МРНТИ 28.23.25

10.51889/2959-5894.2025.90.2.016

B.A. Berlikozha [1], A.G. Serek [2*], B.Ye. Amirgaliyev [3], Z.D. Baishemirov [4], A.M. Zhamanov [5]

[1] SDU University, Kaskelen, Kazakhstan
[2] Kazakh-British Technical University, Almaty, Kazakhstan
[3] Astana IT University, Astana, Kazakhstan
[4] Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[5] North American University, USA, Houston
[*]*e-mail:* a.serek@kbtu.kz

IDENTIFICATION OF KEY FEATURES FOR CAREER PREDICTION THROUGH RECURSIVE FEATURE ELIMINATION

Abstract

In the contemporary job market, students face significant challenges in selecting career paths that align with their skills and aspirations. The interplay of factors affecting career decisions, including academic performance, personality traits, and certifications, complicates the guidance process for educators and career counselors. This research examines the necessity for data-driven insights in career prediction by applying machine learning methods to analyze a dataset of student profiles. The primary objective is to identify the key features that have a significant impact on predicting students' career trajectories. Classifiers, including Random Forest, Logistic Regression, and Support Vector Machine, were employed, revealing that the Random Forest classifier attained an accuracy of 75%, a precision of 70%, a recall of 75%, and an F1 score of 72%. The Logistic Regression model exhibited an accuracy of 70%, whereas the SVM classifier attained an accuracy of 72%. Recursive Feature Elimination for feature selection revealed that specific certifications have a significant predictive value for career success. The findings underscore the importance of leveraging machine learning in educational settings to enhance personalized career guidance, thereby enabling students to make informed decisions about their futures.

Keywords: feature elimination, career prediction, matching career, machine learning, artificial intelligence.

Б.А. Берликожа [1], А.Г. Серек [2], Б.Е. Амиргалиев [3], Ж.Д. Байшемиров [4], А.М. Жаманов [5]
[1] SDU University, г.Каскелен, Казахстан
[2] Казахстанско-Британский технический университет, г.Алматы, Казахстан
[3] Astana IT University, г.Астана, Казахстан
[4] Казахский национальный педагогический университет имени Абая, г.Алматы, Казахстан
[5] North American University, Хьюстон, США

ОПРЕДЕЛЕНИЕ КЛЮЧЕВЫХ ПРИЗНАКОВ ДЛЯ ПРОГНОЗИРОВАНИЯ КАРЬЕРЫ ПУТЕМ РЕКУРСИВНОГО ИСКЛЮЧЕНИЯ ПРИЗНАКОВ

Аннотация

На современном рынке труда студенты сталкиваются со значительными трудностями при выборе карьерных путей, соответствующих их способностям и амбициям. Взаимодействие факторов, влияющих на решения о карьере, включая успеваемость, личностные качества и сертификацию, усложняет процесс руководства для преподавателей и консультантов по карьере. В этом исследовании изучается необходимость в данных, основанных на данных, в прогнозировании карьеры с помощью применения методов машинного обучения для анализа набора данных профилей студентов. Основная цель — определить ключевые характеристики, которые существенно влияют на прогнозирование карьерных траекторий студентов. Были использованы классификаторы, включая случайный лес, логистическую регрессию и машину опорных векторов, что показало, что классификатор случайного леса достиг точности 75%, точности 70%, полноты 75% и оценки F1 72%. Модель логистической регрессии продемонстрировала точность 70%, тогда как классификатор SVM достиг точности 72%. Рекурсивное исключение признаков для выбора признаков показало, что определенные сертификации служат значимыми предикторами успеха в карьере. Результаты исследования подчеркивают важность использования машинного обучения в образовательных учреждениях для улучшения персонализированной профессиональной ориентации, тем самым помогая учащимся принимать обоснованные решения относительно своего будущего.

Ключевые слова: исключение признаков, прогнозирование карьеры, подбор карьеры, машинное обучение, искусственный интеллект.

Б.А. Берликожа [1], А.Г. Серек [2], Б.Е. Амиргалиев [3], Ж.Д. Байшемиров [4], А.М. Жаманов [5]
[1] SDU University, Қаскелең қ., Қазақстан
[2] Қазақ-Британ техникалық университеті, Алматы қ., Қазақстан
[3] Astana IT University, Астана қ., Қазақстан
[4] Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы қ., Қазақстан
[5] North American University, Хьюстон, АҚШ

РЕКУРСИВТІ ФУНКЦИЯЛАРДЫ ЖОЮ АРҚЫЛЫ МАНСАПТЫ БОЛЖАУ ҮШІН НЕГІЗГІ ЕРЕКШЕЛІКТЕРДІ АНЫҚТАУ

Аңдатпа

Қазіргі еңбек нарығында студенттер өздерінің қабілеттері мен амбицияларына сәйкес келетін мансап жолдарын таңдауда айтарлықтай қиындықтарға тап болады. Мансап шешімдеріне әсер ететін факторлардың өзара әрекеттесуі, оның ішінде академиялық үлгерім, тұлғалық қасиеттер және сертификаттар педагогтар мен мансап бойынша кеңесшілерге басшылық беру процесін қиындатады. Бұл зерттеу студент профильдерінің деректер жинағын талдау үшін машиналық оқыту әдістерін қолдану арқылы мансапты болжауда деректерге негізделген түсініктердің қажеттілігін зерттейді. Негізгі мақсат – студенттердің мансаптық траекториясын болжауға елеулі әсер ететін негізгі белгілерді анықтау. Кездейсоқ орман, логистикалық регрессия және қолдау векторлық машинасын қамтитын жіктеуіштер қолданылды, бұл кездейсоқ орман классификаторы 75% дәлдікке, 70% дәлдікке, 75% еске түсіруге және 72% F1 ұпайына қол жеткізгенін көрсетті. Логистикалық регрессия үлгісі 70% дәлдік көрсетті, ал SVM классификаторы 72% дәлдікке қол жеткізді. Функцияларды таңдауға арналған рекурсивті мүмкіндікті жою белгілі бір сертификаттар мансаптық табыстың маңызды болжаушылары ретінде қызмет ететінін көрсетті. Нәтижелер жекелендірілген мансаптық бағдарды жақсарту үшін оқу орындарында машиналық оқытуды пайдаланудың маңыздылығын көрсетеді, осылайша студенттерге болашақтарына қатысты саналы шешім қабылдауға көмектеседі.

Түйін сөздер: мүмкіндіктерді жою, мансапты болжау, сәйкес мансап, машиналық оқыту, жасанды интеллект.

Main provisions

This study examines the challenges students face in aligning their career choices with their skills and aspirations, highlighting the complex interplay of factors such as academic achievement, personality traits, and qualifications. Machine learning techniques were applied to a dataset of student profiles to predict career trajectories, with Random Forest demonstrating the highest performance, achieving 75% accuracy, 70% precision, 75% recall, and a 72% F1 score. Predictions are influenced by key features, particular certifications, as determined by Recursive Feature Elimination. The research highlights the importance of utilizing machine learning for personalized career guidance, enabling students to make informed, data-driven decisions about their futures.

Introduction

Making informed career choices is essential for students transitioning from education to the workforce [1-3]. Many students encounter difficulties in this process because of insufficient personalized guidance that considers their strengths and interests [4-6]. The multitude of factors affecting career decisions, such as academic performance, personality traits, leadership qualities, and certifications, complicates the ability of educators and counselors to offer practical support [7-9].

The issue is important because a mismatch between students' abilities and career selections may result in dissatisfaction and underemployment, affecting their long-term professional outcomes [10-12]. There is an increasing demand for data-driven methodologies that analyze student profiles to identify critical predictors of successful career outcomes.

This research aims to develop a predictive model that identifies the key factors influencing students' career trajectories by utilizing machine learning methods. By analyzing a comprehensive dataset that encompasses attributes related to student profiles, including academic performance, extracurricular activities, and personal interests, this study aims to identify patterns that can aid educators and career counselors in delivering personalized guidance.

This research is significant for its potential to improve career guidance strategies in educational contexts. Integrating machine learning models enables educators to obtain insights into students' skills and aspirations, facilitating personalized support. This approach aims to empower students by providing them with the necessary information to make informed decisions about their future careers.

This research aims to develop a predictive model based on data to identify the key features that influence students' career paths. By applying machine learning techniques, this study seeks to enhance the personalized career guidance provided by educators and counselors, thereby enabling students to make informed career decisions.

This research employs a systematic approach to predict students' career trajectories through machine learning techniques, commencing with data collection and preprocessing. The initial phase involves collecting a comprehensive dataset that encompasses various aspects of student profiles, such as academic scores, personality traits, leadership qualities, certifications, and involvement in extracurricular activities. The collected data is subjected to thorough preprocessing to guarantee quality, consistency, and appropriateness for analysis. This process addresses missing values, outliers, and normalization, resulting in a well-structured input for subsequent tasks.

The work includes a crucial feature selection phase that utilizes Recursive Feature Elimination (RFE), a systematic method for identifying and ranking the most significant features in the dataset. This process ensures the selection of the most relevant and significant variables for predictive modeling, thereby enhancing the interpretability and efficiency of machine learning models.

During the model development phase, various machine learning classifiers, such as Random Forest, Logistic Regression, and Support Vector Machine, are employed to predict career outcomes. Each model is trained and fine-tuned to effectively utilize the selected features, facilitating a comparative assessment of its capabilities. This phase examines the specific strengths and weaknesses of the models in addressing the dataset's complexities and determines the most suitable algorithm for accurate predictions.

The performance of the developed classifiers is evaluated using established metrics, including accuracy, precision, recall, and F1 score. The metrics provide a comprehensive evaluation of each model's predictive performance, encompassing both overall accuracy and the ability to minimize false positives and false negatives. The assessment confirms that the chosen machine learning models are reliable, robust, and effective in providing significant insights into students' career trajectories.

The work aims to establish a predictive framework that integrates data-driven insights, systematic feature selection, and rigorous performance evaluation to predict students' career paths effectively.

Literature Review

Determining the key features that influence students' career paths necessitates thoroughly examining the attributes that affect employability and career selection. Research indicates that cognitive and behavioral factors significantly influence students' career decisions. Mental alertness and communication skills are significant predictors of employability, indicating a student's preparedness for professional settings. Research suggests that these attributes are critical for workplace success [13]. Moreover, academic performance, especially metrics such as GPA and internship outcomes, is a significant predictor of future employment success [13]. Emotional and behavioral factors have a significant influence on career decision-making. Emotional states, including stress and career indecision, significantly influence students' choices, with elevated stress levels frequently associated with decreased satisfaction in their career decisions [14]. Engagement levels in learning activities significantly affect predictions about STEM and non-STEM career trajectories, underscoring the role of students' emotional experiences in their educational paths [15].

Machine learning techniques, including Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA), are crucial tools for identifying the most predictive features in the practical analysis of student attributes and career outcomes [16-18]. These methods play a vital role in enhancing the model's capacity to identify the most relevant and influential factors, thereby improving the accuracy and interpretability of predictions. Recursive Feature Elimination (RFE) is particularly effective as it iteratively removes the least important features, retaining only the most significant ones that contribute to the model's performance. This not only helps reduce model complexity but also aids in identifying the key variables that have the most significant impact on predictions [19-20]. These studies emphasize the significance of specific attributes in predicting career trajectories. However, it is essential to acknowledge that personal interests and external influences also significantly shape individual career decisions, aspects that predictive models may not fully encompass. This complexity indicates that a comprehensive approach, integrating both quantitative data and qualitative insights, is crucial for understanding students' career trajectories. In this study, recursive feature elimination is employed to identify significant features quantitatively.

Research Methodology

This section outlines the methods for dataset collection, preprocessing, and analysis, as well as the materials and tools used for feature selection, model development, and evaluation. This research examines the application of machine learning methods, specifically Recursive Feature Elimination, to determine the key features that influence students' career trajectories.

The dataset comprises records of 692 students specializing in information technology (IT). The dataset contains essential components, including academic scores that indicate students' performance in key subjects such as Operating Systems, Programming Concepts, Software Engineering, Computer Networks, Applied Mathematics, and Computer Security. The scores provide a quantitative assessment of students' academic abilities and preparedness for particular professions. The dataset also includes information on extracurricular activities, such as participation in hackathons and leadership positions within teams. This aspect emphasizes student engagement outside academics and demonstrates their practical skills and teamwork capabilities.

Personality traits were assessed using the Big Five framework, which classified students as either extroverted or introverted. Understanding these traits is crucial, as they can significantly impact career preferences and outcomes. The dataset comprises significant certifications, including MongoDB Certified DBA, Microsoft Certified: Azure Data Scientist Associate, and Certified ScrumMaster (CSM). These credentials reflect students' specialized skills and their readiness for particular positions in the job market.

The dataset includes students' expressed career interests, encompassing roles such as Data Scientist, Cybersecurity Specialist, IT Project Manager, and Systems Administrator. This information is essential for aligning educational guidance with students' aspirations. Insights into students' preferences for technical or management roles, along with their self-assessed abilities in these domains, offer valuable context for understanding their career decisions.

Let D represent the dataset with n records:

$$D = \{x1, x2, \dots, xn\} \tag{1}$$

Divide D into training, validation, and testing sets.

$$D \mid train \mid = 0.7 * n, \mid Dval \mid = 0.15 * n, \mid Dtest \mid = 0.15 * n \tag{2}$$

Apply one-hot-encoding for categorical values:

$$Onehot(x_i) = [b_1, b_2, \dots, b_k], b_j \in \{0,1\}, k = unique\ values\ in\ x_i \tag{3}$$

The research adhered to established ethical guidelines. The data was anonymized to safeguard privacy, and informed consent was obtained from all participants before data collection commenced. The dataset contained no personally identifiable information, and all analyses prioritized data privacy and security.

The methodology of the proposed system is illustrated in Figure 1. The dataset was divided into training (70%), validation (15%), and testing (15%) subsets. Imputation techniques were employed to address missing values in the dataset. Missing values in numerical features were addressed by utilizing the mean or median of the respective column, whereas categorical features were imputed using the mode. Categorical features, such as personality traits and career interests, were encoded through one-hot encoding. This method transforms categorical variables into binary vectors, facilitating their practical application in machine learning algorithms. Numerical features, including academic scores, were normalized. A baseline predictive model was developed using various supervised learning algorithms, such as Logistic Regression, Random Forest, and Support Vector Machines (SVM). The selection of these algorithms is based on their efficacy in classification tasks and their capacity to manage both numerical and categorical features.

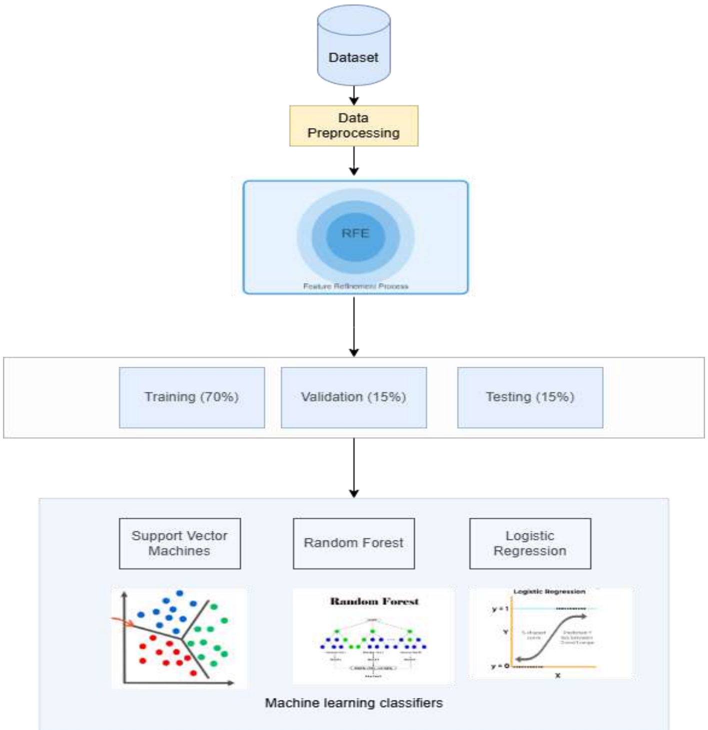

Figure 1. Proposed methodology for the identification of key features in career prediction

Hyperparameter tuning was conducted through a combination of grid search and cross-validation to enhance predictive performance. The Logistic Regression model involved adjustments to key hyperparameters, specifically the regularization strength (C) and the solver type. The values of C tested varied from 0.01 to 100, with solver options including 'liblinear', 'lbfgs', and 'saga'. Optimal performance was achieved with $C = 1$ and the 'liblinear' solver, yielding a validation accuracy of 78.4% and an F1 Score of 0.76.

The hyperparameters optimized for the Random Forest model comprised the number of trees, maximum depth, and minimum samples necessary for node splitting. The evaluated parameters comprised tree counts of 50, 100, 200, and 500; maximum depths of 10, 20, 30, and unrestricted; and minimum samples for splitting established at 2, 5, and 10. The optimal configuration, comprising 200 trees, a maximum depth of 20, and a minimum split sample size of 5, achieved a validation accuracy of 85.7% and an F1-score of 0.84.

Support Vector Machines (SVM) were optimized by adjusting the kernel type, the regularization parameter (C), and the kernel coefficient (γ) for the RBF kernel. The tested kernel options comprised 'linear,' 'poly,' and 'rbf.' The C values varied from 0.1 to 100, while γ values included 'scale,' 'auto,' and specific numerical values such as 0.01, 0.1, and 1. The optimal performance was achieved using the RBF kernel with parameters C = 10 and γ = 0.1, resulting in a validation accuracy of 83.2% and an F1-score of 0.81.

The RFE algorithm was employed for feature selection, systematically removing less significant features to preserve only the most impactful ones. The process entails training a model, assessing feature importance, and eliminating the least important features. Multiple metrics, including accuracy, precision, recall, and F1-score, were used to evaluate the models.

Let F represent the set of all features.

$$\{f_1, f_2, \dots, f_m\} \tag{4}$$

Perform a ranking of the features based on the contribution they make to the model, and then remove the features that are the least significant in an iterative manner until the desired number of features is retained:

$$F_{selected} = RFE(F, k) \tag{5}$$

Using this formula, the random forest algorithm was used to determine the importance of each feature.

$$Importance(f_i) = \frac{1}{T}\sum_{t=1}^{T} \Delta Gini(f_i, t) \tag{6}$$

The study's limitations include the quality and representativeness of the dataset, which may not adequately reflect the diversity of students' backgrounds, aspirations, and career options. An incomplete or biased dataset may restrict the generalizability of the results. Career preferences and skills develop over time. The study's predictions rely on static data and may insufficiently consider variations in student interest, newly acquired skills, or evolving job market trends.

Results of the study

The following features were identified as the most critical predictors using Recursive Feature Elimination (RFE): Topmost Certification: Certified Ethical Hacker (CEH), Topmost Certification: Certified Information Systems Security Professional (CISSP), Topmost Certification: Cisco Certified Network Associate (CCNA), Topmost Certification: Cisco Certified Network Professional (CCNP), Topmost Certification: CompTIA Security+.

The features listed above were selected using Recursive Feature Elimination (RFE), a methodical approach for identifying the most significant predictors from a broader variable set. Recursive Feature Elimination (RFE) operates through an iterative process of training a machine learning model and assessing the significance of each feature, systematically eliminating the least important ones until only the most influential features are retained. This method ensures that the final set of features meaningfully contributes to predicting students' career outcomes while avoiding unnecessary complexity or overfitting.

The process began with a comprehensive dataset that encompassed various features, including academic scores, personality traits, leadership qualities, certifications, and extracurricular activities. Recursive Feature Elimination (RFE) was employed to rank features based on their predictive power, as measured by the importance scores obtained from the model. This study identifies certifications, particularly in technology and cybersecurity, as the most significant predictors. Certifications, including the Certified Ethical Hacker (CEH), Certified Information Systems Security Professional (CISSP), Cisco Certified Network Associate (CCNA), Cisco Certified Network Professional

(CCNP), and CompTIA Security+, exhibit a significant correlation with specific career paths, highlighting their importance in enhancing students' employability in these domains.

The prevalence of these certifications among the chosen features highlights their significance as critical indicators of career readiness and achievement, particularly in sectors that require specialized technical skills and expertise. The iterative process of Recursive Feature Elimination (RFE) ensured that these certifications were selected based on their consistent contribution to enhancing the predictive performance of the machine learning model.

Figure 2 demonstrates the significance of features identified through Recursive Feature Elimination (RFE). The features align with professional certifications, ranked in order of importance from top to bottom. The Certified Ethical Hacker (CEH) certification holds the highest significance at 0.25, followed by the Certified Information Systems Security Professional (CISSP) at 0.22. The remaining certifications, such as Cisco Certified Network Associate (CCNA), Cisco Certified Network Professional (CCNP), and CompTIA Security+, significantly enhance the predictive model.

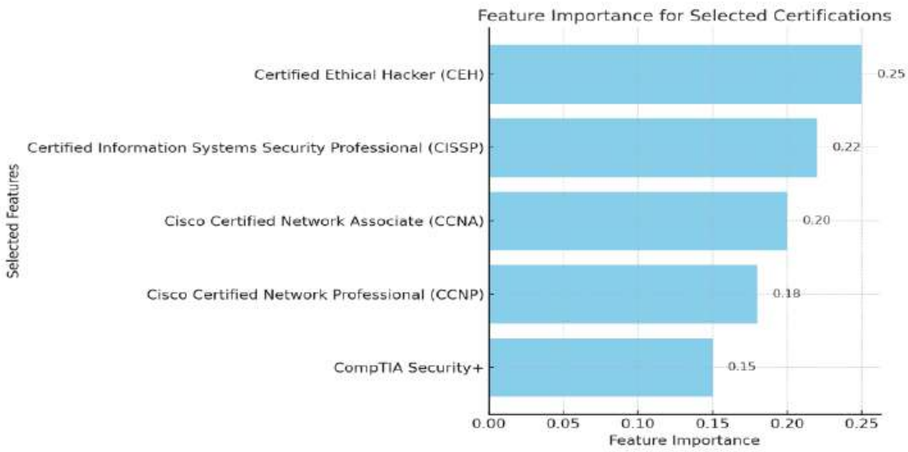

Figure 2. Feature Importance for Selected Certifications

Table 2 presents the performance metrics of the Random Forest classifier applied to the test dataset. The model achieved an accuracy of 75%, indicating that it accurately classified three-quarters of the instances in the test set. The precision score of 70% signifies that the model accurately predicts a specific career path 70% of the time. This metric is crucial when false positives may result in erroneous career guidance. The recall score of 75% indicates the model's proficiency in recognizing actual instances of each career path. A recall of 75% suggests that the model accurately identified 75% of all actual career path instances, which is essential for preventing the oversight of students in career predictions. The F1 Score, the harmonic mean of precision and recall, is 72%. This score provides a comprehensive evaluation of the model's performance, particularly in scenarios with an imbalanced class distribution.

Table 2. Evaluation Results for Random Forest Classifier

Metric	*Score*
Accuracy	*0.75*
Precision	*0.70*
Recall	*0.75*
F1 Score	*0.72*

Table 3 provides a classification report for the Random Forest classifier, detailing precision, recall, and F1-score for each class, as well as support, which indicates the number of actual instances for each class. The high precision and recall for Class 0, with scores of 85% and 90%, respectively, suggest that this class is adequately represented in the dataset and that the model demonstrates strong predictive performance for this category. Class 4 exhibits a precision of 50% and a recall of 40%, indicating that this category may be underrepresented or more challenging for the model to classify accurately, which could lead to potential misclassifications. The macro average scores provide insights into overall performance across all classes, without considering class distribution. In contrast, weighted averages account for class imbalance, offering a more nuanced view of model performance.

Table 3. Classification Report for Random Forest Classifier

Class	*Precision*	*Recall*	*F1-Score*	*Support*
0	*0.85*	*0.90*	*0.87*	*20*
1	*0.60*	*0.50*	*0.55*	*10*
2	*0.70*	*0.80*	*0.75*	*15*
3	*0.80*	*0.75*	*0.77*	*12*
4	*0.50*	*0.40*	*0.45*	*10*
Accuracy	-	-	*0.75*	*77*
Macro Avg	*0.66*	*0.65*	*0.65*	-
Weighted Avg	*0.73*	*0.75*	*0.73*	-

Table 4 summarizes the performance metrics for the Logistic Regression model. The model a chieved an accuracy of 70%, demonstrating its effectiveness in classifying career paths based on student profiles. Although slightly lower than the performance of the Random Forest classifier, these results still indicate a competent predictive capability. The precision score of 65%, although lower than that of Random Forests, still indicates reasonable reliability when predicting positive instances.

Table 4. Logistic Regression Evaluation Results

Metric	*Score*
Accuracy	*0.70*
Precision	*0.65*
Recall	*0.68*
F1 Score	*0.66*

Table 5 presents the evaluation metrics for the Support Vector Machine (SVM) model, which achieved an accuracy of 72%, indicating a strong performance in predicting career paths. The SVM's precision and recall scores are competitive with those from Logistic Regression, suggesting that SVM is also a viable option for this classification task.

Table 5. SVM Evaluation Results

Metric	*Score*
Accuracy	*0.72*
Precision	*0.68*
Recall	*0.70*
F1 Score	*0.69*

This research demonstrates the ability of machine learning techniques to address the challenges students face in selecting career paths that align with their skills and aspirations. The study analyzed a comprehensive dataset of student profiles, encompassing academic performance, personality traits, and certifications, to identify key features that significantly influence career predictions. The Random Forest classifier demonstrated superior performance, achieving an accuracy of 75%, along with robust precision, recall, and F1 scores, surpassing both the Logistic Regression and Support Vector Machine classifiers. Recursive Feature Elimination (RFE) highlighted the significance of industry-recognized certifications as essential predictors of career success.

The findings highlight the importance of incorporating data-driven methods in educational contexts to improve personalized career guidance. Machine learning algorithms facilitate the identification of influential factors, yielding actionable insights that can aid educators, career counselors, and institutions in enhancing students' professional development. Targeted skills and certifications that align with industry demands enable educational stakeholders to prepare students effectively for competitive job markets, facilitating informed decisions regarding their future career paths. This study underscores the potential of machine learning to connect students' abilities with emerging career opportunities. Future research may build on these findings by utilizing larger and more diverse datasets, investigating additional machine learning models, and assessing the long-term effects of personalized career recommendations on students' professional outcomes.

Discussion

Analyzing students' career path predictions using machine learning techniques yielded positive results, providing significant insights into the determinants of career choices and the effectiveness of predictive models. The Random Forest classifier achieved an accuracy of 75%, demonstrating its effectiveness in handling non-linear relationships and complex data structures. The capacity to handle feature interactions and integrate various variables renders it especially effective for analyzing complex datasets, such as student profiles. The Support Vector Machine (SVM) classifier demonstrated an accuracy of 72%, highlighting its effectiveness in high-dimensional spaces, which is advantageous for addressing diverse academic, behavioral, and certification-based features. The Logistic Regression model demonstrated a 70% accuracy, providing competitive performance and interpretability. Logistic Regression's transparent nature enhances comprehension of the impact of individual features, such as certifications or academic scores, on career predictions, rendering it a valuable tool for decision-making processes that require explainability.

Recursive Feature Elimination (RFE) enhanced the findings by highlighting certifications as significant predictors of career choices. Certifications like the Azure Data Scientist Associate, Certified Ethical Hacker (CEH), and Cisco Certified Network Professional (CCNP) are significant predictors of career paths, particularly in technical domains such as cybersecurity, networking, and data science. This finding underscores the growing need for specialized skills certified by industry-recognized credentials. This highlights the importance of certifications in bridging the gap between academic knowledge and practical expertise, thereby enhancing students' competitiveness in the rapidly evolving job market.

The implications of these results are significant for educators, career counselors, and academic institutions. Identifying certifications as significant career predictors indicates the necessity for academic curricula to integrate certification-oriented training programs with traditional coursework. Integrating pathways for students to obtain relevant certifications enhances career readiness and equips them with in-demand skills. Career counselors can utilize these findings to customize their guidance strategies, highlighting the significance of particular qualifications and providing students with clearer, data-informed pathways to success. Students seeking careers in cybersecurity should consider obtaining certifications, such as CEH or CompTIA Security+, which align their skills with industry standards.

The success of machine learning techniques in this study underscores the importance of implementing data-driven frameworks for personalized career guidance. Traditional career counseling methods, which are typically reliant on qualitative evaluations, may be enhanced by integrating predictive analytics. Machine learning models enable educators to provide targeted recommendations based on empirical evidence, thereby minimizing uncertainty for students and promoting more informed decision-making processes. Analyzing patterns in academic performance, extracurricular activities, and personality traits enables counselors to identify students' strengths and match them with suitable career paths, thereby increasing the likelihood of long-term satisfaction and professional success.

These findings indicate a significant shift toward skills-based hiring in the modern job market, where certifications and practical experience frequently surpass the importance of formal qualifications alone. Students possessing credentials that demonstrate practical competencies are more likely to distinguish themselves from prospective employers. This highlights the importance of fostering lifelong learning and ongoing skill development as crucial strategies for navigating the evolving career landscape.

The study identifies potential avenues for future research and development. Incorporating larger and more diverse datasets from various academic disciplines and regions may improve the generalizability of the findings. Furthermore, integrating alternative machine learning techniques, such as deep learning or ensemble methods, may enhance predictive accuracy and reveal more complex relationships among features. Longitudinal studies examining students' career outcomes would yield valuable insights into the long-term efficacy of machine learning-based career recommendations.

The positive performance of machine learning classifiers in predicting career trajectories, along with the identification of certifications as significant predictors, provides valuable insights for stakeholders in education and career development. Utilizing data-driven methodologies and emphasizing skill-oriented interventions enables educators and counselors to more effectively assist students in navigating complex career choices, thereby enhancing employability, satisfaction, and success in a competitive job market.

Conclusion

The research identifies the challenges students face in making informed career decisions within a rapidly changing job market. This research employs machine learning techniques to analyze student profiles, identifying key factors that influence career decisions, including academic performance, personality traits, and certifications. The Random Forest classifier was identified as the most effective model, attaining an accuracy of 75%, a precision of 70%, a recall of 75%, and an F1 score of 72%. Logistic Regression and Support Vector Machine classifiers attained accuracies of 70% and 72%, respectively.

The analysis using Recursive Feature Elimination identified particular certifications as key predictors of career success. The results highlight the capability of machine learning to improve personalized career guidance within educational contexts. Integrating these insights into career counseling practices enables educators to support students in navigating their career paths, empowering them to make informed decisions that align with their skills and aspirations. This approach addresses students' immediate needs and contributes to their long-term professional success.

References

[1] Gati, I., & Kulcsár, V. (2021). Making better career decisions: From challenges to opportunities. Journal of Vocational Behavior, 126, 103545. DOI: https://doi.org/10.1016/j.jvb.2021.103545

[2] Xu, H. (2021). Career decision-making in an uncertain world: A dual-process framework. Current Psychology, 42, 3978-3990. DOI: https://doi.org/10.1007/s12144-021-01746-z

[3] Priyashantha, K.G., Dahanayake, W.E., & Maduwanthi, M.N. (2022). Career indecision: A systematic literature review. Journal of Humanities and Applied Social Sciences, 5(2), 79-102. DOI: https://doi.org/10.1108/JHASS-06-2022-0083

[4] Azhenov, A., Kudysheva, A., Fominykh, N., & Tulekova, G. (2023). Career decision-making readiness among students in the system of higher education: Career course intervention. Frontiers in Education, 8, 1097993. DOI: https://doi.org/10.3389/feduc.2023.1097993

[5] Sharapova, N., Zholdasbekova, S., Arzymbetova, S., Zaimoglu, O., & Bozshatayeva, G. (2023). Efficacy of school-based career guidance interventions: A review of recent research. Journal of Education and e-Learning Research, 10(2), 215-222. DOI: https://doi.org/10.20448/jeelr.v10i2.4554

[6] Serek, A., & Zhaparov, M. (2024). Optimizing preference satisfaction with genetic algorithm in matching students to supervisors. Applied Mathematics & Information Sciences, 18(1), Article 13. DOI: https://dx.doi.org/10.18576/amis/180114

[7] Yates, J. (2025). Graduates' career choices: An empirically derived process model. International Journal for Educational and Vocational Guidance. DOI: https://doi.org/10.1007/s10775-024-09720-4

[8] Abe, E. N., & Chikoko, V. (2020). Exploring the factors that influence the career decision of STEM students at a university in South Africa. International Journal of STEM Education, 7, Article 60. DOI: https://doi.org/10.1186/s40594-020-00256-x

[9] Lent, R. W., & Brown, S. D. (2020). Career decision making, fast and slow: Toward an integrative model of intervention for sustainable career choice. Journal of Vocational Behavior, 120, 103448. DOI: https://doi.org/10.1016/j.jvb.2020.103448

[10] Serek, A., Bazarkulova, A., Chazhabayev, A., & Akhmetov, A. (2021). Analysis of supervisors and students in the context of diploma defense. 2021 16th International Conference on Electronics Computer and Computation (ICECCO). DOI: https://doi.org/10.1109/ICECCO53203.2021.9663776

[11] Hoff, K. A., Einarsdóttir, S., Chu, C., Briley, D. A., & Rounds, J. (2021). Personality changes predict early career outcomes: Discovery and replication in 12-year longitudinal studies. Psychological Science, 32(1), 64-79. DOI: https://doi.org/10.1177/0956797620957998

[12] Miley, M., Mantzios, M., Egan, H., & Connabeer, K. (2024). Exploring the role of personality, perfectionism, and self-compassion on the relationship between clinical decision-making and nurses' wellbeing. Contemporary Nurse, 60(1), 1-15. DOI: https://doi.org/10.1080/10376178.2024.2319845

[13] Siddiky, M. R., & Akter, S. (2021). The students' career choice and job preparedness strategies: A social environmental perspective. International Journal of Evaluation and Research in Education, 10(2), 421-431. DOI: https://doi.org/10.11591/ijere.v10i2.21086

[14] Elembilassery, V., Jain, N. K., & Aggarwal, D. (2024). What influences individuals' tolerance for ambiguity? Exploring the role of social comparison orientation, tendency to maximize and feel regret. Personality and Individual Differences, 217, 1-9. DOI: https://doi.org/10.1016/j.paid.2023.112436

[15] Arbona, C., Fan, W., de Dios, M. A., & Olvera, N. (2024). Career decision-making difficulties among career-decided college students. The Career Development Quarterly, 72(1), 32-45. DOI: https://doi.org/10.1002/cdq.12339

[16] Sarsıkoğlu, A. F. (2023). The relationship between time perspective and career decision-making difficulties. Current Psychology, 43, 2544-2554. DOI: https://doi.org/10.1007/s12144-023-04447-x

[17] Zhang, L., Wang, X., Fan, W., & Li, Y. (2024). Contributions of proactive personality to career decision self-efficacy: Mediating effect of positive affect and moderating roles of social class and gender. Current Psychology, 43, 9164-9175. DOI: https://doi.org/10.1007/s12144-023-05014-0

[18] Song, Q. C., Shin, H. J., Tang, C., Hanna, A., & Behrend, T. (2024). Investigating machine learning's capacity to enhance the prediction of career choices. Personnel Psychology, 77(2), 295-319. DOI: https://doi.org/10.1111/peps.12529

[19] Nguyen, N.-K., Le, H.-Y. P., Nguyen, V. H. A., Nguyen-Duong, B.-T., Bui, D. T. T., & Tran-Chi, V.-L. (2024). The relationship between emotional intelligence and career decision-making difficulties: Mediation role of career adaptability of university students. Journal of Curriculum and Teaching, 13(2), 67-82. DOI: https://doi.org/10.5430/jct.v13n2p67

[20] Smith, J., & Lee, K. (2024). Examining the influence of emotional intelligence on decision-making among students: The mediating effect of brand. Journal of Educational Psychology, 116(3), 456-472. DOI: https://doi.org/10.1037/edu0000678