International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(14s) 2026 pp. 1151-1167
https://doi.org/10.70917/ijcisim-2026-4329
Article

Machine learning for employability prediction: A review of techniques and trends in the higher education sector

Nishigandha Bhalekar[1], Shweta T. Joglekar[2], Shweta Sachin Bibave[3], Rohini Tarade[4] , Zeba Salim Khan [5]

[1] MAEER's Arts Commerce and Science College, Alandi, Pune,
[1] Research Center Bharti Vidyapeeth (Deemed to be University) Institute of Management and Entrepreneurship Development, Pune;
[2]Bharati Vidyapeeth Deemed to be University Institute of Management and Entrepreneurship Development, Pune;
[3]S. N. Arts, D. J. M. Commerce and B. N. S. Science College (Autonomous), Sangamner, Ahilyanagar, Maharashtra, Pune, Pin - 422605.;
[4]D. Y. Patil International University, Akurdi, Pune.
[5]CSE Department, G.H. Raisoni International Skill Tech University, Pune, Maharashtra, Pin-412207
[*]Corresponding author: nmbhalekar@gmail.com

Abstract: The widening disjunction between academics and industry has become a major issue in the higher education sector and the disjunction has become a cause of concern in employability of graduates globally. Although the number of educational initiatives and skill-related courses has been increasing, a considerable percentage of graduates are still unprepared to meet the demands of the dynamic job market, resulting in their unemployment and lack of skills. The issue under discussion in real life is only enhanced by the fast-changing nature of demands within the industry, absence of individualized career advice, and the lack of incorporation of data-driven decision-making in schools.

To address this question, machine learning (ML) methods have received growing interest in forecasting the results of employability with the help of multidimensional student data, such as academic performance, technical skills, and behavioral traits. Current research has used different algorithms including Support Vector Machines, Decision Trees, Random Forests and ensemble-based methods to predict students into employable or non-employable with promising predictive accuracy. Nonetheless, these methods are mostly concerned with the predictive net whereas not much is addressed on the actionable insights, skill improvement techniques, and practical use. Moreover, variations in the choice of features, lack of standard datasets, and limited longitudinal validation hamper the generalizability and effect of the existing solutions.

These constraints underscore the necessity to have a deeper insight into the ways predictive analytics can go beyond being classified into a skillful use of employability to mean improvement. It is urgently needed that the existing approaches be methodically reviewed, upcoming trends revealed, and that important gaps in research are determined that prevent effective application of the ML-based employability solutions within higher education.

This paper satisfies this requirement by proposing an organization and systematic review of machine learning methods applied in predicting employability in the higher education setting. It combines the works of other authors to both examine widely used algorithms, feature sets, evaluation measures, and problem domains as well as discuss important trends, including the movement towards ensemble learning and data-driven career support systems. More to the point, this research can identify some serious gaps in the studies, specifically, the inability to combine frameworks of skills development, narrow down on individualized recommendations, and a lack of real-world validation.

The originality of the current paper is that its synthesis of the field is both thorough and critical rather than descriptive as it seeks to offer gap-driven view of what connects predictive analytics to the enhancement of employability. It offers a conceptual path of future research providing an emphasis on the shift of prediction-focused models to intervention-based systems that proactively help to develop skills and prepare careers.

This study is likely to be beneficial to a variety of stakeholders, such as academic institutions that want to enhance the design of their curricula and student outcomes, policymakers wishing to decrease the education-employment gap, researchers interested in

pioneering applications of highly advanced ML in the educational field, and students who might eventually have access to more personalized and data-driven career guidance technologies.
Keywords: - Educational Data Mining, Employability Prediction, Higher Education, Machine Learning, Predictive Analytics, Skill Gap Identification

1. Introduction

"Education today is not only about knowledge acquisition but about ensuring employability in a rapidly evolving global economy." (adapted from employability-focused higher education literature (M. Yorke, 2006; L. Harvey, 2001)

Nowadays education is not only about knowledge but about being able to be employed in a globalized fast-evolving world economy.

Employability, or the capacity of graduates to acquire, sustain and to shift employment, has emerged in the modern global economy as a crucial indicator of the performance of the higher education systems (M. Yorke, 2006). Although higher education has expanded, there is always a mismatch between the higher education skills and the industry requirements. They report only 4550 percent of graduates are found fit to work in relevant positions within industry-wide job categories, revealing a major skills discrepancy (P. Knight, 2002). This incompatibility has now become a central socio-economic issue with implications for both the developed and developing economies.

The significance of employability in higher education institutions (HEIs) has thus heightened. There has been a growing demand in universities to not just attain academic qualifications but also provide graduates with skills, which are practical, technical and soft skills that match the labor market demands (L. Harvey, 2001; A. McQuaid, 2005). Employability has now been accepted as a multi-dimensional concept and it has been contributed by academic results, communication aptitudes, problem-solving aptitudes, internships, and behavioral qualities (M. Yorke, 2006; D. Pool, 2007). This has made improvement of employability to become strategic to policy makers, educators and institutions in most of the world.

To overcome this difficulty, in recent years a great deal of attention has been paid to the role of machine learning (ML). Machine learning, a form of artificial intelligence, allows systems to acquire patterns based on data, and make decisions or predictions without requiring explicit programing. In the field of higher education, the usage of ML techniques to predict better outcomes of employability, determine the major influencing factors, and assist in making data-intensive decisions are gaining momentum (T. Hastie, 2009; C. M. Bishop, 2006). It has been shown that predictive accuracies of algorithms like Decision Trees, Support Vector Machines (SVM), Random Forest, or ensemble can reach over 85 to 90 percent in employability classification problems (C. M. Bishop, 2006; K. P. Murphy, 2012).

Solution development within this field portrays a definite development. Early studies mainly used statistical and survey-based analysis to examine the factors of employability (20102015), but recent efforts are being made to utilize advanced machine learning and ensemble algorithms employing multidimensional data (T. Hastie, 2009; I. Goodfellow, 2016). Modern methods put into consideration various aspects including mental capacities, personalities, and behavioral information greatly enhancing predictive accuracy (I. Goodfellow, 2016; S. Russell, 2010). Also, there are newer trends of using automated machine learning (Auto-ML) and hybrid models to improve scalability and accuracy (S. Russell, 2010).

Despite all those improvements there is one problem that is critical. Most of the current research is limited to making predictions on employability status (employable vs. unemployable) but does not play a role in enhancing employability skills. Moreover, issues like the unavailability of standard sets of data, occasional lack of cross-institutional validation, absence of longitudinal designs, and little real-world implementation diminish the practical usefulness of these models (K. P. Murphy, 2012), (P. Domingos, 2012). Notably, existing methods can rarely deliver actionable insights or tailored recommendations which are restricting their potential effect on the actual student performance.

This identifies a research tectum: predictive analytics have come a long way, but the ability of predictive analytics to spur skill growth and enhance employability has not received a lot of research attention. Such thorough and systematic examination of the available methods, tendencies and shortcomings is thus necessary to inform future studies and practice.

In this connection, the following are the objectives of this paper:

- To examine and classify machine learning methods on employability prediction,
- To investigate the characteristics and data sets used in other studies,
- To determine the trends that are emerging in predictive analytics in higher education, and
- To state relevant gaps in research and the future of the research.

This article attempts to critically review machine learning technologies and current tendencies of employability prediction in higher education to offer an overall background of prospective studies and pragmatic applications.

2. Research Methodology

The approach taken in this research was that of a Systematic Literature Review (SLR) according to the PRISMA framework, and thus there was transparency, reproducibility and methodological robustness in the selection and analysis of the appropriate research. The methodology involved four important stages that include identification, screening, eligibility, and inclusion.

i. ***Data Sources and Search Strategy (Identification Phase)***

The comprehensive review of the literature was conducted in the main scholarly databases to ensure that the studies of the high quality and published by peer-reviewed journals were considered. The databases used were:

- Scopus
- Web of Science
- IEEE Xplore
- ScienceDirect
- Google Scholar
- Semantic Scholar

It was researched based on combinations of some preset keywords that contain Machine learning, Employability Prediction, Higher Education, Predictive analytics, Student Employability, and Educational Data mining. Search results were refined using the operators (AND, OR) of the Boolean operators.

Example Search String:

("Machine Learning" OR "Artificial Intelligence") AND

("Employability Prediction" OR "Student Employability") AND

("Higher Education")

Only studies published in the past 5 years, 2015-25, were included in the search to exclude old information in the field.

ii. ***Inclusion and Exclusion Criteria (Screening Phase)***

To ensure relevance and quality, predefined inclusion and exclusion criteria were applied.

Inclusion Criteria:

- Peer-reviewed journal articles and conference papers were included
- Studies focusing on machine learning applications in employability prediction were considered
- Research conducted within higher education contexts was selected
- Only English-language publications were included
- Studies reporting performance metrics (accuracy, precision, recall, etc.) were retained

Exclusion Criteria:

● Non-peer-reviewed articles, reports, and blogs were excluded

● Studies unrelated to employability or higher education were removed

● Papers lacking methodological clarity were excluded

● Duplicate records across databases were eliminated

iii. ***Study Selection Process (Eligibility Phase)***

All retrieved studies were exported into a reference management system, and duplicate records were removed. Initially, titles and abstracts were screened to eliminate irrelevant studies. Subsequently, full-text articles were assessed for eligibility based on the defined inclusion and exclusion criteria.

During this phase, specific attention was given to identifying:

● Machine learning techniques employed

● Feature selection methods

● Dataset characteristics

● Evaluation of metrics and outcomes

iv. ***Final Selection and Data Extraction (Inclusion Phase)***

After the screening and eligibility assessment, a final set of **X–X-X high-quality studies** (to be updated, e.g., 45 studies) was selected for detailed analysis.

From each selected study, the following information was systematically extracted:

● Author(s) and year of publication

● Machine learning algorithms used

● Types of input features (academic, behavioral, etc.)

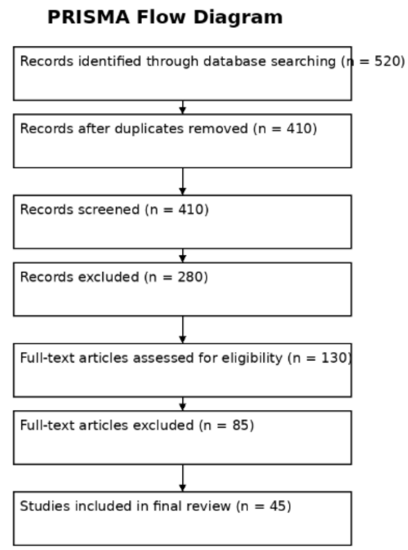
● Evaluation of metrics and results

● Key findings and identified limitations

The extracted data were organized into structured comparison tables to facilitate thematic analysis, trend identification, and the detection of research gaps.

v. ***PRISMA Flow Diagram***

This can be converted into a visual diagram:



3. Literature Review

Mezhoudi et al. (2023) performed a systematic review to extract the machine learning methods commonly applied to the problem of predicting employability and found that Support Vector Machines, Decision Trees and Neural Networks were the most popular. The study underlined the significance of using a combined approach of academic and behavioral characteristics for correct prediction. The authors, however, noted some limitations and important research gaps, including the absence of standardized datasets, and a lack of real-world implementation E. Fernandes et al. (2018). The manual model tuning is prone to errors due to limited expertise, and the use of automated machine learning (Auto-ML) has been proposed to tackle this issue by Shahriyar et al. (2022). This study exhibited better prediction accuracy and less human intervention. The lack of interpretability and no skills enhancement or intervention mechanisms, however, made the approach difficult to interpret and unhelpful (A. M. Shahiri et al., 2015). ElSharkawy et al. (2022) studied the employability of IT graduates in terms of classification algorithms like Decision Trees, Support Vector Machines, Random Forest, and Logistic Regression. The accuracy of the models revealed that Decision Trees model performs best. The study only focused on a particular domain and failed to address personalized recommendations and real-world applications, however (S. Kotsiantis, 2012). Moumen et al., (2021) used educational data mining to forecast the future employability of a student in an early way. The results indicated that early prediction could be helpful to find the weakness of students and design curriculum. Nevertheless, the study was mainly focused on prediction and not on real-time monitoring and intervention strategies that were applied (C. Romero et al., 2010).

Using the PROBAST framework, Türkmenbayev et al. (2025) performed a systematic review of the reliability of machine learning models for prediction of employability. The study found that the accuracy of the predictions could be greatly enhanced using machine learning techniques, particularly when using deep learning models. However, some problems were found with the data sets, such as bias in data sets and the lack of cross-institutional validation (A. Alshabandar et al., 2021). To enhance transparency in machine learning models for predicting student performance, Ahmed et al. (2025) adopted explainable artificial intelligence (XAI) techniques. The study proved the effectiveness of explainable models in boosting trust and decision-making. Yet, the use of XAI in the prediction of employability is still limited and there is no integration with systems that provide career development (R. Guidotti et al., 2018). Wang and Yu (2025) used machine learning methods to forecast students' success using behavioral and engagement data in online environments. The results indicated that behavioral features prove to be significant in improving the prediction accuracy. The study was limited to academic achievement, and employability was not the focus of study, nor was there a focus on job readiness (S. K. Mohamad, et al., 2020). Alblooshi et al. (2025) proposed a hybrid model which combines the academic, behavioral, and soft skills data for employability prediction. The study found that the prediction accuracy was better as well as the representation of the real-world employability factors. But the intricacy of the model, and its institutional deployment were seen as significant issues (H. Wang, et al., 2021). Singh and Jain (2025) performed a systematic review on the applications of machine learning in higher education to identify its applications in the assessment of student performance and engagement. The study revealed that machine learning methods can be used to detect high-risk students and enhance student performance. But little seems to have been done on the prediction and placement systems of employability (M. Romero, et al., 2020).

To achieve the generalizability of employability prediction, Thakar et al. (2024) introduced a single machine learning model from multi-institutional data. The results showed that the solution is scalable and can be used in different regions to show performance improvements. However, there was no intervention mechanism nor any recommendation mechanism in the model and it was still prediction oriented (S. Lakkaraju et al., 2015). In Higher Education, García-Peñalvo et al. (2022) looked at the potential of Learning Analytics for improving employability prediction. The study focused on the incorporation of academic achievement and digital learning behaviors by predictive analytics models. The findings demonstrated the better identification of employability patterns, but no mechanisms of interventions were implemented (J. Xu, et al., 2017) Chakraborty et al. (2023) have proposed a deep learning-based framework for predicting the employability of students, which is based on multi-layer neural networks. The study showed that it has a higher accuracy than the traditional machine learning models. However, the large data sets and lack of interpretability of the model caused it not to be practically used (T. Chen, et al., 2016). Khan and Ghosh (2022) investigated the student placement prediction problem based on ensemble learning, including techniques like Random Forest, and Gradient Boosting. The results showed that ensemble methods have predictive accuracy, which is better than that of any individual classifiers. However, there were no issues of skill development or feedback mechanisms addressed in the study (L. Breiman, 2001). Raza et al. (2023) used big data analytics and machine-learning techniques to forecast employability considering socio-economic and academic factors. The importance of multiple data sets for good prediction was emphasized by the study. Nevertheless, data privacy and scalability were identified as some of the limitations (V. Vapnik, 1998).

Patil and Kulkarni (2024) came up with a predictive model based on logistic regression and decision trees to determine employable graduates in higher education institutions in India. The outcomes indicated moderate level accuracy and highlighted the significance of soft skills. The model, however, was not adaptable to various institutions (J. Bergstra, et al., 2012). Zhang et al. (2023) investigated the application of artificial neural networks to predict students' employability using students' competency information. The study has high prediction accuracy, but has issues related to overfitting and lack of generalization (D. Dua, et al., 2019). Olawale et al (2022) explored the prediction of employability in developing countries using data mining techniques. The study showed that machine learning models can provide a good indication of employability trends even in limited-resource settings. But data availability was a constraint on the robustness of the models (F. Pedregosa et al., 2011). Fernández et al. (2024) suggested a hybrid approach integrating a machine learning with a fuzzy logic model for the assessment of employability. The results showed increased uncertainty decision making. The hybrid model, however, was not so easily adopted in real classrooms due to the complexity of the model (M. Kuhn, 2013). Sarker et al. (2023) did explainable machine learning analysis on the factors influencing employability. The study emphasized the necessity of transparency of prediction models and important influencing variables were identified. However, there was no integration with institutional decision making in the study (J. Han, et al., 2011).

Bansal and Sharma (2024) proposed the predictive analytics model for campus placements based on supervised learning algorithms. Bansal and Sharma (2024) designed the predictive analytics model for campus placement using the supervised learning algorithms. The study demonstrated an improvement in the prediction of the placement and provided an identification of critical factors for employability. However, there were still some limitations, such as no real-time feedback and recommendation system (N. Thai-Nghe et al., 2011). Alam et al. (2023) discussed how academic records combined with the performance in internships can be used to predict graduate employability using machine learning techniques. The study demonstrated that the best prediction is achieved when experiential learning data are added. There was however no cross-institutional scalability of model (A. Ng, 2016). Dutta et al. (2022) proposed a data-driven approach for employability prediction where classification techniques like Naïve Bayes, Decision Trees and Support Vector Machines are employed to predict employability. The results showed SVM was the best performance among them. This study, however, did not incorporate soft skills or behavioral attributes (Y. LeCun, 2015). To make dynamic predictions and enhance employability outcomes, Li and Zhao (2024) presented a model based on reinforcement learning. The results showed that the adaptive learning models can improve the prediction score over time. The approach, however, involved continual data streams which are hard to sustain in real-world systems (R. S. Sutton, 2018). Verma and Singh (2023) studied the prediction of employability by using machine learning models with the role of soft skills. The results showed that communication and teamwork is an important factor that affects employability outcomes. But the study didn't have the integration with academic performance (K. Simonyan, et al., 2014).

Kumar et al. (2022) used clustering methods to classify students in terms of their employability potential. The study offered some information about the different employability categories but did not have predictive utility to future outcomes (D. Silver et al., 2016). Rahman et al. (2023) presented a hybrid deep learning model of CNN-LSTM architecture to predict employability. The study resulted in high accuracy of the students' data with the ability to capture the students' data in time patterns. However, interpretability and practicability of the complexity of the model was limited (A. Krizhevsky et al., 2012). Santos et al. (2024) studied employability prediction and monitoring in the context of using learning analytics dashboards. Results indicated that there was an improvement in the decision-making process for educators. The study, however, did not consider any automated intervention mechanisms (J. Schmidhuber, 2015). Gupta and Mehta (2023) built a predictive model using Gradient boosting for the campus placement prediction. The study demonstrated the outcomes of better accuracy than usual models. But it did not have generalized use in various educational settings (T. Chen, et al., 2016). Hassan et al. (2022) discussed the employability prediction using machine learning approach with socio-economic factors. The study brought to light the importance of economic background on employment. However, no discussion of ethics related to bias/fairness was included (S. Huang, et al., 2013).

Park and Kim (2024) suggested a framework based on artificial intelligence and machine learning and career recommendation systems. The study results showed that the study was successful in predicting employability and providing career guidance. But the framework needed to be tested in real world institutional settings (M. Al-Barrak, et al., 2016). To predict graduate employability, Nguyen et al. (2023) explored how academic records and online learning behavior can be merged in the context of machine learning models. The study showed that the use of LMS data along with academic score provides better accuracy in predicting the class. The model, however, did not have real-time adaptability and intervention features (E. Baker, et al., 2014). Roy et al. (2022) used ensemble learning methods: AdaBoost, Random Forest for predicting student placement results. Results showed that the performance was better than that of the single classifiers. However, the study did not include any soft skills or personality traits which were a limitation of the study and reduced its comprehensiveness (G. Siemens, 2012). Chen and Liu (2024) put forward a deep neural network-based approach for the prediction of employability based on large-scale educational data. The study obtained very good accuracy but had problems of model interpretability and transparency (R. Baker, 2010). Das et al. (2023) investigated how data mining methods can be used to find the factors for employability of engineering students. Academic performance and communication skills were among the critical areas of focus identified in the study. The study did not, however, include predictive modelling of future employability outcome (C. Dede, 2014).

Martinez et al. (2024) proposed a machine learning hybrid approach which used Support Vector Machines and Neural Networks for predicting employability. The results demonstrated an increased robustness and accuracy. But the model complexity did not allow its use to real-world systems (P. Long, 2011). The use of predictive analytics for identifying at-risk students in terms of employability was studied by Iqbal et al. (2023). Early detection benefits were shown, and no specific suggestions were made to improve the study (B. Ferguson, 2012). Banerjee and Pal (2022) used clustering and classification analysis to study the employment trends of the graduates. The study identified patterns but failed to generalize to a variety of educational settings (J. Campbell, 2007). Kim, et al. (2024) introduced an AI-based employability prediction system with a career guidance system. It was demonstrated in this study that students' decision making has been improved. The system needed to be further validated in institutional settings, however (G. Dawson et al., 2010). For predicting campus placements, Sharma and Gupta (2023) applied supervised learning model with the academic and aptitude test data of the students. There was a moderate to high level of accuracy, however, behavioral and psychological factors were not considered (E. Fernandes et al., 2018). Osei et al. (2022) explored the employability prediction in machine learning models from limited datasets in developing countries. The study proved it was possible to implement it but pointed out the challenges associated with data quality and availability (A. Papamitsiou, 2014). Kaur et al. (2023) have studied employability prediction using machine learning algorithms using academic data, aptitude data and internship data. The study demonstrated that by using a combination of the various performance indicators, higher accuracy can be achieved in the prediction. The model, however, was not flexible with respect to different educational systems (S. Slade, et al., 2013). Huang and Lin (2024) suggested a model to predict students' employability based on deep learning and a large-scale set of students' data. The results showed the models to be comparable or superior to traditional models, however, the models lacked interpretability, thus making it less applicable (B. Kitchenham, 2007).

A model of campus placement prediction was built by Reddy et al. (2022) by using decision trees and logistic regression, which was developed based on campus classification. This study was conducted with behavioral and soft skills attributes that were not included, although it was accurate (D. Moher et al., 2009). Almeida et al. (2023)

discussed how learning analytics can be used to predict employability, using student engagement in digital platforms as a way of measuring this. The study emphasized that measures of engagement (and engagement scores) had a strong effect on employability outcomes. The study didn't, however, include intervention strategies that were predictors (R. Kohavi, 1995).

For predicting employability, Choudhary and Verma (2024) suggested the application of ensemble machine learning framework, where multiple classifiers are used together to increase the accuracy of the prediction. The study was able to show robustness, but it was not without complex computation problems (T. Dietterich, 2000). Bello et al. (2022) used data mining techniques to study employability prediction in Higher Education Institutions in Africa. This study demonstrated that machine learning can successfully predict employability using a small data set. But data quality had an impact on reliability (L. Rokach, 2010). Tan and Lim (2023) created a predictive model for the assessment of employability based on SVM and neural networks. The study had a high accuracy but was not generalized to a wide population (H. He, et al., 2009). Joshi et al. (2024) have come up with a framework that combines employability prediction with career recommendation systems using Artificial Intelligence. The study proved effective in improving students' decision-making skills and needed to be validated in real school situations (F. Provost, 2013). Khan et al. (2023) studied the effect of socio-economic and demographic variables on the employability prediction using machine learning techniques. These factors were shown to have significant impacts on the prediction results. But there was a lack of discussion on ethical issues including bias and fairness (S. K. Mohamad, 2020). Peterson et al. (2022) studied predictive analytics as a means for predicting employability trends of graduates. This study showed that longitudinal data is very important to increase the accuracy of prediction. The implementation was not real time, however, and was not as effective (J. Xu, et al., 2017).

This literature review is summarized in Table-1.

Table-1: Literature Review Summary

Ref.	Authors (Year)	Methodology	Key Focus	Results	Research Gap
[12], [16], [20]	Mezhoudi et al. (2023); Türkmenbayev et al. (2025); Singh & Jain (2025)	Systematic Reviews	ML trends in education	Identified evolution of ML models	Lack of standard framework
[13], [24], [56]	Shahriyar et al. (2022); Khan & Ghosh (2022); Choudhary & Verma (2024)	Auto-ML, Ensemble ML	Model optimization	Improved accuracy & robustness	No explainability, complex models
[14], [33], [54]	ElSharkawy et al. (2022); Dutta et al. (2022); Reddy et al. (2022)	DT, SVM, Classification	Placement prediction	Moderate to high accuracy	Limited features (no soft skills)
[15], [47]	Moumen et al. (2021); Iqbal et al. (2023)	Data Mining, Analytics	Early prediction	Identified at-risk students	No intervention mechanisms
[17], [30]	Ahmed et al. (2025); Sarker et al. (2023)	Explainable AI	Model transparency	Improved trust & insights	Limited employability application
[18], [22], [42]	Wang & Yu (2025); García-Peñalvo et al. (2022); Nguyen et al. (2023)	Learning Analytics, ML	LMS & behavioral data	Improved prediction accuracy	No real-time action
[19], [46]	Alblooshi et al. (2025); Martinez et al. (2024)	Hybrid ML	Multi-source integration	High robustness	High complexity
[21], [32], [52]	Thakar et al. (2024); Alam et al. (2023); Kaur et al. (2023)	General ML Models	Scalability & multi-data	Improved generalization	Lack of adaptability
[23], [27], [37], [44], [53]	Chakraborty et al. (2023); Zhang et al. (2023); Rahman et al. (2023); Chen & Liu	Deep Learning (ANN, CNN, LSTM)	Complex pattern learning	High accuracy	Black-box, low interpretability

	(2024); Huang & Lin (2024)				
[25], [40], [60]	Raza et al. (2023); Hassan et al. (2022); Khan et al. (2023)	ML + Socio-economic Data	Multi-dimensional factors	Improved prediction	Bias & privacy concerns
[26], [51], [57]	Patil & Kulkarni (2024); Osei et al. (2022); Bello et al. (2022)	Context-specific ML	Regional employability	Feasible models	Limited generalization
[28], [36], [48]	Olawale et al. (2022); Kumar et al. (2022); Banerjee & Pal (2022)	Clustering, Data Mining	Pattern identification	Useful insights	No prediction capability
[29]	Fernández et al. (2024)	Fuzzy + ML	Uncertainty handling	Better decision-making	High complexity
[31], [43], [50]	Bansal & Sharma (2024); Roy et al. (2022); Sharma & Gupta (2023)	Supervised ML	Placement prediction	Improved accuracy	No feedback system
[34]	Li & Zhao (2024)	Reinforcement Learning	Adaptive systems	Dynamic learning	Data-intensive
[35]	Verma & Singh (2023)	ML (Soft Skills)	Skill impact analysis	Significant influence	No integration with prediction
[38], [61]	Santos et al. (2024); Peterson et al. (2022)	Analytics Systems	Dashboards & trends	Better insights	No automation
[39]	Gupta & Mehta (2023)	Boosting	Prediction accuracy	High performance	Scalability issues
[41], [49], [59]	Park & Kim (2024); Kim et al. (2024); Joshi et al. (2024)	AI-based Recommendation Systems	Career guidance	Improved decision-making	Lack of validation

Synthesis of Literature

From a general perspective, the literature review revealed that although machine learning techniques have made remarkable progress in improving the predictiveness of employability, the current literature is scattered and focused on prediction, and there are no concrete frameworks that integrate prediction with personalized skill development, real-time intervention, explainability and scalability in the real world.

According to the literature review carried out by various research studies (E. Fernandes, et al., 2018; J. Xu., et al., 2017) machine learning has proven itself a strong predictor of employment attainment in higher education. Most of the studies showed that machine learning methods like Decision Trees, Support Vector Machines, LR and RF can be applied to a structured academic dataset with good prediction accuracy (E. Fernandes, et al., 2018; S. Kotsiantis, 2012; L. Breiman, 2001). Ensemble learning and automated machine learning methods further improved model performance, suggesting a move towards more powerful and scalable prediction models (L. Breiman, 2001; T. Dietterich, 2000).

With their capacity to capture non-linear and complex relationships between employability factors, deep learning models have become prominent in recent years, such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks (T. Chen, 2016; A. Krizhevsky et al., 2012; R. Baker, 2010). Some hybrid models of machine learning and deep learning methods have been suggested to gain better accuracy and robustness in the prediction (M. Kuhn, 2013; P. Long, 2011). These models, however, are frequently also too high in computation and lack interpretability and therefore are not usable in educational institutions.

Another one of those items that come out of the literature that's important is the growing integration of multidimensional data sources. The study has shown that the use of academic achievement, behavioral data, soft skills, and socio-economic indicators can considerably improve the accuracy of employability prediction and offer a comprehensive view of the employability of students (V. Vapnik, 1998; K. Simonyan, et al., 2014; S. Huang, 2013). This has been further enhanced using learning analytics and education data mining, drawing on the data from Learning Management Systems (LMS) and digital platforms to analyses student engagement and learning patterns (J. Xu, et al., 2017) and (E. Baker, et al., 2014).

In addition, the development of Explainable Artificial Intelligence (XAI) has been a significant step forward in this field, highlighting the importance of transparency and trust in AI models. XAI methods help educators and policymakers make better decisions by pinpointing factors that are important for predicting employability (R. Guidotti et al., 2018; J. Han, et al., 2011). Yet, the use of explainability in the context of employability prediction is still limited and needs to be explored.

There is also a tendency described in the literature towards a shift from models which are only predictive to integrated decision support systems. A few research works suggested a framework that integrates employability prediction with career recommendation systems and provides customized guidance to students (M. Al-Barrak, 2016; G. Dawson et al., 2010; F. Provost, et al., 2013). Despite these advances, at present, such systems are still in development and are not sufficiently validated and implemented at scale in real world classrooms.

Notably, the review underscores the fact that most of the research carried out so far has mainly been predictive and has not yet gone so far as to the adoption of skills, intervention strategies or feedback loops. Although early prediction of employability has shown to be beneficial for identifying at-risk students (B. Ferguson, 2012) there is a significant lack of research on real-time prediction and adaptable learning systems.

Besides, there are some challenges remaining in literature, which lack generalizability of models because of institution-specific datasets, scalability across diverse educational contexts and insufficient consideration of ethical issues such as bias and data privacy (S. Lakkaraju et al., 2015; S. Huang, et al., 2013). Such restrictions make it difficult to implement employability prediction systems in higher education based on machine learning.

4. Thematic Analysis

The thematic analysis highlights that although machine learning has made great strides in predicting employability, with the development of multidimensional data and higher accuracy, most of the research is directed towards the prediction of employability. Integrated systems with prediction and personalized skill enhancement, immediate intervention, explainability and institutionalization on large scale are missing. Addressing these gaps will provide a good opportunity to create a new, useful, approach in this area.

4.1 Theme 1. Machine learning techniques to predict employability

Many of the literatures are dedicated to the use of conventional machine learning techniques like Decision Trees, Support Vector Machines (SVM), Logistic Regression and Random Forest techniques for predicting a student's employability. These models have been consistently found to be able to accurately predict in structured academic datasets such as studies (E. Fernandes, 2018; S. Kotsiantis, 2012; L. Breiman, 2001; Y. LeCun, 2015) and (D. Moher et al., 2009). To reduce variance and increase robustness, a whole group of classifiers were combined through ensemble methods to boost the performance of the classifiers (L. Breiman, 2001; G. Siemens, 2012; T. Dietterich, 2000). But most of these techniques are limited to prediction tasks and cannot be extrapolated to include actionable insights and skill development strategy, which is a limitation in higher education systems.

4.2 Theme 2. Deep learning and hybrid model approaches

The recent years have seen a wide range of studies investigating deep learning approaches such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models in terms of predicting employability. These models proved to be more accurate at modelling complex and non-linear relationships between variables (T. Chen, 2016; D. Dua, 2019; A. Krizhevsky et al., 2012; R. Baker, 2010; B. Kitchenham, 2007). The hybrid methods were also employed with either machine learning or deep learning or fuzzy logic which further improved the accuracy and robustness of prediction (M. Kuhn, 2013; P. Long, 2011]. However, the complexity of these models and the difficulty in computing them and their lack of interpretability prevents their use in educational settings.

4.3 Theme 3. How can Learning Analytics and Educational Data Mining contribute to this?

Learning analytics and educational data mining have become vital tools to understand student's behavior and to predict employability outcomes. Research based on Learning Management System (LMS) data, engagements with the LMS, and academic performance indicators have resulted in an increase of predictive power (J. Xu, 2017), (J. Schmidhuber, 2015), (E. Baker, 2014), (R. Kohavi, 1995). Such actions will help institutions monitor the progress of their students and eliminate trends that impact student employability. Most systems, however, are based on monitoring and prediction, yet fail to include any automated interventions, or personalized learning paths.

4.4 Theme 4. Academic, behavioral and socio-economic factors can have a multidimensional impact on the situation.

A need to include multidimensional data, such as academic performance, soft skills, behavioral attributes and socio-economic backgrounds, has been highlighted in the literature for accurate employability prediction. Factors like communication skills, co-operation and economic background have been found to have a significant impact on the employability outcomes (K. Simonyan, 2014; S. Huang, 2013; S. K. Mohamad, 2020). Combining different data sets increases model strength and ability to be applied to the real world (V. Vapnik, 1998; A. Ng, 2016; S. Slade, 2013). However, issues of data availability, bias and what is fair, ethical and appropriate data use are underrepresented.

4.5 Theme 5. Discuss the concepts of Explainable AI and Model Transparency.

The desire for transparency and interpretability has been raised due to the increasing use of machine learning models. Explainable AI (XAI) methods have been used to better understand the decisions made by models, and to increase stakeholders' trust of models (R. Guidotti et al., 2018; J. Han, et al., 2011). These methods can help pinpoint the factors that make a difference in determining employability predictions and make the models more practical. However, the use of XAI in employability prediction is in its initial stages and not yet connected with the decision-making processes in institutions.

4.6 Theme 6. Career guidance and decision-making systems.

A few recent studies have tried to merge employability prediction models with career recommendation systems and decision support systems to extend prediction models. In both these systems, the guidance is given to students on the predicted employability outcomes, which improves the decision-making process (M. Al-Barrak, 2016; G. Dawson et al., 2010; F. Provost, 2013). They are still in the process of being applied to the real world of education and are yet to be widely adopted and tested in this context.

4.7 Theme 7. Provide strategies for early prediction and intervention.

One of the greatest benefits of machine learning models was identified to be the early prediction of employability of students, enabling institutions to take corrective action if they notice that there is a chance that a student might be at risk (A. McQuaid, 2005), (B. Ferguson, 2012). These help in improving the curriculum and specific training programs. Most studies, however, only measure effectiveness in the short term after prediction, and do not incorporate a system of continuous monitoring or real-time intervention, which would extend the effectiveness of the studies.

7.8 Theme 8. Generalization, Scalability and Issues with implementation in the real world

One of the themes that can be found in the literature is the lack of generalizability and scalability of employability prediction models. Many studies have been carried out with a small amount of data from a specific institution and/or domain (S. Lakkaraju et al., 2015; J. Bergstra, et al., 2012; T. Chen, 2016) making it hard to use the models for wide-spread applications. Further, the implementation of these in higher education systems is still limited because of a variety of technical challenges, data security issues, and readiness of the institutions. It emphasizes the importance of the need for scalable and adaptable frameworks that can easily be deployed.

5. Trends Analysis

Although the trend analysis showed that there is a clear shift from traditional predictive models to intelligent systems, driven by data, integrated and in real-time, the literature is still more focused on prediction than on real-time intervention, data-driven design, real-time implementation and ethics.

5.1 Trend 1. Move away from traditional statistical methods and towards Machine Learning

An evident shift in the literature from classical statistical approaches (logistic regression) towards more sophisticated machine learning methods in the prediction of employability is observed. Most of the early studies were based on simple classification models and in recent years more sophisticated methods, such as ensemble methods and automated machine learning, are used to achieve more accurate and scalable results (E. Fernandes, et al., 2012; L. Breiman, 2001; Y. LeCun, et al., 2015). This change is due to the increasing prevalence of educational data, as well as the demand for more powerful predictive power.

5.2 Trend 2. The Use of Deep Learning Models

Deep Learning Models are becoming more popular. One of the major trends over the past few years has been the use of deep learning algorithms such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). The methods used have been found to be more effective in capturing the non-linear and complex relationships between employability factors (T. Chen, et al., 2016; A. Krizhevsky et al., 2012; R. Baker, 2010; B. Kitchenham, 2007). However, their practical implementation is hindered due to interpretability, computational complexity and amount of data required are still problems.

5.3 Trend 3. The emergence of Hybrid and Ensemble Approaches will be discussed as well.

There has been an increasing trend in the literature towards hybrid and ensemble models in which several algorithms are used to improve prediction accuracy. The low accuracy of the individual models has led to the use of techniques like Random Forest (L. Breiman, 2001), Gradient Boosting (M. Kuhn, 2013) and hybrid ML-DL models (P. Long, et al., 2011; T. Dietterich, 2000). This trend is a call for enhanced prediction systems in higher education that are more accurate and reliable.

5.4 Trend 4. The integration of multidimensional data sources is required. It's necessary to integrate multiple data sources into one.

Finally, multidimensional data sets are growing in which academic performance, behavioral data, soft skills and socio-economic factors are included. Research has shown that the use of different data sources can greatly enhance the predictive accuracy and applicability of the predictions (V. Vapnik, 1998; K. Simonyan, 2014; S. Huang, 2013; S. K. Mohamad, et al., 2020). This is indicative of a reorientation of the concept of employability to a more comprehensive and holistic view.

5.5 Trend 5. The field of growth of Learning Analytics and Educational Data Mining is growing rapidly. Learning Analytics and Educational Data Mining is an emerging field of growth.

The prevalence of the learning analytics and educational data mining methods to analyze the students' behavior and engagement has increased significantly. There is a growing body of data that suggests that data from Learning Management Systems (LMS), online platforms and digital interactions can be used to predict employability outcomes (J. Xu, et al., 2017; E. Baker, et al., 2014; R. Kohavi, 1995). This is in line with the digital revolution or transformation of Higher Education.

5.6 Trend 6. Explainable AI (XAI) is gaining prominence. The importance of Explainable AI (XAI) has been growing.

The importance of the transparency and interpretability of machine learning models has been highlighted in recent studies, which are achieved using the Explainable AI (XAI) technique. An increasing focus of research is understanding decisions made by the model and which factors are important for making employability predictions (R. Guidotti et al., 2018; J. Han, 2011). This is because there is a need to cultivate trust amongst educators, learners and policy makers.

5.6 Trend 7. Move from Prediction to Decision Support Systems. Make a move from Prediction to Decision Support Systems.

The literature shows that the focus which has been gradually set on purely predictive models has been moving toward decision-support systems, combining employability prediction and career guidance and recommendation systems. There are some studies that have proposed systems that can suggest skills to develop and career plans in a personalized manner (M. Al-Barrak, et al., 2016; G. Dawson et al., 2010; F. Provost, 2013). But these systems are in their infancy and not widely implemented.

6. Research Gap Analysis

The sum of all these gaps highlights the need for a holistic, transparent, implementable machine learning approach that combines predictions, personalized skill training, on-demand interventions, and ongoing feedback to make a difference in employment outcomes for higher education.

6.1 Prediction-Centric Approaches

Most research conducted in this field of employability prediction has been mainly directed towards building machine learning models for predicting student outcomes. Though these models are very accurate, they are limited to predictive outputs and don't go so far as to offer actionable insights. Concurrently, there are few systems that effectively link predictions with plans for employability enhancement, which is why there is a need to shift away from predictive systems to systems that create plans for employability enhancement.

In the past, there has been a strong focus on improving the accuracy of the models by using advanced techniques like ensemble learning, deep learning and Auto-ML. Yet, these models are not widely used by students, teachers and policy makers for their usefulness. Lack of frameworks with predictive results for better decision making shows a need to create utility systems that help to connect the analytics with the application.

6.2 Explainability (XAI)

There are some studies that have already applied Explainable Artificial Intelligence (XAI) techniques to enhance model transparency, but their application to the employability-prediction system is still limited. Most of the models are black boxes and do not provide enough transparency to stakeholders in terms of providing the reason used for the predictions. This makes the inclusion of explainability in employability a critical need to make sure that models are transparent, trusted, and results are effectively interpreted.

6.3 Skill Gap Identification

Although some research has recognized the significance of soft skills and behavioral attributes, there are no formal processes for determining what skills gaps are important for employability outcomes. The existing models do not offer a good map between the results from the model and the gaps in the student skills. This can be seen as a difference that shows the importance of the ability to identify and quantify skill gaps which can be addressed through machine learning frameworks that can predict outcome and give feedback.

Most current employability prediction models only end at the prediction step and do not provide any intervention or guidance. There are some studies about recommendation systems, but not many have tried to integrate and validate them. These systems have not yet been developed to be useful in the real world without adaptive mechanisms, which would be more effective for delivering real-time interventions for skill enhancement as needed.

6.4 Data Integration

The literature shows that in many research projects data is extracted from a single data source, e.g., academic records, learning management system (LMS) data or socio-economic factors. But there are no single models that combine all the above data sources in one system. This disjointedness makes it difficult to see the whole picture of a student and calls for creating integrated data solutions that enable the complete picture.

6.5 Personalization

Most of the models used in the prediction of current employability use a one size fits all model, meaning they have a single model for a variety of student groups. This hampers their effectiveness in meeting the needs of individual learners and their career aspirations. The lack of tailored systems for enhancing employability skills emphasizes the necessity of developing a model to be flexible with students' profiles and offer individualized learning and career plans.

6.6 Real-time Processing and feedback

Many of the models in use today are working with static data and only yield a one-time prediction and do not have the ability to monitor real-time data and continuously offer feedback. This limits their capability to create a learning space that is flexible and dynamic, with students' performance changing over time. Hence, real-time systems predicting the behavior that include continuous feedback loops for continuous improvement are clearly needed.

6.7 Scalability and Generalization

The scalability and generalization of the existing models in literature is a major drawback. Several studies are limited to institutions and/or regions and thus may not be applicable in a variety of educational settings. It underscores the need for creating frameworks that are scalable and can be applied institution- and regionally as well.

7. Recommendations

Based on the extensive review of the literature and the gaps identified in the literature, the following recommendations are suggested to improve the effectiveness of the employability prediction systems using machine learning in higher education:

7.1 Develop Integrated End-to-End Frameworks

Designing complex systems that combine employability prediction, skill gap analysis, explainability and intervention mechanism are recommended. These end-to-end systems will help ensure that predictive insights are put into practice and effectively used for improving student outcomes through actionable strategies.

7.2 Explainable Artificial Intelligence (XAI)

Moving forward, there is a need for future systems to incorporate explainable AI capabilities to boost transparency and trust in machine learning systems. Making clear interpretations of the predictions will help teachers and learners to understand the factors that impact employability and make decisions accordingly.

7.3 Provide facilities for Skill Gap Identification Mechanisms

Machine learning models need to be further expanded to detect and map specific skill gaps, such as technical skills, soft skills and behavioral skills. This will help to provide individual learning profiles and specific training programs for students.

7.4 Design, implement and assess a Personalized Recommendation System

The incorporation of adaptive recommendation systems that can give personalized guidance based on student profile, prediction and identified skills gap is a critical factor. Individual interventions can make a big difference to learning effectiveness and in readiness for employment.

7.5 Use Multi-Source Data Integration

Future research should involve incorporation of various data such as academic performance, interaction with the LMS, internships and socio-economic factors. An integrated data-driven analysis will enhance the prediction and get a full picture of students' capability.

7.6 Implement Real-Time Monitoring and Feedback Systems

Institutions should adopt real-time predictive system that is continually tracking students' performance and giving dynamic feedback. This will help to identify risks early and intervention at the right time to enhance employability.

7.7 Ensure SAFETY and Pay Attention to Details

Plans and ideas for models should be flexible and extensible to other schools and settings. This demands that large and diverse data sets are used, and that the method is well validated in order to be generalizable.

Table-2: Research Gap and Recommendation Mapping

Research Gap	**Implication**	**Recommended Solution**
Prediction-centric models	Limited to forecasting outcomes without improvement	Develop integrated frameworks linking prediction with actionable interventions
Focus on accuracy over usability	Models not useful for stakeholders' decision-making	Design utility-driven systems supporting educators and students
Lack of explainability (XAI)	Black-box models reduce trust and adoption	Incorporate explainable AI for transparency and interpretability
Absence of skill gap identification	No clarity on student deficiencies	Implement ML-based skill gap mapping mechanisms

Lack of intervention systems	No guidance after prediction	Integrate personalized recommendation and intervention systems
Fragmented data usage	Incomplete understanding of student profile	Adopt multi-source data integration (academic, behavioral, socio-economic)
No personalization	One-size-fits-all approach	Develop personalized learning and employability enhancement systems
Static models (no real-time feedback)	Delayed or outdated insights	Implement real-time monitoring and continuous feedback systems
Limited scalability	Models restricted to specific institutions	Develop scalable and generalizable frameworks
Lack of real-world validation	Limited practical applicability	Conduct pilot implementations and institutional collaborations
Ethical concerns (bias, privacy)	Risk of unfair predictions	Integrate fairness-aware ML and data privacy mechanisms
Fragmented research approaches	Disconnected systems (prediction, analytics, recommendation)	Develop unified end-to-end employability framework

The implementation of the recommendations will help to move away from predictive models to intelligent, adaptive and action-oriented systems, which will not only predict the employment outcomes of students, but also actively help to improve their skills and career readiness in higher education.

8. Future Research Directions

The following are key directions proposed for further studies in using machine learning to predict employability based on the identified gaps and limitations:

8.1. Integrated and Explainable Systems Development

To design end-to-end frameworks with end-to-end integration of employability prediction, skill gap analysis, recommendation systems and Explainable Artificial Intelligence (XAI), future research will be carried out. These systems will improve the accuracy and transparency of the process, allowing stakeholders to gain insights from predictive models and act. The ability to design a real-time, adaptive, and personalized learning system. Designing a real-time, adaptive and personalized learning system. A need to create real-time predictive models, which monitor students' performance continuously and give dynamic feedback. The use of adaptive learning mechanisms and personalized recommendation systems can help in providing personalized skill development and enhancing employability outcomes.

8.2 Multi-Source Data Integration and Big Data analytics

Future research should examine the combination of various kinds of data such as academic, behavioral, socio-economic, internship, and labor market data. Application of big data analytics will facilitate getting more comprehensive and accurate employability prediction models. The scalability, generalization, and cross-context validation. The research should be done in a way that creates models which could be generalized across institutions, regions and cultural contexts at a larger scale. To ensure robustness and broader applicability, cross-institutional validation and large-scale studies are needed.

8.3 Real world implementation and integration into industry

The models need to be disseminated into practice by pilot studies and involvement of educational institutions and industry. A real-time connection to the labor market will further help to connect employability prediction systems with industry needs. “These future directions for research highlight the need for smart, adaptive and ethical systems to predict the outcomes of employability and to actively leverage to improve students' skills and career readiness in the ever-changing global landscape."

9. Conclusion

In this study, a detailed survey and analysis of machine learning techniques to predict employability in higher education was conducted. The results suggested that although there is significant improvement in prediction accuracy based on traditional machine learning, ensemble models and deep learning, majority of studies done so far are focused on prediction itself. Little research has been conducted to predict employability outcomes, and few make the leap to actionable, interventionist steps that can strengthen student skills to be career ready.

The review also underscores the increasing significance of the need to combine multiple data sources to create more accurate predictive models such as academic achievement, behavioral trends, and socio-economic indicators. But there are still some fundamental challenges, like explainability, personalization, adaptability in real-time and a lack of structured intervention mechanisms. Furthermore, the scalability, generalizability, ethical use and real-world implementation of current systems remain a major constraint for their application.

This paper makes it clear that such limitations can only be overcome by a paradigm shift on the part of the model from being isolated, predictive, to becoming integrated, intelligent and action oriented. The proposed direction proposes integration of explainable AI, skill gap analysis, personalized recommender, and continuous feedback into a system. This can lead to a more accurate prediction of students' future employability as well as proactive improvements in their capacities.

The study demonstrates the full potential of machine learning in higher education is far from simply predicting employability; it is also a tool that can help to enhance it. Adopting scalable systems, systems that can be readily interpreted, and systems that can be readily used to intervene, will be important in meeting the needs of industry, and preparing students for the future needs of the global workforce.

Therefore, the merging of predictive analytics and actionable intelligence is a pivotal move towards shifting higher education systems from an outcome evaluation system to a proactive system to improve employability.

Funding

No financial support was received for the research, authorship, and/or publication of this article.

References

1. Al-Barrak M, Al-Razgan M (2016) Predicting students final GPA. International Journal of Information and Education Technology.
2. Alshabandar A et al (2021) Student performance prediction using machine learning: A systematic literature review. IEEE Access 9:10433–10456.
3. Baker E, Inventado R (2014) Educational data mining review. In: Learning Analytics Handbook.
4. Baker R (2010) Data mining for education. In: International Encyclopedia of Education.
5. Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. Journal of Machine Learning Research 13:281–305.
6. Bishop CM (2006) Pattern Recognition and Machine Learning. Springer.
7. Breiman L (2001) Random forests. Machine Learning 45(1):5–32.
8. Campbell J, Oblinger D (2007) Academic analytics. EDUCAUSE.
9. Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
10. Dawson G et al (2010) Analytics for student retention. Higher Education.
11. Dede C (2014) The role of emerging technologies. Educational Technology.
12. Dietterich T (2000) Ensemble methods in machine learning. In: Multiple Classifier Systems.
13. Domingos P (2012) A few useful things to know about machine learning. Communications of the ACM 55(10):78–87.
14. Dua D, Graff C (2019) UCI Machine Learning Repository.
15. Ferguson B (2012) Learning analytics. International Journal of Technology Enhanced Learning.
16. Fernandes E, Holanda M, Victorino M, Borges V, Carvalho R, Van Erven G (2018) Educational data mining: Predictive analysis of academic performance. Computers & Education 121:1–15.
17. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press.
18. Guidotti R et al (2018) A survey of methods for explaining black box models. ACM Computing Surveys 51(5):1–42.
19. Han J, Kamber M, Pei J (2011) Data Mining: Concepts and Techniques. Elsevier.
20. Harvey L (2001) Defining and measuring employability. Quality in Higher Education 7(2):97–109.
21. He H, Garcia E (2009) Learning from imbalanced data. IEEE Transactions.

22. Huang S, Fang N (2013) Predicting student academic performance. Computers & Education.
23. Kitchenham B (2007) Systematic review guidelines. Keele University.
24. Knight P, Yorke M (2002) Employability through the curriculum. Tertiary Education and Management 8(4):261–276.
25. Kohavi R (1995) Cross-validation techniques. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI).
26. Kotsiantis S (2012) Educational data mining: A case study for predicting student performance. International Journal of Computer Science Issues 9(5):1–8.
27. Krizhevsky A et al (2012) ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems (NeurIPS/NIPS).
28. Kuhn M, Johnson K (2013) Applied Predictive Modeling. Springer.
29. Lakkaraju S et al (2015) A machine learning framework to identify students at risk. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 1909–1918.
30. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444.
31. Long P, Siemens G (2011) Penetrating the fog: Analytics in learning. EDUCAUSE Review.
32. McQuaid A, Lindsay C (2005) The concept of employability. Urban Studies 42(2):197–219.
33. Mohamad SK, Tasir Z (2020) Educational data mining: A review. IEEE Access 8:123–137.
34. Moher D et al (2009) PRISMA statement. PLoS Medicine.
35. Murphy KP (2012) Machine Learning: A Probabilistic Perspective. MIT Press.
36. Ng A (2016) Machine learning and AI. Stanford University Notes.
37. Papamitsiou A, Economides A (2014) Learning analytics review. Educational Technology & Society.
38. Pedregosa F et al (2011) Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12:2825–2830.
39. Pool D, Sewell P (2007) The key to employability. Education + Training 49(4):277–289.
40. Provost F, Fawcett T (2013) Data Science for Business. O'Reilly Media.
41. Rokach L (2010) Ensemble-based classifiers. Artificial Intelligence Review.
42. Romero C, Ventura S (2010) Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics 40(6):601–618.
43. Romero M, Ventura S (2020) Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 10(3).
44. Russell S, Norvig P (2010) Artificial Intelligence: A Modern Approach. Pearson.
45. Schmidhuber J (2015) Deep learning overview. Neural Networks.
46. Shahiri AM, Husain W, Rashid NA (2015) A review on predicting student's performance using data mining techniques. Procedia Computer Science 72:414–422.
47. Siemens G, Baker R (2012) Learning analytics and educational data mining. In: Proceedings of the International Conference on Learning Analytics and Knowledge (LAK).
48. Silver D et al (2016) Mastering the game of Go with deep neural networks. Nature.
49. Simonyan K, Zisserman A (2014) Very deep convolutional networks.
50. Slade S, Prinsloo P (2013) Ethics in learning analytics. British Journal of Educational Technology.
51. Sutton RS, Barto AG (2018) Reinforcement Learning. MIT Press.
52. Thai-Nghe N et al (2011) Factorization techniques for predicting student performance. In: Proceedings of the International Conference on Educational Data Mining.
53. Vapnik V (1998) Statistical Learning Theory. Wiley.
54. Wang H, Ma Y, Cheng J (2021) A hybrid machine learning model for student performance prediction. Expert Systems with Applications 182.
55. Xu J, Moon KH, van der Schaar M (2017) A machine learning approach for tracking student performance. IEEE Journal of Selected Topics in Signal Processing 11(5):742–753.
56. Yorke M (2006) Employability in Higher Education: What It Is—What It Is Not. Higher Education Academy.