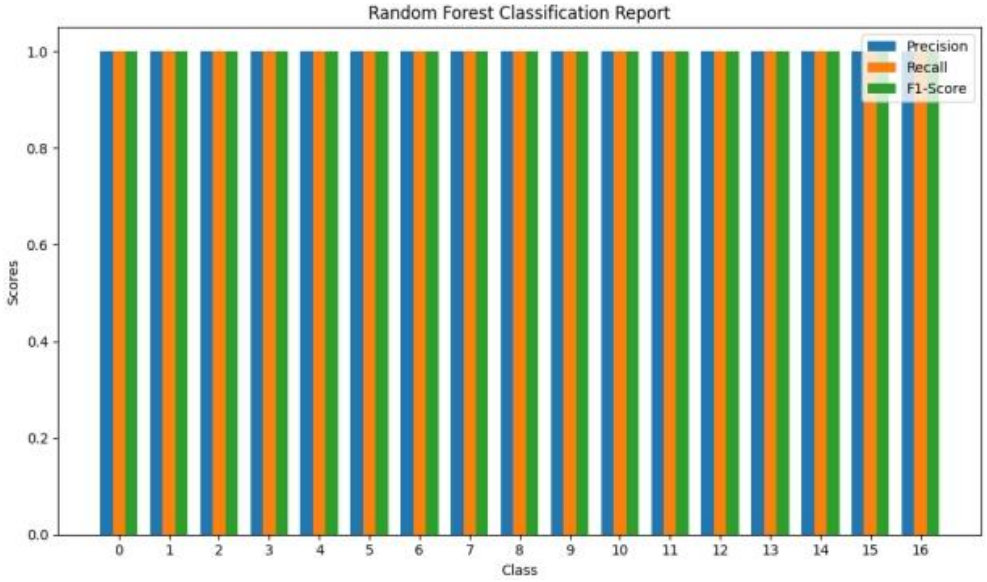

Figure 2. *Barchart of Random Forest classification report.*

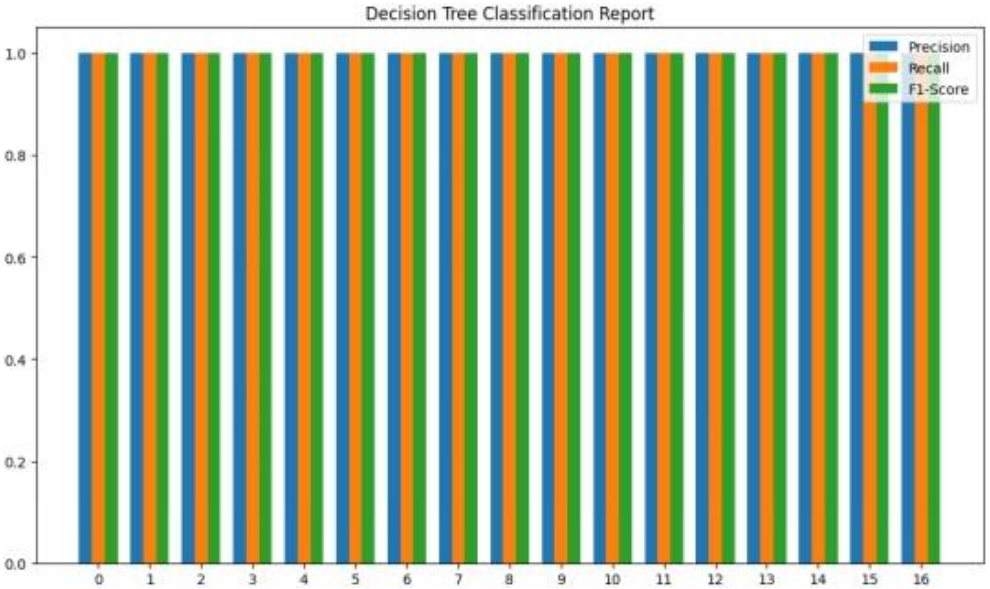

Figure 3. *Barchart of Decision Tree classification report.*

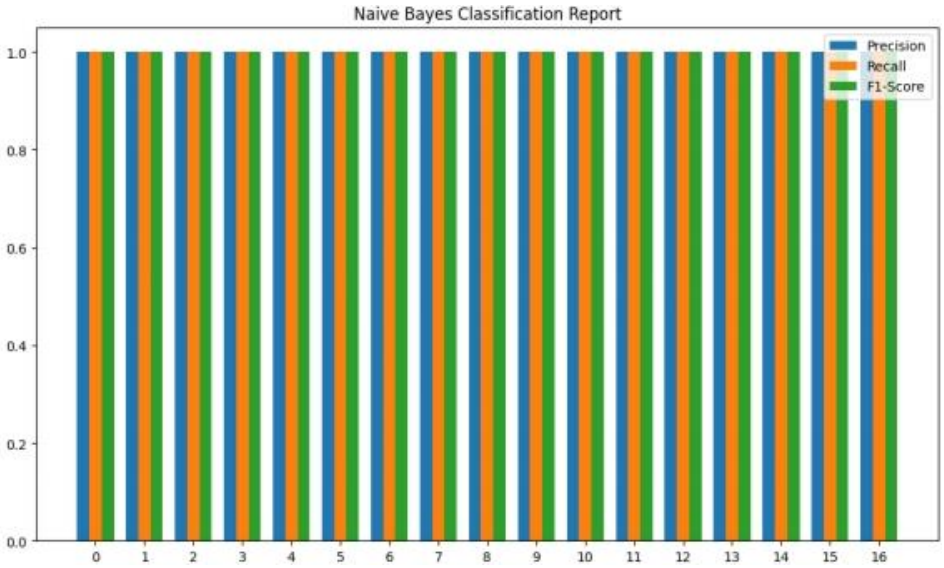

Figure 4. *Barchart of Naive Bayes classification report.*

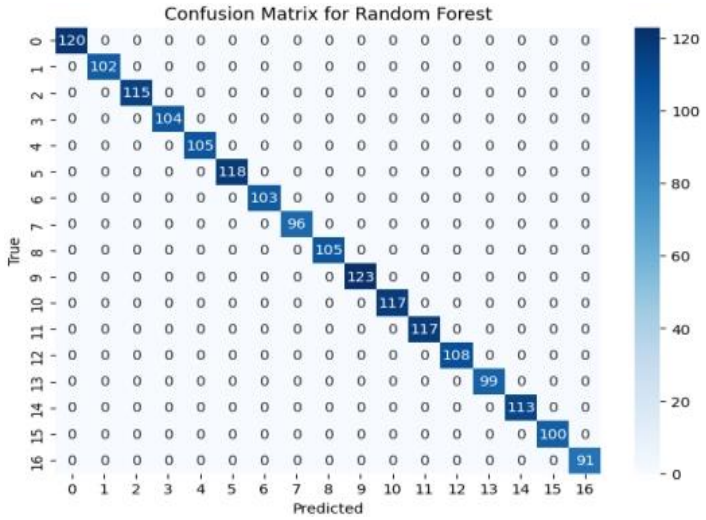

Figure 5. *Confusion matrix for Random Forest classifier.*

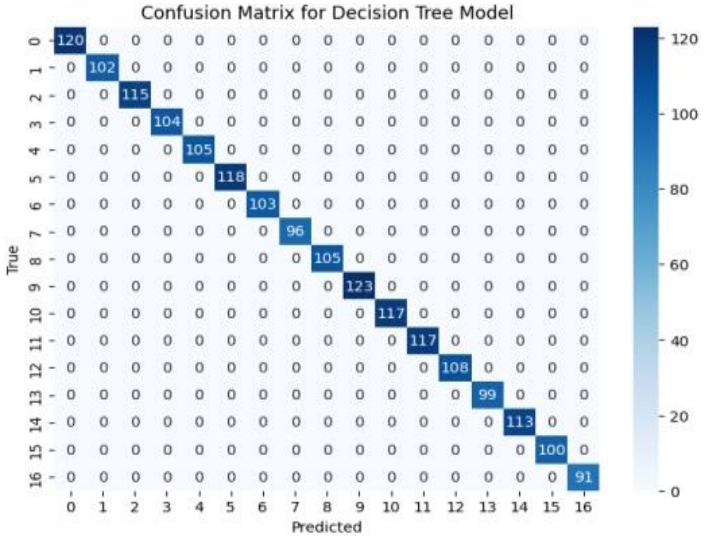

Figure 6. *Confusion matrix for Decision Tree.*

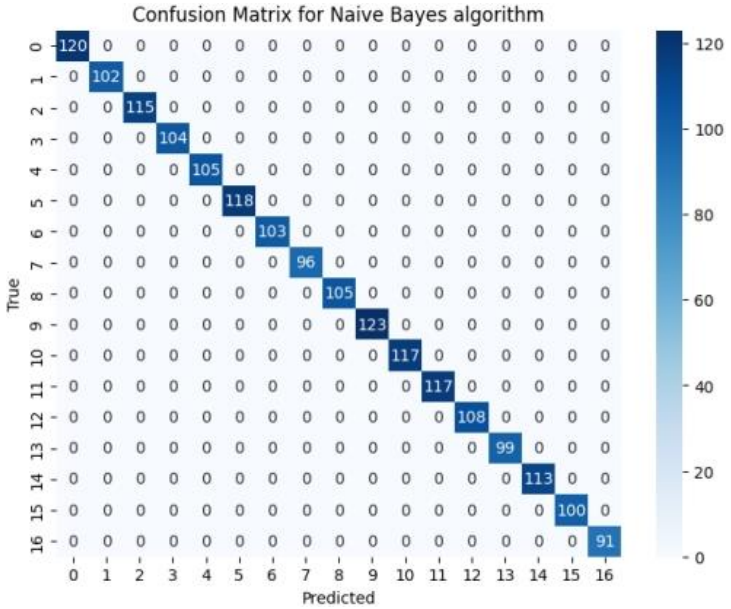

Figure 7. *Confusion matrix for Naive Bayes.*

The classification reports offer valuable insights into the performance of the Random Forest, Decision Tree, and Naive Bayes classifiers, respectively. Notably, the Random Forest classifier achieved a remarkable accuracy of 100%, indicating its robustness in accurately predicting career paths based on student attributes. This high accuracy underscores the effectiveness of the Random Forest model in capturing complex relationships between student interests, passion, and career choices.

The bar charts presented in Figures 2, 3, and 4 elucidate the precision, recall, and F1-score for each class in the classification reports. These visualizations provide a comprehensive overview of the models' performance across different classes, highlighting areas of strength and potential improvement. Despite achieving perfect accuracy, it's essential to analyze the precision, recall, and F1-score for individual classes to identify any specific areas where the model may excel or struggle.

Figures 5, 6, and 7 showcase the confusion matrices for the Random Forest, Decision Tree, and Naive Bayes classifiers, respectively. These matrices offer insights into the models' abilities to correctly classify instances and identify any patterns of misclassification. By examining the confusion matrices, researchers can gain a deeper understanding of the models' performance and identify areas for refinement or further investigation.

The trained machine learning models was integrated with the interactive web application interface developed using Django in Visual Studio Code enabling seamless user interaction and real-time access to personalized career guidance.

The intuitive design and user-friendly interface of the web app facilitate easy navigation and enable students to receive tailored career recommendations and insights effortlessly. The chatbot engages users in interactive dialogue, guiding them through the career exploration process and providing personalized recommendations based on their interests and preferences. Overall, the presented figures collectively demonstrate the successful development and integration of a comprehensive career guidance system, leveraging machine learning algorithms and interactive technology to empower students in making informed career decisions.

5. Conclusion

In conclusion, the development and deployment of the comprehensive career guidance system mark a significant advancement in the field of educational technology. By harnessing the power of machine learning and interactive interfaces, the system empowers students to make informed career decisions aligned with their individual strengths and aspirations. The high accuracy achieved by the machine learning models, coupled with the user-friendly design of the web application and chatbot interface, underscores the effectiveness and potential impact of the system in guiding students towards fulfilling career paths.

Abbreviations

TF-IDF	Term Frequency-Inverse Document Frequency
LR	Logistic Regression
DT	Decision Tree
SVM	Support Vector Machine
KNN	K-Nearest Neighbour
SMOTE	Synthetic Minority Over-Sampling Technique
MLPNN	Multi-Layer Perceptron Neural Network

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Kabari, L. G., & Agaba, F. (2019). An Intelligent Career Advisor Expert System. International Journal of Advanced Research and Publications, 3(4), 91-94.

[2] Macharia, M. S. (2019). Dr. Reengineering mass career acquisition through technical vocational education training counseling in Kenya. International Journal of Research in Business and Social Science, 212-218.

[3] Subahi, A. F. (2018). Data Collection for Career Path Prediction Based on Analysing Body of Knowledge of Computer Science Degrees. Journal Software, 533-546.

[4] Idakwo, J., Babatunde, A. J., & Kolajo, T. (2022). Development of a Hybrid Student's Career Path Recommender System Using Machine Learning Techniques. FUW Trends in Science & Technology Journal, 7(3), 236-242.

[5] Reddy, K. Sai Vivek, Reddy, M. Akhila, Kaur, Veerpal, & Kaur, Gagandeep. (2022). Career Guidance System Using Ensemble Learning. Proceedings of the Advancement in Electronics & Communication Engineering 2022. http://dx.doi.org/10.2139/ssrn.4157249

[6] Kamal, A., Naushad, B., Rafiq, H., & Tahzeeb, S. (2021). Smart Career Guidance System. 2021 4th International Conference on Computing & Information Sciences (ICCIS), Karachi, Pakistan, 1-7. https://doi.org/10.1109/ICCIS54243.2021.9676408

[7] Repaso, J. A. A., & Caparino, E. T. (2020). Analyzing and Predicting Career Specialization using Classification Techniques. International Journal of Advanced Trends in Computer Science and Engineering, 9(13), 342-348.

[8] Biswas, A. A., Majumder, A., Jueal Mia, M., & Basri, R., Sabab Zulfiker, M. (2021). Career Prediction with Analysis of Influential Factors Using Data Mining in the Context of Bangladesh. In: Kaiser, M. S., Bandyopadhyay, A., Mahmud, M., Ray, K. (eds) Proceedings of International Conference on Trends in Computational and Cognitive Engineering. Advances in Intelligent Systems and Computing, 1309. Springer, Singapore. https://doi.org/10.1007/978-981-33-4673-4_35

[9] Panthee, S., Rajkarnikar, S., & Begum, R. (2023). Career Guidance System Using Machine Learning. Journal of Advanced College of Engineering and Management, 8(2), 113–119. https://doi.org/10.3126/jacem.v8i2.55947

[10] Vignesh, S. P., Shivani Priyanka, C., Shree Manju, H., & Mythili, K. B. (2021). An Intelligent Career Guidance System using Machine Learning. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 1, 987-990.

[11] Westman, S., Kauttonen, J., Klemetti, A., Korhonen, N., Manninen, M., Mononen, A., Niittymaki, & Paananem, H. (2021). Artificial Intelligence for Career Guidance – Current Requirements and Prospects for the Future. IAFOR Journal of Education: Technology in Education, 9(4), 43-62.

[12] Ouassil, M. A., Errami, M., Rachidi, R., Hamida, S., Cherradi, B., & Raihani, A. (2023). A Classification System of Career Guidance Questions Based on Natural Language Processing and Supervised Machine Learning Techniques. 2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Mohammedia, Morocco, pp. 01-08. https://doi.org/10.1109/IRASET57153.2023.10152874

[13] Lohar, N., Sanadhya, Y., Paliwal, R., Singh, M., & Sachdev, P. (2023). Web-Based Career Guidance System. International Advanced Research Journal in Science, Engineering and Technology, 10(2), 122-125.

[14] Jash, H. P., et al. (2021). Career Guidance System Using Machine Learning. International Journal of Creative Research Thoughts (IJCRT), 9(5), 57-60.

[15] Daharwal, A., Gore, S., Bhagwat, A., Dethe, S., & Chavan, S. (2020). Career Guidance System using Machine Learning For Engineering Students (CS/IT). International Research Journal of Engineering and Technology, 7(6), 3417-3420.

[16] Guleria, P., & Sood, M. (2022). Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling. Education and Information Technologies, 28(1), 1081–1116. https://doi.org/10.1007/s10639-022-11221-2

[17] Thirumalaiselvi, R., & Narayanan, P. (2021). An Effective Model for Choosing Career in Schools Using Data Mining Techniques. I-manager's Journal on Software Engineering, 15(3), 21.

[18] Nuanmeesri, S., & Poomhiran, L. (2023). Improving efficient student Career Recommendations Using Machine Learning. Home, 58(2).