ARTIFICIAL INTELLIGENCE IN EDUCATION: A SYSTEMATIC LITERATURE REVIEW OF MACHINE LEARNING APPROACHES IN STUDENT CAREER PREDICTION

Fabricio Trujillo[*] (ID)**, Marcelo Pozo** (ID)**, Gabriela Suntaxi** (ID)

Escuela Politécnica Nacional (Ecuador)

*Corresponding author: danilo.trujillo@epn.edu.ec
marcelo.pozo@epn.edu.ec, gabriela.suntaxi@epn.edu.ec

Abstract

This paper presents a systematic literature review of using Machine Learning (ML) techniques in higher education career recommendation. Despite the growing interest in leveraging Artificial Intelligence (AI) for personalized academic guidance, no previous reviews have synthesized the diverse methodologies in this field. Following the Kitchenham methodology, we analyzed 38 studies selected from an initial pool of 1,296 articles, retrieved using a custom-built web scraper leveraging the CrossRef API. Data were extracted based on ML techniques, data types, and validation metrics. Our findings reveal that Random Forest, Support Vector Machines (SVM), and Neural Networks are the most frequently employed models to improve the accuracy and personalization of career recommendations in higher education. These systems typically use academic performance, personal interests, and demographic data as the primary data types. The review also highlights key validation metrics like precision, recall, and F1-score, which reflect the effectiveness of these models. However, limitations were identified, such as the lack of access to open datasets and the scarcity of studies with longitudinal data that evaluate the long-term impact of recommendations. Additionally, ethical considerations, particularly regarding fairness, transparency, and data privacy, were highlighted as critical challenges. This systematic literature review provides a solid foundation for improving career recommendation systems using advanced ML techniques. By integrating ML with traditional counseling approaches, this research underscores the potential to revolutionize academic guidance and better align students with their career goals.

Keywords – Systematic review, Web scraper, Machine learning, Career recommendation, Higher education, Predictive modeling.

To cite this article:

----

1. Introduction

In recent decades, artificial intelligence, particularly machine learning (ML), has revolutionized various fields, including higher education. One of the most promising applications in this domain is the prediction

of student careers, a process that seeks to identify the most suitable academic trajectories for students based on the analysis of large volumes of historical and behavioral data (Song, Shin & Shin, 2024).

In higher education, decision-making regarding academic careers is one of the most critical processes that students face. Traditionally, this process has been based on guidance from academic counselors, students' personal aspirations, and, in some cases, external influences such as family expectations and labor market trends. However, in recent years, a new perspective has emerged based on using Machine Learning algorithms for career prediction, aiming to offer more personalized recommendations based on historical and behavioral data. This approach can potentially revolutionize the way vocational guidance is conducted, offering predictions that not only reflect past academic performance but also other complex factors that may influence a student's future success in a particular career (Maulana, Idroes, Kemala, Maulydia, Sasmita, Tallei et al., 2023).

The purpose of this study is to conduct a comprehensive systematic review of the literature on the use of Machine Learning techniques in predicting and recommending higher education careers. This research aims to synthesize the current state of knowledge in this field, identify the most effective ML techniques, analyze the data types used, and evaluate the validation metrics employed in these predictive models. The context of this study is particularly relevant given the increasing complexity of the job market and the growing need for precise, data-driven career guidance in higher education. The justification for this research lies in the potential of ML to revolutionize career counseling by providing more accurate, personalized, and scalable recommendations to students. By critically analyzing the existing body of literature, this study seeks to inform future research directions, highlight current limitations, and provide insights that can lead to the development of more effective career prediction systems in higher education.

The availability of large volumes of educational data and advances in data processing techniques have driven the predictive capacity of ML in this context. These techniques allow the analysis of complex patterns in student data, such as their course choices, academic performance, extracurricular interests, and even socioeconomic factors, to identify the most promising academic trajectories (Musso, Hernández & Cascallar, 2020). Recent studies have shown how ML models can accurately surpass traditional academic counseling methods by incorporating a greater diversity of variables in their analyses (Hilbert, Coors, Kraus, Bischl, Lindl, Frei, et al., 2021). For example, (Song et al., 2024) demonstrated that applying ML models in career selection can significantly improve the match between students' skills and selected careers, thus reducing dropout rates and improving academic outcomes.

We use the Kitchenham methodology (Kitchenham, Pearl-Brereton, Budgen, Turner, Bailey & Linkman, 2009) to conduct a system literature review on student career prediction using ML models to identify the most effective techniques, challenges faced, and future opportunities in this field of research. This method is widely used in software engineering and has been adapted to conduct systematic literature reviews in other fields, including educational research. This method provides a rigorous structure for identifying, evaluating, and synthesizing relevant literature, allowing researchers to obtain a comprehensive and critical view of the current state of knowledge in a specific area (Koval, Knollmeyer, Mathias, Asif, Uzair-Akmal, Grossmann et al., 2024).

The use of ML in career prediction has generated growing interest in the academic community, resulting in a proliferation of studies exploring various techniques and approaches. The literature review shows that there is no clear consensus on which ML technique is most effective for this purpose. However, supervised learning techniques, such as logistic regression, support vector machines (SVM), and decision trees, are commonly used due to their ability to handle categorical and continuous data (Kuzey, Uyar & Delen, 2019). Nevertheless, more recent advances have seen an increase in the use of more sophisticated approaches, such as deep neural networks and ensemble models, like Random Forests, which have shown to be effective in capturing non-linear and complex relationships between predictor variables and career outcomes (Badal & Sungkur, 2023).

A crucial aspect of the effectiveness of these models is the selection of predictor variables. Studies have identified that in addition to academic grades, factors such as online behavior, participation in extracurricular activities, and demographic data play a significant role in career prediction (Namoun & Alshanqiti, 2021). Namoun and Alshanqiti highlighted the importance of using a supervised learning approach that not only relies on academic data but also includes variables related to the student's motivation and personal interest, which can offer a more holistic and accurate prediction.

Despite the demonstrated potential of ML models in career prediction, several challenges limit their widespread adoption. One of the main challenges is the quality and availability of data. In many educational institutions, relevant data is fragmented or not collected uniformly, which can introduce biases in predictive models (Himanen, Geurts, Foster & Rinke, 2019). Additionally, the lack of standardization in data collection and processing between different institutions makes it difficult to compare results across studies directly.

Another significant challenge is interpreting results generated by complex models, such as deep neural networks. Although accurate, these models often function as a "black box," meaning it is difficult for users to understand how decisions are made (Baker & Hawn, 2022). In the educational context, where transparency and justification of recommendations are essential for their acceptance by students and educators, this lack of interpretability can limit trust in ML-based systems.

Furthermore, studies have pointed out the need to develop models that are not only accurate but also fair. There are concerns about potential biases inherent in ML algorithms, which may perpetuate existing inequalities if not properly managed (Zhang, Lee, Ali, DiPaola, Cheng & Breazeal, 2023). For example, if an ML model is trained on a dataset that reflects historical or social biases, it could replicate these biases in its predictions, leading to unfair or discriminatory decisions in vocational guidance (Barredo-Arrieta, Díaz-Rodríguez, Del-Ser, Bennetot, Tabik, Barbado et al., 2020).

Despite these challenges, the opportunities for using ML in career prediction are vast. The integration of ML with other emerging technologies, such as Big Data and explainable artificial intelligence (XAI), promises to improve both the accuracy and transparency of predictions (Tasmin, Muhammad & Nor-Aziati, 2020). The use of Big Data allows incorporating more data into predictive models, including unstructured data such as essay texts and social media posts, which could offer a more comprehensive view of a student's academic potential (Farrow, 2023).

On the other hand, explainable artificial intelligence (XAI) seeks to make ML models more interpretable and transparent. This is particularly relevant in education, where decisions must be understandable and justifiable. Implementing XAI could help educators and academic counselors better understand the recommendations generated by ML models, which would facilitate their adoption in educational settings (Kaspersen, Bilstrup, Van Mechelen, Hjort, Bouvin & Petersen, 2022).

Another area that promotes future research is the development of ML models capable of adapting to different educational contexts. Many models are designed to function within a specific context, which limits their applicability in different environments. Creating more generalizable models that can adjust to the particularities of different institutions and student populations is an important challenge for the future of ML-based vocational guidance (Braiki, 2023).

The lack of a comprehensive review of methodologies for applying ML models to predict career pathways within higher education highlights a notable gap in the field. Research by (Namoun & Alshanqiti, 2021) and (Badal & Sungkur, 2023) illustrates how ML models outperform traditional educational models by delivering personalized recommendations based on diverse datasets. (Chen, Chen & Lin, 2020) presented a comprehensive review of Artificial Intelligence (AI) in education, highlighting its impact on administrative tasks, instructional methods, and learning processes. These works provide valuable insights into predicting students' performance and ML-based applications for education but lack a systematic synthesis of methodologies specific to career prediction.

While (Chen et al., 2020) explored general applications of AI in education, their work does not focus on consolidating methodologies or addressing specific challenges associated with ML-based career recommendation systems. This gap leaves a fragmented understanding of how ML techniques can enhance career prediction, limiting scalability, ethical integration, and practical implementation. (Hilbert et al., 2021) further emphasize that the absence of synthesized research hinders progress in addressing critical issues such as data quality, interpretability, and the ethical challenges of using ML in education.

Addressing this gap through a systematic review is crucial. By focusing on the methodologies, data types, and validation metrics used in career prediction systems, this work aims to provide a targeted contribution to the field. Such efforts can offer practical insights to enhance student outcomes, reduce dropout rates, and foster better career alignment, enabling more scalable and ethically sound ML-based applications in higher education.

The remaining of the document is organized as follows. Section 2 describes the methodology used to conduct the systematic literature review, including the search strategy, inclusion and exclusion criteria, and data extraction process. Section 3 presents the results of the review, focusing on the most frequently used Machine Learning techniques, types of data, and validation metrics in career recommendation systems. Section 4 discusses the implications of these findings, highlighting both the strengths and limitations of current approaches, as well as potential ethical challenges. Section 5 concludes the paper by summarizing the key contributions, identifying research gaps, and proposing future research directions to improve the use of ML in student career prediction.

2. Methodology

This research employed a systematic literature review methodology following the guidelines proposed by Kitchenham (Kitchenham et al., 2009). The review process was structured into three main phases: planning, conducting, and reporting the review.

2.1. Planning the Review

In this phase, we identify the need for the review and describe the review protocol.

2.1.1. Identification of the Need for Review

Despite the increasing interest in applying machine learning to predict student career outcomes, a preliminary search reveals a significant gap in the literature. There is no comprehensive review that consolidates and evaluates the various models used in this area. Given the rapid advancements in machine learning techniques and their potential to enhance decision-making in education and career counseling, this gap represents a missed opportunity for researchers and educators. Therefore, a systematic literature review is crucial to provide an overview of the methodologies, highlight the most effective approaches, and identify challenges and areas for future research. This review will offer a critical synthesis of the existing work, helping to guide both academic inquiry and practical implementation in predicting student career trajectories.

2.1.2. Development of the Review Protocol
a) Research Questions

We start by formulating the following research questions:

RQ1: What Machine Learning techniques are used in career recommendation systems for higher education?

RQ2: What data types are used to train models to recommend higher education careers?

RQ3: What validation metrics are used to evaluate the effectiveness of these recommendation systems?

b) Search Strategy

The search terms were designed to capture relevant studies, combining key concepts from machine learning, academic advising, and higher education. Specifically, the following search phrase was used: [((("machine learning" OR "artificial intelligence") AND ("career recommendation" OR "major selection" OR "academic advising") AND ("higher education" OR "university"))]

c) Data Sources

To ensure a thorough and systematic review of the literature on machine learning models for student career prediction, we employed a comprehensive search strategy using a combination of academic digital libraries and repositories: Scopus, IEEE, MDPI, IOPscience, ERIC, EBSCO, Web of Science, Sciendo, ResearchGate, arXiv, Google Scholar, and doctoral thesis repositories.

d) Inclusion and Exclusion Criteria

We define the following inclusion and exclusion criteria:

Inclusion criteria:

- Articles published between 2010 and 2024
- Articles written in English
- Studies using Machine Learning techniques to recommend university careers or majors

Exclusion criteria:

- Non-peer-reviewed articles
- Duplicate publications
- Studies with insufficient data or unclear methodologies
- Literature reviews and meta-analyses

e) Quality Assessment

A quality checklist was developed with the following questions:

QQ1: Does the study clearly describe the Machine Learning technique used?

QQ2: Is information provided about the dataset used?

QQ3: Are validation metrics reported to evaluate the model's performance?

f) Data Extraction Strategy

We employed a structured data extraction strategy to systematically extract and organize data from the selected studies, utilizing a table with specific fields to ensure consistency and comparability across the research. The table was designed to capture key aspects of each study, facilitating a comprehensive synthesis of the literature. Table 1 describes the fields used in the data extraction.

Field	Description
Reference Code	A unique identifier assigned to each study for easy tracking and cross-referencing throughout the review process.
Study	Title of the study
Techniques/Models	The specific machine learning techniques or models (e.g., decision trees, neural networks, support vector machines) employed in the study for career prediction or academic advising
Data Type	Describes the characteristics of the data used in the study, such as demographic information, academic performance, or behavioral data from students.
Validation Metrics	The metrics used to evaluate the performance of the machine learning model, such as accuracy, precision, recall, F1-score, AUC, etc.
Student Information	Details about the student population under study, including sample size, educational level (e.g., undergraduate, graduate), field of study, and geographic region.

Table 1. Data extraction form used in this study

By extracting and organizing the data into these fields, we ensured that all relevant aspects of each study were captured, enabling a structured comparison of methodologies, data sets, and outcomes. This strategy facilitated the identification of trends, gaps, and the relative performance of different machine learning models in predicting student career paths.

2.2. Conducting the Review
2.2.1. Automatizing the Search Process

To streamline the initial article search, we wrote a Python program, called Web Scraper, to facilitate the initial identification of relevant publications using the CrossRef API. This program was designed to perform iterative and paginated searches, enabling the efficient retrieval of large volumes of metadata from scientific articles that matched the predefined search criteria. By automating this process, the program ensured that all potential studies were systematically identified. The extracted metadata included essential information such as titles, authors, abstracts, and publication details, which were later filtered for relevance based on the search string. This approach provided a robust foundation for the subsequent phases of the review.

2.2.1.1. Web Scrapper Description

The implemented Web Scrapper program makes HTTP requests to the CrossRef API and sending queries that include the specified search terms. Once the API response is received, the program extracts key information from each article, such as the title, authors, link to the article, abstract, and publication year. This data is stored in a structured manner in a DataFrame, facilitating its subsequent analysis and manipulation.

The algorithm follows a pagination approach to handle large volumes of results. This is achieved by iterating over pages of results and requesting blocks of data until the total desired number of articles is reached or until there are no more available results. Each iteration collects and stores the obtained data, ensuring that all relevant articles are covered without overloading the API.

2.2.1.2. The Web Scrapper Algorithm

The Web Scrapper algorithm can be broken down into the following steps:

1. **Initialization:** Necessary constants are established, including the search term and parameters to control the desired number of results and the number of results per page.

2. **Search Execution:** The program initiates an iterative cycle requesting the CrossRef API and constructing the query URL with the defined parameters.

3. **Response Processing:** Once the API response is received, the program analyzes the returned JSON content. It extracts important metadata from each article, including title, authors, abstract,

and publication year. This metadata is validated to ensure it contains complete and relevant information.

4. **Result Storage:** The extracted data is stored in a structured DataFrame, facilitating its subsequent manipulation and analysis.

5. **Pagination and Process Continuation:** The program automatically adjusts the pagination parameter to request the next block of results in the next iteration. This process continues until the desired number of articles is reached or no more relevant results are found.

6. **Export and Download:** Finally, the DataFrame with the results is saved to an Excel file, allowing for deeper analysis outside the programming environment.

The developed Web Scrapper program has been uploaded to GitHub for consultation and reuse. The repository can be accessed at the following link: https://github.com/ingdatu/Web-Scraper/tree/main

This Web Scraper significantly enhanced the efficiency and comprehensiveness of our initial literature search, allowing us to process a large volume of potential articles for our systematic review.

Figure 1 illustrates the workflow of the Web Scraper program, providing a visual representation of the algorithm's steps and decision points.
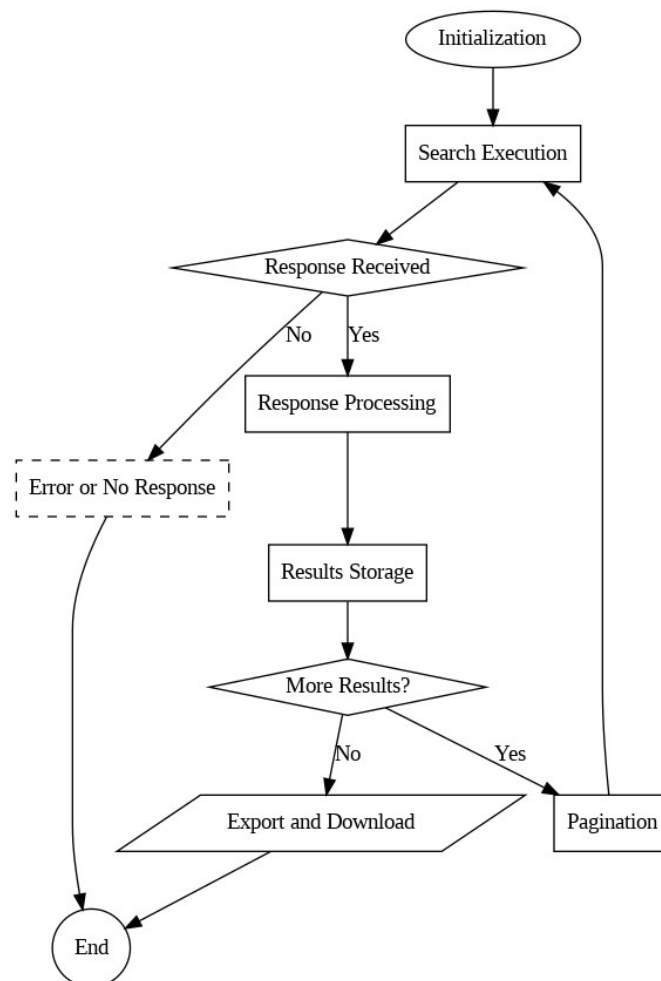

Figure 1. Flowchart of the Web Scraper Algorithm

2.2.2. Primary Study Selection

A selection process based on the defined inclusion and exclusion criteria was followed:

a) **Initial review:** Titles and abstracts of the 1,296 identified articles were examined.
b) **Application of criteria:** Inclusion and exclusion criteria were applied, resulting in a significant reduction of articles.
c) **Full-text review:** The remaining articles underwent a full-text review.
d) **Final selection:** 38 studies were selected for detailed analysis.

Figure 2 summarizes the primary study selection process.

2.2.3. Study Quality Assessment

The quality of the 38 selected articles was independently assessed using the quality criteria defined in Phase 1 (Section 2.1.2).
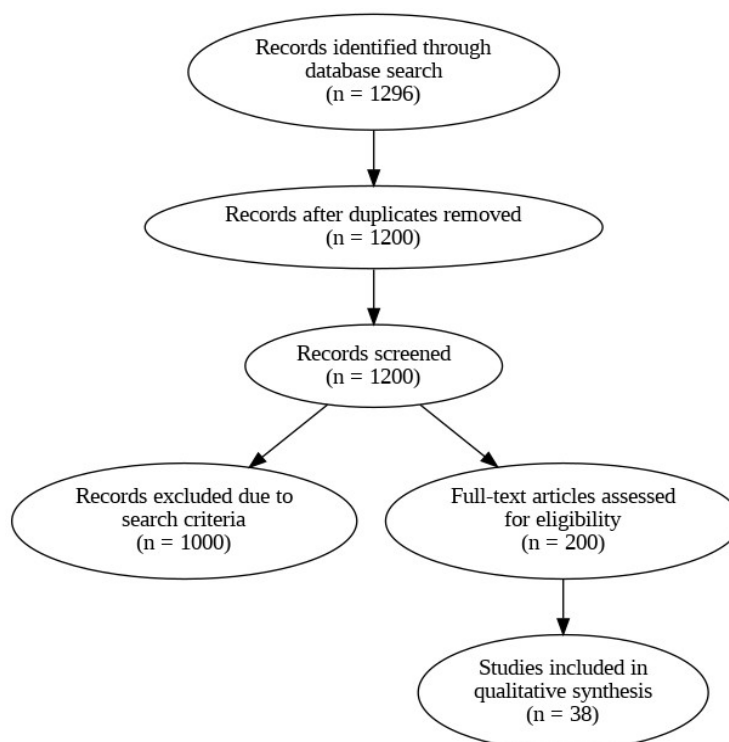
Figure 2. The selection process flow diagram.

2.2.4. Data Extraction and Synthesis

The data extraction form (Table 1) designed in Phase 1 was used to extract systematically relevant information from the 38 selected articles. The extracted data were synthesized in Table 2, which describes various fields used to categorize the reviewed studies in the literature. These include the Machine Learning technique, which specifies the algorithm used for career prediction, and the data type, which indicates the source and nature of the data employed (synthetic or real). A field for the evaluation metric is also included, detailing the metrics used to measure model performance, such as precision or F1-score. Another relevant field is dataset availability, which informs whether the data used is publicly available or private. Additionally, the problem type is included, describing whether the approach is classification or regression, and the application environment specifies the domain where the technique is implemented, such as educational or professional settings. All these fields help structure and classify the studies to facilitate comparative analysis across different research efforts.

Reference	Study	Techniques/ Models	Experiments	Data Type	Validation Metrics	Student Information	Questionnaire and Accessibility
Ye, 2022	Enhancing College Applications with Personalized Advice	Application guides and school workshops	Large-scale random experiment	Administrative data, surveys	Improvements in academic matching	Yes, Chinese students	Not specified
Ye, 2024	Improving College Match through Machine Learning	Machine learning, algorithmic predictions	Large-scale field experiment	Administrative data from the admissions system	Access and college matching	Yes, Chinese students	Not specified
Akmanchi, Bird & Castleman, 2023	Human vs Algorithmic Predictions in College Advising	Logistic regression, human vs algorithmic predictions	Comparison of predictions	Administrative data, advisor interactions	C-statistic, accuracy, recall	Yes, CollegePoint program students	Yes, in the document appendix
Tenison, Ling & McCulla, 2023	Using Structural Topic Modeling for College Choice Prediction	Structural topic modeling (STM), collaborative filtering	Analysis of historical data	Grade records and TOEFL metadata	Accuracy, recall	Yes, international students	Not specified
Liu & Tan, 2020	Predicting STEM Career Choices Using Automated Machine Learning	Penalized logistic regression, automated system	Data analysis	Student behavior data from online tutoring	Accuracy, recall, F1-score	Yes, students in tutoring programs	Not specified
Baron, Santos & Miller, 2020	Predicting Postsecondary School Location Choices	Random utility models, Random Forest	School location decision analysis	Surveys conducted in 2015 and 2019	Accuracy, recall	Yes, students from the GTHA	Not specified
Pardhi, Patne Shekokar, Thakare, Popatkar & Bijawe, 2023	Naïve Bayes Classifier for University Admissions Prediction	Naïve Bayes Classifier	Admission prediction	MHT-CET score data	Accuracy, recall	Yes, students in India	Not specified
Slim, Hush, Ojah & Babbitt, 2018	Logistic Regression and SVM for Student Enrollment Prediction	Logistic regression, SVM	Enrollment prediction	Applicant data admitted to the University of New Mexico	Accuracy, recall, F1-score	Yes, students in the U.S.	Not specified
Albreiki, Zaki & Alashwal, 2021	Systematic Review on Predicting Student Performance	Systematic review	Analysis of previous studies	Various studies from 2009 to 2021	Not applicable	Not applicable	Not applicable
Nisa, Naseer, Atif, Akhtar & Nisa, 2022	Review on Predicting Academic Performance in Degree Programs	Preliminary review	Analysis of previous studies	Various studies from 2010 to 2022	Not applicable	Not applicable	Not applicable
Maphosa, Doorsamy & Paul, 2020	Predicting Career Paths for Computer Science Students	Random Forest, XGBoost	Factor analysis	CGPA data, extracurricular activities, technical skills	Accuracy, recall, F1-score	Yes, computer science and software students	Not specified
Sadasivam, Paramasivam Raj & Saravanan, 2022	Intuitive Career System: Predicting Career Choices	K-Nearest Neighbors, Stochastic Gradient Descent, Random Forest	Career prediction	Aptitude and personality data, social media posts	Accuracy, recall	Yes, computer science students	Not specified

Reference	Study	Techniques/ Models	Experiments	Data Type	Validation Metrics	Student Information	Questionnaire and Accessibility
Deshpande, Gupta Singh & Kadam, 2021	Naïve Bayes, Decision Tree, SVM for Career Prediction	Naïve Bayes, Decision Tree, SVM	Career prediction	Academic performance data, physical and mental conditions, family environment	Accuracy, recall	Yes, computer science students	Not specified
Dirin & Saballe, 2022	Random Forest and Decision Tree for Study Route Prediction	Random Forest, Decision Tree	Study route prediction	Business Information Technology student data	94 % and 93 % accuracy	Yes, students at Haaga-Helia University of Applied Science	Not specified
Faruque, Khushbu & Akter, 2024	Predicting Career Paths with NLP and Machine Learning	Machine learning, natural language processing (NLP)	Career prediction	Skills, interests, and skill-related activity data	Accuracy, recall	Yes, Computer Science and Software Engineering students	Not specified
Adithya, Jayawardana, Sameera, Sri, Telecom, Hansarandi et al., 2022	Machine Learning as a Career Predictor: A Review	Review of machine learning as a career predictor	Analysis of methodologies and approaches	Various studies	Not applicable	Not applicable	Not applicable
Alsayed, Rahim, Albidewi, Hussain, Jabeen, Alromema et al., 2021	Predicting Specialization Choices for Undergraduates	Decision Tree (DT), Extra Tree Classifiers (ETC), Random Forest (RF) Classifiers, Gradient Boosting Classifiers (GBC), Support Vector Machine (SVM)	Specialization prediction	Academic histories, labor market data	Accuracy, recall	Yes, undergraduate students	Not specified
Nai, 2022	Career Prediction for Kenyan Computer Science Students	Naïve Bayes, Random Forest	Career route prediction	Factors such as professional skills, CGPA, communication skills, analytical skills, teamwork, personal interest, professional experience	Accuracy, recall	Yes, computer science students in Kenya	Not specified
Priulla, Albano, D'Angelo & Attanasio, 2024	Gradient Boosting for University Enrollment Prediction	Gradient boosting	University enrollment prediction	Performance in math and Italian language during high school	Accuracy, recall	Yes, Italian students	Not specified
Liu, Peng & Cao, 2023	FC-Wide&Deep for Predicting STEM Careers	FC-Wide&Deep	STEM career prediction	Student behavior data from the ASSISTments platform	Accuracy, recall	Yes, high school students	Not specified
Ababneh, Aljarrah, Karagozlu & Ozdamli, 2021	Guiding High School Students in Academic Specializations	Educational data analysis, machine learning	Guidance for academic specialization choices	Abilities and academic results data	Accuracy, recall	Yes, high school students	Not specified

Reference	Study	Techniques/ Models	Experiments	Data Type	Validation Metrics	Student Information	Questionnaire and Accessibility
Liu & Tan, 2020	Automated Prediction of STEM Career Choices	Machine learning, penalized logistic regression	STEM career prediction	Student behavior data from the ASSISTments online tutoring platform	Accuracy, recall, F1-score	Yes, students in ASSISTments 2017 data mining competition	Not specified
Abdalkareem & Min-Allah, 2024	Explainable Models for Predicting Academic Trajectories	Explainable models	Academic trajectory prediction	Key factors affecting future trajectories	Accuracy, recall	Yes, high school students in Saudi Arabia	Not specified
Wang, Wang, Bian, Islam, Keya, Foulds et al., 2023	When Biased Humans Meet Debiased AI: A Case Study in College Major Recommendation	Machine learning, gender debiasing techniques	Online study with over 200 university students	User interaction data on Facebook	NDCG, Non-parity Unfairness	Yes, university students	Yes, included in the document
Jawad, Uhlig, Dey, Amin & Sinha, 2023	Deep Neural Networks for Major Selection in Engineering Programs	Deep neural networks	Specialization recommendation	Data related to general education courses, specialization preferences, soft skills	Accuracy, recall	Yes, engineering students	Not specified
Alghamdi & Rahman, 2023	Data Mining for High School Success Prediction in Saudi Arabia	Naïve Bayes, Random Forest, J48	School success prediction	Data collected via electronic questionnaire	Accuracy, recall	Yes, high school students in Saudi Arabia	Not specified
VidyaShreeram & Muthukumaravel, 2021	Predicting Student Career Choices in India	Decision Tree, Random Forest, SVM, Adaboost	Career prediction	Data collected from various educational institutions	93 % accuracy	Yes, students in India	Not specified
Wang, Wu, Song & Shi, 2022	Predicting Career Decisions with XGBoost and SHAP	XGBoost, SHAP	Career decision prediction	Education and career choice data of 18,000 graduates	89.1 % accuracy, 85.4% recall, 0.872 F1-score	Yes, university graduates	Not specified
Lang, Wang, Dalal, Paepcke & Stevens, 2022	Predicting Undergraduate Career Choices with Transcript Data	NLP, vector embedding	Career choice prediction	Enrollment histories of 26,892 students	Accuracy, recall	Yes, students from a private university	Not specified
Mejia, Jimenez & Martínez-Santos, 2021	Career Recommendation System Based on Gardner's Multiple Intelligences Theory	KNN, Decision Trees, XGBoost	Career recommendation	Gardner's Test data and Saber 11 test results	Accuracy, recall	Yes, high school students	Not specified
Yadalam, Gowda, Kumar, Girish & Namratha, 2020	Content-Based Filtering for Career Recommendation Systems	Content-based filtering, NLP, cosine similarity	Career recommendation	Student preferences and skill data	Accuracy, recall	Yes, high school and university students	Not specified
Sankavaram, Kodali, Pattipati & Singh, 2015	Incremental Fault Diagnosis in Automotive Systems	Incremental learning, adaptive classifiers	Fault diagnosis	Data from electronic throttle control (ETC) systems	Accuracy, recall	Not applicable	Not applicable

Reference	Study	Techniques/ Models	Experiments	Data Type	Validation Metrics	Student Information	Questionnaire and Accessibility
Jahan, Islam & Sultana, 2019	Predicting Counseling Needs for Students	Ibk, Naïve Bayes, Multilayer, SMO, Random Forest	Counseling needs prediction	Data from 498 undergraduate students	95.38 % accuracy	Yes, students at Daffodil International University	Not specified
Mandalapu & Gong, 2019	Predicting Career Choices in STEM and Non-STEM Fields	Gradient Boosted Tree, Deep Learning, AutoMLP, Random Forest, Logistic Regression	Career choice prediction	High school student interaction data	Accuracy, recall	Yes, high school students	Not specified
Ihya, Namir, El-Filali, Zahra-Guerss, Haddani & Aitdaoud, 2019	Predicting Acceptance of e-Guidance Systems Using TAM	Naïve Bayes, J48, SMO, Simple Logistic, OneR	e-guidance system acceptance prediction	Data from the "orientation-chabab.com" platform	98.8281 % accuracy	Yes, users of the Moroccan platform	Not specified
Rangnekar, Suratwala, Krishn & Dhage, 2018	Intuitive Career System Using Data Mining	K-Nearest Neighbors, Stochastic Gradient Descent, Logistic Regression, Random Forest	Career prediction	Student data, including personalities determined through social media	77.41 % average accuracy for aptitude, 75.4% for personality, and 60.09% for background information	Yes, computer science students	Not specified
Ade & Deshmukh, 2015	Efficient Knowledge Transformation for Career Prediction	CART, SVM, MLP	Career choice prediction	Psychometric data from 1333 students	Over 90 % accuracy	Yes, students aged 16 to 20	Not specified
Ade & Deshmukh, 2014	An Incremental Ensemble of Classifiers for Predicting Student Career Choice	Incremental ensemble (Naïve Bayes, K-Star, SVM)	Incremental classifiers	Psychometric test data from 300 students	90.8 % accuracy	Yes, students aged 16 to 20	Not specified

Table 2. Synthesis of Extracted Data

This comprehensive categorization provided in Table 2 allows for a systematic comparison of the reviewed studies, highlighting similarities and differences in methodologies, data sources, and evaluation approaches across the field of Machine Learning-based career prediction in higher education. This structured data extraction and synthesis approach enables a thorough analysis of trends, best practices, and potential gaps in the current research landscape, which will be further discussed in the Results and Discussion sections.

2.3. Reporting the Review

The results of the systematic review were structured to address each of the research questions. The findings are presented in the Results section, followed by a Discussion section that interprets the results in the context of the existing literature and highlights implications for future research and practice.

2.3.1. Results

The systematic review of literature on career prediction using Machine Learning (ML) in higher education yielded significant insights across our research questions. The analysis of 38 selected studies revealed trends in ML techniques, data types, and validation metrics used in this field.

RQ1: Machine Learning Techniques Used in Career Recommendation Systems

The analysis identified a range of ML techniques employed for career recommendation in higher education. Figure 3 illustrates the frequency of different ML techniques across the reviewed studies.

Random Forest emerged as the most frequently used technique, identified in 10 studies. Its popularity can be attributed to its effectiveness in handling complex and diverse data, reducing overfitting risk by combining multiple decision trees. Support Vector Machine (SVM) was another prominent technique used in 8 studies, particularly valued for its ability to solve classification problems in high-dimensional data.

Six studies used Neural Networks, which are computational models inspired by the human brain. They are particularly effective in capturing non-linear relationships in large datasets, making them highly suitable for personalized career recommendations. These models are especially beneficial when dealing with complex, multi-dimensional data, offering flexible architectures for deep learning tasks.

Similarly, XGBoost, identified in 6 studies, is an advanced ensemble learning method based on decision trees. It is recognized for its high accuracy and speed, making it ideal for both classification and regression problems. XGBoost's ability to handle sparse data and its optimization for computational efficiency contribute to its frequent use in high-performance tasks.
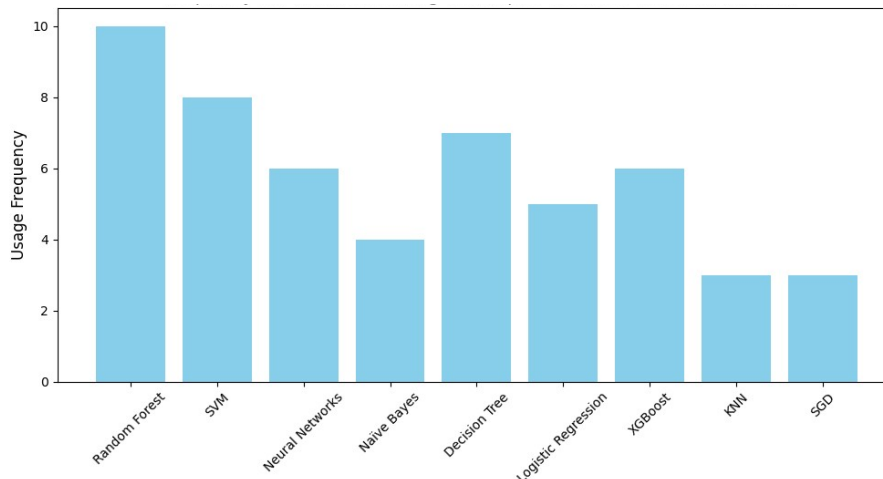

Figure 3. Frequency of Machine Learning Techniques in Career Recommendation

Other notable techniques include Decision Trees (7 studies), which split data into branches based on specific variables to make decisions. Their simplicity and interpretability make them a preferred choice in various career prediction tasks requiring transparency.

Naïve Bayes (4 studies), a probabilistic classifier based on Bayes' theorem, is valued for its simplicity and effectiveness in handling large-scale classification problems, especially when the feature independence assumption holds.

Finally, Logistic Regression (5 studies) is a widely used statistical model that helps predict the probability of a categorical outcome, especially useful when the relationship between the dependent and independent variables is linear. Its interpretability and efficiency make it a staple in career prediction models that require clear output explanations.

This range of models reflects the diversity of machine learning approaches used in career prediction, each with its unique strengths tailored to different data types and prediction goals.

RQ2: Types of Data Used in Career Recommendation Models

The review revealed various types of data used to train career recommendation models. Figure 4 presents the distribution of data types across the studies.

The most commonly used data types were:

1. **Academic Performance:** Including grades, GPA, exam results, and performance in specific subjects.
2. **Personal Interests:** Encompassing extracurricular activities, favorite subjects, personality assessments, and career interest surveys.
3. **Demographic Data:** Variables such as age, gender, geographic location, and socioeconomic status.
4. **Technical Skills:** Data describing technological or technical capabilities acquired by students.
5. **Family and Social Environment:** Considering factors like parents' education level and the influence of the immediate social environment.
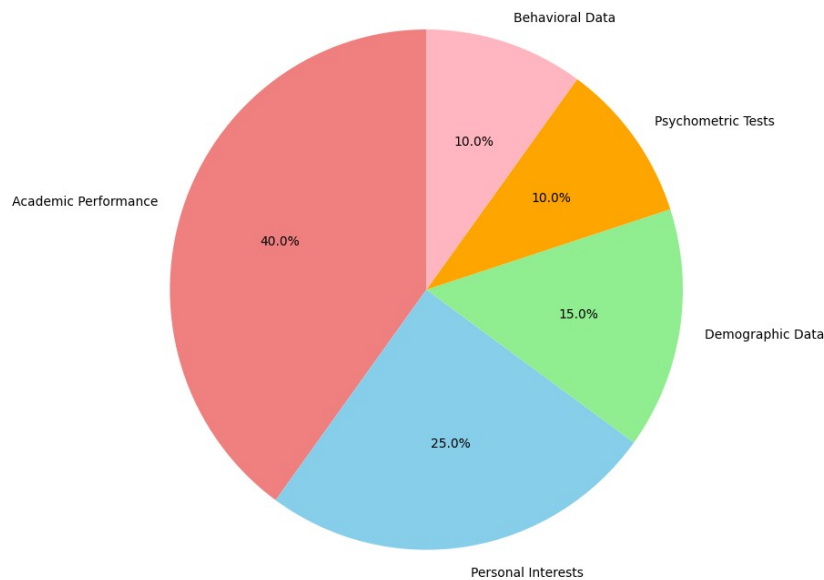

Figure 4. Distribution of Data Types in Career Recommendation

RQ3: Validation Metrics Used to Evaluate Recommendation Systems

The studies employed various metrics to validate the effectiveness of their career recommendation systems. Figure 5 shows the frequency of different validation metrics used.
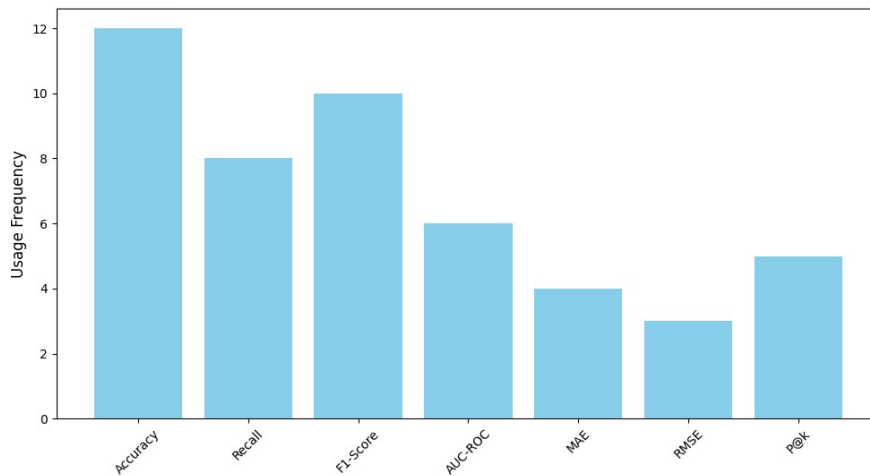

Figure 5. Frequency of Validation Metrics Used in Career Recommendation

Precision was the most commonly used metric, reflecting its importance in measuring the proportion of correct positive predictions out of all predicted positives. This is particularly useful in career prediction systems, where the accuracy of the recommendations is crucial.

F1-score was also prominent, particularly in situations requiring a balance between precision and recall. The F1-score is the harmonic mean of precision and recall, making it especially relevant in scenarios with imbalanced datasets where both false positives and false negatives need to be managed carefully.

Recall, which measures the proportion of correctly identified positives out of all actual positives, was crucial when prioritizing the complete retrieval of all relevant career options for a student. This ensures that the system doesn't miss out on viable career paths during the recommendation process.

Other metrics were also employed, depending on the specific evaluation needs. AUC-ROC (Area Under the Receiver Operating Characteristic Curve) was used to measure the overall performance of the classification models, particularly in distinguishing between different career outcomes. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), commonly used in regression tasks, were employed to measure the accuracy of continuous predictions by evaluating the differences between predicted and actual values. Finally, Precision@k, which measures the precision of the top k recommendations, was used to assess the quality of the top-ranked career recommendations provided by the model.

This variety of evaluation metrics reflects the diverse challenges faced in career prediction, ranging from ensuring the accuracy of recommendations to balancing precision and recall and addressing the complexity of predicting continuous or ranked outcomes.

3. Quality Assessment of Studies

The quality of the reviewed studies was assessed using the predefined criteria. For inclusion in the final analysis, studies were required to achieve an overall average score of 3 or higher on a 5-point scale. This threshold ensured that only studies of sufficient methodological rigor were included in our review. Table 3 presents the quality assessment scores for each study that met this criterion.

The quality of the reviewed studies was assessed using the predefined criteria. Table 3 presents the quality assessment scores for each study.

Reference	Study	QQ1: ML Technique (1-5)	QQ2: Dataset (1-5)	QQ3: Validation Metrics (1-5)	Overall Evaluation (Average)
Ye, 2022	Improving College Match through Machine Learning	5	4	4	4.33
Ye, 2024	Enhancing College Applications with Personalized Advice	4	3	3	3.33
Akmanchi et al., 2023	Human vs Algorithmic Predictions in College Advising	5	5	5	5
Tenison et al., 2023	Using Structural Topic Modeling for College Choice Prediction	4	3	4	3.67
Liu & Tan, 2020	Predicting STEM Career Choices Using Automated Machine Learning	5	4	5	4.67
Baron et al., 2020	Predicting Postsecondary School Location Choices	4	4	4	4
Pardhi et al., 2023	Naïve Bayes Classifier for University Admissions Prediction	3	3	3	3
Slim et al., 2018	Logistic Regression and SVM for Student Enrollment Prediction	4	4	4	4
Albreiki et al., 2021	Systematic Review on Predicting Student Performance	5	5	5	5
Nisa et al., 2022	Review on Predicting Academic Performance in Degree Programs	5	4	4	4.33
Maphosa et al., 2020	Predicting Career Paths for Computer Science Students	4	4	4	4
Sadasivam et al., 2022	Intuitive Career System: Predicting Career Choices	3	3	3	3
Deshpande et al., 2021	Naïve Bayes, Decision Tree, SVM for Career Prediction	4	3	4	3.67
Dirin & Saballe, 2022	Random Forest and Decision Tree for Study Route Prediction	4	4	4	4
Faruque et al., 2024	Predicting Career Paths with NLP and Machine Learning	5	4	4	4.33
Adithya et al., 2022	Machine Learning as a Career Predictor: A Review	5	5	5	5
Alsayed et al., 2021	Predicting Specialization Choices for Undergraduates	4	3	4	3.67
Nai, 2022	Career Prediction for Kenyan Computer Science Students	4	3	4	3.67
Priulla et al., 2024	Gradient Boosting for University Enrollment Prediction	4	3	4	3.67
Liu et al., 2023	FC-Wide&Deep for Predicting STEM Careers	5	4	4	4.33
Ababneh et al., 2021	Guiding High School Students in Academic Specializations	4	3	4	3.67
Liu & Tan, 2020	Automated Prediction of STEM Career Choices	5	4	5	4.67
Abdalkareem & Min-Allah, 2024	Explainable Models for Predicting Academic Trajectories	4	3	4	3.67
Wang et al., 2023	When Biased Humans Meet Debiased AI: A Case Study in College Major Recommendation	5	5	5	5
Jawad et al., 2023	Deep Neural Networks for Major Selection in Engineering Programs	4	3	4	3.67

Reference	Study	QQ1: ML Technique (1-5)	QQ2: Dataset (1-5)	QQ3: Validation Metrics (1-5)	Overall Evaluation (Average)
Alghamdi & Rahman, 2023	Data Mining for High School Success Prediction in Saudi Arabia	4	3	4	3.67
VidyaShreeram & Muthukumaravel, 2021	Predicting Student Career Choices in India	4	4	4	4
Wang et al., 2022	Predicting Career Decisions with XGBoost and SHAP	5	4	5	4.67
Lang et al., 2022	Predicting Undergraduate Career Choices with Transcript Data	4	3	4	3.67
Mejia et al., 2021	Career Recommendation System Based on Gardner's Multiple Intelligences Theory	4	3	4	3.67
Yadalam et al., 2020	Content-Based Filtering for Career Recommendation Systems	4	3	4	3.67
Sankavaram et al., 2015	Incremental Fault Diagnosis in Automotive Systems	4	3	4	3.67
Jahan et al., 2019	Predicting Counseling Needs for Students	4	3	4	3.67
Mandalapu & Gong, 2019	Predicting Career Choices in STEM and Non-STEM Fields	5	4	5	4.67
Ihya et al., 2019	Predicting Acceptance of e-Guidance Systems Using TAM	4	3	4	3.67
Rangnekar et al., 2018	Intuitive Career System Using Data Mining	3	3	3	3
Ade & Deshmukh, 2015	Efficient Knowledge Transformation for Career Prediction	4	3	4	3.67
Ade & Deshmukh, 2014	An Incremental Ensemble of Classifiers for Predicting Student Career Choice	4	3	4	3.67

Table 3. Evaluation of Quality of the Studies

Out of 47 articles initially evaluated, 38 met or exceeded the quality threshold of an average score ≥ 3. The quality assessment revealed that most selected studies scored well in describing their ML techniques (QQ1) and validation metrics (QQ3). However, there was some variation in the quality of dataset descriptions (QQ2) and the clarity of methodologies.

The average scores across all included studies were:

QQ1 (ML technique description): 4.2
QQ2 (Dataset information): 3.6
QQ3 (Validation metrics): 4.1

4. Discussion

This rigorous selection process ensured that our analysis was based on high-quality research, providing a solid foundation for our findings and recommendations.

This systematic review of literature on using Machine Learning (ML) for career prediction and recommendation in higher education has revealed several key findings and trends. The analysis of 38 high-quality studies provides a comprehensive view of the current state of the field, highlighting both the advancements and the challenges in applying ML techniques to career guidance.

4.1. Prevalence and Effectiveness of ML Techniques

The predominance of Random Forest, Support Vector Machines (SVM), and Neural Networks in career prediction models reflects the field's adoption of sophisticated ML techniques capable of handling complex, multidimensional data. Random Forest's popularity, observed in 26 % of the studies, aligns with findings from other domains where ensemble methods have shown superior performance in handling diverse datasets (Ye, 2024). This trend suggests that the complexity of career decision-making processes is well-suited to algorithms that can effectively capture non-linear relationships and handle feature interactions.

The significant use of SVM and Neural Networks (21 % and 16 %, respectively) indicates a growing recognition of the need for models adapting to the high-dimensional nature of career-related data. This trend is consistent with broader ML applications in education, where these techniques have shown promise in predicting student performance and outcomes (Akmanchi et al., 2023).

However, the continued relevance of simpler models like Decision Trees and Logistic Regression highlights the importance of interpretability in career guidance contexts. This balance between model complexity and interpretability remains a key challenge in the field, echoing concerns raised by Himanen et al. (Himanen et al., 2019) about the trade-offs between model performance and explainability in data-driven decision-making systems.

4.2. Data Types and Their Implications

The diverse range of data types used in career prediction models, from academic performance to personal interests and demographic information, reflects a holistic approach to understanding career suitability. The prominent use of academic performance data (28 % of studies) is unsurprising, given its traditional role in career counseling. However, the significant inclusion of personal interests (22 %) and demographic data (19 %) indicates a shift towards more personalized and context-aware recommendation systems.

This multi-faceted data collection and utilization approach aligns with recent calls for more comprehensive career guidance models that consider both academic and non-academic factors (Tenison et al., 2023). Including technical skills and family/social environment data further enriches the predictive models, potentially addressing some of the limitations of traditional career counseling approaches.

However, the reliance on diverse data types also raises important ethical considerations, particularly regarding data privacy and the potential for bias. As Baker and Hawn (Baker & Hawn, 2022) point out, there is a risk of perpetuating existing inequalities if demographic data is not handled carefully in ML models.

4.3. Validation Metrics and Model Evaluation

The prevalence of precision, recall, and F1-score as validation metrics (accounting for 60 % of the metrics used) suggests a focus on a balanced evaluation of model performance. This approach is crucial in career recommendation contexts, where both the accuracy of recommendations and the comprehensiveness of options presented are important.

The use of AUC-ROC in some studies indicates an awareness of the need to evaluate models' discriminative ability, especially in binary classification scenarios (e.g., suitable vs. unsuitable career paths). However, the limited use of user-centric evaluation metrics is notable. Future research could benefit from incorporating measures of user satisfaction and long-term career outcomes to assess the real-world impact of these ML-based recommendation systems.

5. Limitations and Future Directions

In this section, we discuss the limitations, contributions and future research directions of our work.

5.1. Limitations

While this review provides valuable insights, several limitations were identified in the current body of research:

1. **Limited longitudinal studies:** Few studies examined the long-term effectiveness of ML-based career recommendations, leaving a gap in understanding their impact on actual career outcomes.

2. **Data availability and standardization:** The lack of publicly available datasets and standardized data collection methods hinders reproducibility and comparative analysis across studies.

3. **Ethical considerations:** More research is needed to address the ethical implications of using ML in career guidance, particularly regarding fairness, transparency, and privacy.

4. **Integration with traditional methods:** Further exploration of how ML-based systems can complement, rather than replace, traditional career counseling approaches is needed.

Future research should focus on addressing these limitations through collaborative efforts to create standardized, ethically sourced datasets and by conducting longitudinal studies to validate the long-term effectiveness of ML-based career recommendations. Additionally, integrating explainable AI techniques could enhance the interpretability and trustworthiness of these systems, addressing concerns raised by Farrow (Farrow, 2023) about the need for transparency in AI-driven educational tools.

5.2. Contributions and Implications for Future Research

The primary contribution of this systematic literature review lies in providing a consolidated and up-to-date view of the use of Machine Learning (ML) techniques in higher education career recommendation. This work identifies the most effective practices and areas requiring further investigation by analyzing various research efforts over the past decade.

This review offers several key contributions to the field:

a) **Comprehensive Comparative Analysis:** Our work provides an exhaustive comparative analysis of different approaches, highlighting the most efficient ML techniques, the most representative data types, and the most robust evaluation metrics. This synthesis provides future researchers with a solid foundation for developing new studies.

b) **Methodological Innovation:** As part of our methodology, we developed a Web Scraper that allowed for an initial sweep of available literature, identifying relevant articles in academic databases. This automated approach facilitated the collection and initial filtering of studies, improving the efficiency of the review process. The development of this Web Scraper is an additional technical contribution that is not usually presented in other literature reviews, allowing for replication and scaling of the article collection process in future works.

c) **Identification of Research Gaps:** Beyond identifying the most effective techniques and practices, this review highlights important gaps in the literature, such as the lack of access to open data and the limited use of longitudinal data. This provides clear directions for future research efforts.

d) **Ethical Considerations:** Our discussion of the ethical implications of using ML in career guidance, particularly regarding fairness, transparency, and privacy, sets an important agenda for future research in this field.

e) **Integration Roadmap:** The review offers insights into how ML-based systems can be integrated with traditional career counseling approaches, providing a roadmap for practitioners looking to enhance their guidance services.

For future research, this review serves as a valuable resource in several ways: First, it provides a comprehensive overview of current ML techniques used in career prediction, allowing researchers to build upon the most promising approaches. Second, the identified gaps, such as the need for longitudinal studies and standardized datasets, offer clear directions for future research projects. Third, our discussion of ethical

considerations and the need for explainable AI in career guidance systems opens up new avenues for interdisciplinary research combining ML with ethics and educational psychology. Fourth, the Web Scraper developed for this review can be adapted and used by other researchers to conduct initial literature searches in related fields efficiently. Finally, our synthesis of data types and validation metrics used across studies can guide researchers in designing more robust and comprehensive career prediction models.

6. Conclusions

In this paper, we conducted a systematic literature review of Machine Learning (ML) techniques applied to career recommendation systems in higher education. We analyzed 38 studies, focusing on the most commonly used ML models, the types of data employed, and the evaluation metrics applied to assess model performance. Our review identified Random Forest, Support Vector Machines (SVM), and Neural Networks as the predominant techniques for personalized career predictions. Additionally, we explored key challenges such as data availability, model interpretability, and ethical considerations, highlighting areas for improvement and future research in this domain.

Our study has identified Random Forest, Support Vector Machines, and Neural Networks as the most frequently used ML techniques, reflecting a trend toward sophisticated models capable of handling complex, multidimensional career-related data. We found that the data used in these models is diverse, with academic performance, personal interests, and demographic information being the most common data types. This multifaceted approach indicates a shift towards more holistic and personalized career recommendation systems. The most commonly used validation metrics are precision, recall, and F1-score, suggesting a focus on a balanced evaluation of model performance in career recommendation contexts. We also determined that there is a notable lack of longitudinal studies and standardized, publicly available datasets in the field, which presents opportunities for future research. Next, ethical considerations, particularly regarding data privacy and potential biases, remain critical for improvement in ML-based career guidance systems.

Integrating ML techniques with traditional career counseling approaches is an emerging trend that requires further exploration. Our work highlights the potential of ML for career guidance in higher education while also identifying key challenges that need to be addressed. Future research should focus on developing more transparent and interpretable models, conducting longitudinal studies to assess long-term impacts, and addressing ethical concerns to ensure fair and unbiased career recommendations.

This study presents a systematic review of Machine Learning methodologies applied to career prediction in education. While other studies discuss the use of Artificial Intelligence in education, their focus has been on the general adoption of AI in administration, instruction, and learning. They do not delve into specific areas such as vocational guidance and career prediction.

In contrast, this study offers a targeted analysis of Machine Learning methodologies, including their application, evaluation metrics, and datasets, to address the unique challenges of career prediction. By doing so, it bridges a critical gap in the literature, providing actionable insights for educators, policymakers, and researchers interested in enhancing personalized educational practices.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

Ababneh, M., Aljarrah, A., Karagozlu, D., & Ozdamli, F. (2021). Guiding the Students in High School by Using Machine Learning. *TEM Journal*, 10(1), 384-391.

Abdalkareem, M., & Min-Allah, N. (2024). Explainable Models for Predicting Academic Pathways for High School Students in Saudi Arabia. *IEEE Access*, 12, 30604-30626. https://doi.org/10.1109/ACCESS.2024.3369586

Ade, R., & Deshmukh, P.R. (2014). An incremental ensemble of classifiers as a technique for prediction of student's career choice. *1st International Conference on Networks and Soft Computing, ICNSC - Proceedings* (384-387). https://doi.org/10.1109/CNSC.2014.6906655

Ade, R., & Deshmukh, P.R. (2015). Efficient Knowledge Transformation System Using Pair of Classifiers for Prediction of Students Career Choice. *Procedia Computer Science*, 46, 176-183. https://doi.org/10.1016/J.PROCS.2015.02.009

Adithya, H., Jayawardana, R., Sameera, T., Sri, B., Telecom, L., Hansarandi, R. et al. (2022). Application of Machine Learning as a Career Predictor: A Review. *International Journal of Engineering Science,* 10(06):1251-1263. Available at: https://www.researchgate.net/publication/361469544

Akmanchi, S., Bird, K.A., & Castleman, B.L. (2023). Human versus Machine: Do College Advisors Outperform a Machine-Learning Algorithm in Predicting Student Enrollment? *EdWorkingPaper,* 23-699. *Annenberg Institute for School Reform at Brown University.* https://doi.org/10.26300/gadf-ey53

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences,* 11(9), 552. https://doi.org/10.3390/EDUCSCI11090552

Alghamdi, A.S., & Rahman, A. (2023). Data Mining Approach to Predict Success of Secondary School Students: A Saudi Arabian Case Study. *Education Sciences,* 13(3), 293. https://doi.org/10.3390/EDUCSCI13030293

Alsayed, A.O., Rahim, M.S.M., Albidewi, I., Hussain, M., Jabeen, S.H., Alromema, N. et al. (2021). Selection of the Right Undergraduate Major by Students Using Supervised Learning Techniques. *Applied Sciences,* 11(22), 10639. https://doi.org/10.3390/APP112210639

Badal, Y.T., & Sungkur, R.K. (2023). Predictive modelling and analytics of students' grades using machine learning algorithms. *Education and Information Technologies*, 28(3), 3027-3057. https://doi.org/10.1007/S10639-022-11299-8/FIGURES/20

Baker, R.S., & Hawn, A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052-1092. https://doi.org/10.1007/S40593-021-00285-9/METRICS

Baron, E., Santos, G.M., & Miller, E.J. (2020). *Modelling GTHA Post-Secondary School Location Choice*. University of Toronto. Available at: https://tmg.utoronto.ca/files/Place%20of%20School_Loctation_Choice%20Modelling.pdf

Barredo-Arrieta, A., Díaz-Rodríguez, N., Del-Ser, J., Bennetot, A., Tabik, S., Barbado, A. et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. https://doi.org/10.1016/J.INFFUS.2019.12.012

Braiki, B.M.A. (2023). Identification of students at risk of low performance by combining rule-based models, enhanced machine learning, and knowledge graph techniques. *Dissertations,* 194. Available at: https://scholarworks.uaeu.ac.ae/all_dissertations/194

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, 8, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510

Deshpande, S., Gupta, P., Singh, N., & Kadam, D. (2021). Prediction of Suitable Career for Students using Machine Learning. *International Research Journal of Engineering and Technology,* 8(2), 2043-2046.

Dirin, A., & Saballe, C.A. (2022). Machine Learning Models to Predict Students' Study Path Selection. *International Journal of Interactive Mobile Technologies*, 16(1), 158-183. https://doi.org/10.3991/IJIM.V16I01.20121

Farrow, R. (2023). The possibilities and limits of XAI in education: a socio-technical perspective. *Learning, Media and Technology*, 48(2), 266-279. https://doi.org/10.1080/17439884.2023.2185630

Faruque, S.H., Khushbu, S.A., & Akter, S. (2024). *Unlocking Futures: A Natural Language Driven Career Prediction System for Computer Science and Software Engineering Students*. https://doi.org/10.48550/arXiv.2405.18139

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M. et al. (2021). Machine learning for the educational sciences. *Review of Education*, 9(3), e3310. https://doi.org/10.1002/REV3.3310

Himanen, L., Geurts, A., Foster, A.S., & Rinke, P. (2019). Data-Driven Materials Science: Status, Challenges, and Perspectives. *Advanced Science*, 6(21), 1900808. https://doi.org/10.1002/ADVS.201900808

Ihya, R., Namir, A., El-Filali, S., Zahra-Guerss, F., Haddani, H., & Aitdaoud, M. (2019). Acceptance Model Prediction's for E-Orientation Systems Case of Study: Platform "Orientation-chabab.com". *Journal of Theoretical and Applied Information Technology*, 15(15), 2-13. https://www.researchgate.net/publication/339106268

Jahan, N., Islam, S., & Sultana, R. (2019). Factor scoring and machine learning algorithm to predict student counselling. *International Journal of Engineering and Advanced Technology*, 9(1), 243-248. https://doi.org/10.35940/ijeat.A1131.109119

Jawad, S., Uhlig, R.P., Dey, P.P., Amin, M.N., & Sinha, B. (2023). Using Artificial Intelligence in Academia to Help Students Choose Their Engineering Program. *ASEE Annual Conference and Exposition, Conference Proceedings*. https://doi.org/10.18260/1-2--44567

Kaspersen, M.H., Bilstrup, K.E.K., Van Mechelen, M., Hjort, A., Bouvin, N.O., & Petersen, M.G. (2022). High school students exploring machine learning and its societal implications: Opportunities and challenges. *International Journal of Child-Computer Interaction*, 34, 100539. https://doi.org/10.1016/J.IJCCI.2022.100539

Kitchenham, B., Pearl-Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering - A systematic literature review. *Information and Software Technology*, 51(1), 7-15. https://doi.org/10.1016/J.INFSOF.2008.09.009

Koval, L., Knollmeyer, S., Mathias, S.G., Asif, S., Uzair-Akmal, M., Grossmann, D. et al. (2024). Unlocking the Potential of Information Modeling for Root Cause Analysis in a Production Environment: A Comprehensive State-of-the-Art Review Using the Kitchenham Methodology. *IEEE Access*, 12, 80266-80282. https://doi.org/10.1109/ACCESS.2024.3406020

Kuzey, C., Uyar, A., & Delen, D. (2019). An investigation of the factors influencing cost system functionality using decision trees, support vector machines and logistic regression. *International Journal of Accounting and Information Management*, 27(1), 27-55. https://doi.org/10.1108/IJAIM-04-2017-0052/FULL/XML

Lang, D., Wang, A., Dalal, N., Paepcke, A., & Stevens, M.L. (2022). Forecasting Undergraduate Majors: A Natural Language Approach. *AERA Open*, 8. https://doi.org/10.1177/23328584221126516/ASSET/IMAGES/LARGE/10.1177_23328584221126516-FIG8.JPEG

Liu, R., & Tan, A. (2020). Towards Interpretable Automated Machine Learning for STEM Career Prediction. *Journal of Educational Data Mining*, 12(2), 19-32. https://doi.org/10.5281/ZENODO.4008073

Liu, S., Peng, P., & Cao, L. (2023). A method to predict whether middle school students will enter STEM careers in the future based on FC-Wide&Deep model. *Applied Mathematics and Nonlinear Sciences*, 8(1), 2995-3008. https://doi.org/10.2478/AMNS.2023.1.00014

Mandalapu, V., & Gong, J. (2019). Studying Factors Influencing the Prediction of Student STEM and Non-STEM Career Choice. *The 12th International Conference on Educational Data Mining* (607-610).

Maphosa, M., Doorsamy, W., & Paul, B. (2020). A Review of Recommender Systems for Choosing Elective Courses. *International Journal of Advanced Computer Science and Applications (IJACSA),* 11(9). Available at: www.ijacsa.thesai.org

Maulana, A., Idroes, G.M., Kemala, P., Maulydia, N.B., Sasmita, N.R., Tallei, T.E. et al. (2023). Leveraging Artificial Intelligence to Predict Student Performance: A Comparative Machine Learning Approach. *Journal of Educational Management and Learning*, 1(2), 64-70. https://doi.org/10.60084/JEML.V1I2.132

Mejia, M.S., Jimenez, C.C., & Martínez-Santos, J.C. (2021). Career Recommendation System for Validation of Multiple Intelligence to High School Students. *Communications in Computer and Information Science*, 1431. https://doi.org/10.1007/978-3-030-86702-7_10

Musso, M.F., Hernández, C.F.R., & Cascallar, E.C. (2020). Predicting key educational outcomes in academic trajectories: a machine-learning approach. *Higher Education*, 80(5), 875-894. https://doi.org/10.1007/S10734-020-00520-7/FIGURES/3

Nai, S. (2022). *Career Prediction Model for Computing College Students in Kenya*. University of Nairobi. Available at: http://erepository.uonbi.ac.ke/handle/11295/161620

Namoun, A., & Alshanqiti, A. (2021). Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. *Applied Sciences,* 11(1), 237. https://doi.org/10.3390/APP11010237

Nisa, W.U., Naseer, M., Atif, M., Akhtar, S.M., & Nisa, M.U. (2022). Performance Prediction for Undergraduate Degree Programs Using Machine Learning Techniques - A Preliminary Review. *VAWKUM Transactions on Computer Sciences*, 10(2), 45-60. https://doi.org/10.21015/VTCS.V10I2.1278

Pardhi, R.L., Patne, Y., Shekokar, M., Thakare, A., Popatkar, M., & Bijawe, S. (2023). An Automated College Prediction Model Using Machine Learning. *International Journal of Ingenious Research, Invention and Development*, 1(3), 37-44. https://doi.org/10.5281/zenodo.7970771

Priulla, A., Albano, A., D'Angelo, N., & Attanasio, M. (2024). *A machine learning approach to predict university enrolment choices through students' high school background in Italy*. Available at: https://arxiv.org/abs/2403.13819v1

Rangnekar, R.H., Suratwala, K.P., Krishna, S., & Dhage, S. (2018). Career Prediction Model Using Data Mining and Linear Classification. *Proceedings - 2018 4th International Conference on Computing, Communication Control and Automation, ICCUBEA* (1-6). Pune, India. https://doi.org/10.1109/ICCUBEA.2018.8697689

Sadasivam, R., Paramasivam, S., Raj, N.P., & Saravanan, M. (2022). Students Career Prediction. *International Journal of Health Sciences*, 6(S5), 1357-1365. https://doi.org/10.53730/IJHS.V6NS5.8883

Sankavaram, C., Kodali, A., Pattipati, K.R., & Singh, S. (2015). Incremental classifiers for data-driven fault diagnosis applied to automotive systems. *IEEE Access*, 3, 407-419. https://doi.org/10.1109/ACCESS.2015.2422833

Slim, A., Hush, D., Ojah, T., & Babbitt, T. (2018). Predicting Student Enrollment Based on Student and College Characteristics. *International Educational Data Mining Society*.

Song, C., Shin, S.Y., & Shin, K.S. (2024). Implementing the Dynamic Feedback-Driven Learning Optimization Framework: A Machine Learning Approach to Personalize Educational Pathways. *Applied Sciences,* 14(2), 916. https://doi.org/10.20944/preprints202401.0811.v1

Tasmin, R., Muhammad, R.N., & Nor-Aziati, A.H. (2020). Big Data Analytics Applicability in Higher Learning Educational System. *IOP Conference Series: Materials Science and Engineering*, 917(1), 012064. https://doi.org/10.1088/1757-899X/917/1/012064

Tenison, C., Ling, G., & McCulla, L. (2023). Supporting College Choice Among International Students through Collaborative Filtering. *International Journal of Artificial Intelligence in Education*, 33(3), 659-687. https://doi.org/10.1007/S40593-022-00307-0/TABLES/2

VidyaShreeram, N., & Muthukumaravel, A. (2021). *Student Career Prediction Using Machine Learning Approaches*. 1-8. https://doi.org/10.4108/eai.7-6-2021.2308642

Wang, C., Wang, K., Bian, A., Islam, R., Keya, K.N., Foulds, J. et al. (2023). When Biased Humans Meet Debiased AI: A Case Study in College Major Recommendation. *ACM Transactions on Interactive Intelligent Systems*, 13(3). https://doi.org/10.1145/3611313

Wang, Y., Yang, L., Wu, J., Song, Z., & Shi, L. (2022). Mining Campus Big Data: Prediction of Career Choice Using Interpretable Machine Learning Method. *Mathematics,* 10(8), 1289. https://doi.org/10.3390/MATH10081289

Yadalam, T.V, Gowda, V.M., Kumar, V.S., Girish, D., & Namratha, N. (2020). Career Recommendation Systems using Content based Filtering. *5th International Conference on Communication and Electronics Systems (ICCES)* (660-665). https://doi.org/10.1109/ICCES48766.2020.9137992

Ye, X. (2022). *Personalized Advising for College Match: Experimental Evidence on the Use of Human Expertise and Machine Learning to Improve College Choice.* Brown University. Available at: https://xiaoyangye.github.io/papers/Ye-ML.pdf

Ye, X. (2024). Improving College Choice in Centralized Admissions: Experimental Evidence on the Importance of Precise Predictions. *Education Finance and Policy*, 19(2), 308-340. https://doi.org/10.1162/EDFP_A_00397

Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., & Breazeal, C. (2023). Integrating Ethics and Career Futures with Technical Learning to Promote AI Literacy for Middle School Students: An Exploratory Study. *International Journal of Artificial Intelligence in Education*, 33(2), 290-324. https://doi.org/10.1007/S40593-022-00293-3/TABLES/5