Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 13 No. 2 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijep.v13i2.33643

PAPER

Applying Data Mining in Graduates' Employability: A Systematic Literature Review

Héritier Nsenge Mpia(✉), Lucy Waruguru Mburu, Simon Nyaga Mwendia

School of Technology, KCA University, Nairobi, Kenya

2004988@students.kcau.ac.ke

ABSTRACT

Envisaging an adequate IT/IS solution that can mitigate the employability problems is imperative because nowadays there is a high rate of unemployed graduates. Thus, the main goal of this systematic literature review (SLR) was to explore the application of data-mining techniques in modeling employability and see how those techniques have been applied and which factors/variables have been retained to be the most predictors or/and prescribers of employability. Data-mining techniques have shown the ability to serve as decision-support tools in predicting and even prescribing employability. The review determined and analyzed the machine-learning algorithms used in data mining to either predict or prescribe employability. This review used the PRISMA method to determine which studies from the existing literature to include as items for this SLR. Hence, we chose 20 relevant studies, 16 of which are predicting employability and 4 of which are prescribing employability. These studies were selected from reliable databases: ScienceDirect, Springer, Wiley, IEEE Xplore, and Taylor and Francis. According to the results of this study, various data-mining techniques can be used to predict and/or to prescribe employability. Furthermore, the variables/factors that predict and prescribe employability vary by country and the type of prediction or prescription conducted research. Nevertheless, all previous studies have relied more on skill as the main factor that predicts and/or prescribes employability in developed countries, and no studies have been conducted in unstable developing countries. Therefore, there is a need to conduct research on predicting or prescribing employability in such countries by trying to use contextual factors beyond skill as features.

KEYWORDS

data mining, employability, machine learning, predictive analysis, prescriptive analysis, graduate skills, lack of contextual factors

1 INTRODUCTION

The past research has applied different techniques to discover which factors influence employability or which data mining techniques can effectively predict employability. Studies focused on the latter approach have discovered hidden patterns that

Mpia, H.N., Mburu, L.W., Mwendia, S.N. (2023). Applying Data Mining in Graduates' Employability: A Systematic Literature Review. *International Journal of Engineering Pedagogy (iJEP)*, 13(2), pp. 86–108. https://doi.org/10.3991/ijep.v13i2.33643

Article submitted 2022-06-27. Resubmitted 2022-11-15. Final acceptance 2022-11-18. Final version published as submitted by the authors.

depict different levels of influence by extracting information from massive datasets. Nevertheless, these past studies have predominantly been focused on understanding the skills and competencies of graduates and have overlooked other key success and failure factors of employability [1].

Data mining, a major component of database knowledge discovery, is an analytical process that allows the discovery of new knowledge. The purpose of data mining is to apply machine-learning (ML) algorithms to discover intrinsic relationships among the explored data [2]. In recent years, data-mining techniques have outperformed other traditional investigation methods in many domains, such as business, engineering, and social sciences. In this paper, an investigation was conducted into the application of data mining in employability prediction in order to understand the effectiveness and applicability of different techniques, as well as to discover the factors that have been prioritized in previous studies for predicting employability of individuals after their academic training [3].

As generally defined by the majority of reviewed research articles, employability implies the consideration of employment prospects of an individual, and the individual's capacity to proactively face the demands of the labor markets. Graduate employability is of great concern for institutions offering higher education. Consequently, a lot of research has been conducted to understand the employability issues in different contexts through data-mining techniques [4]. However, the increasing need for accuracy of prediction and the generalizability of the prescriptive models makes it paramount for researchers to prioritize contextual attributes. Inclusion of such attributes can promote the understanding of individualized employment profiles of graduates. This will help researchers to apply developed models to complex backdrops, such as areas affected by war, marginalization, limited industrialization, and restriction due to pandemics [5–7].

The objectives of this systematic literature review (SLR) were:

1. To investigate the extent to which data-mining techniques have been applied to determine graduate employability.
2. To assess the factors affecting graduate employability.
3. To explore the existing predictive and prescriptive models of employability and the generalizability of these models to other contexts.
4. To make recommendations for future research into employability prediction and prescription in the Democratic Republic of Congo (DRC), an unstable developing country.

The research was guided by the following research questions in a bid to reach the above objectives:

RQ1. To what extent has data mining been used to predict graduate employability?
RQ2. Which are the factors identified in previous literature for predicting graduate employability?
RQ3. What is the applicability of existing models for predicting graduate employability in unstable socio-economic and political backdrops?

To answer these research questions, keyword searches into several journal databases was conducted to identify the scientific articles. Articles of interest were classified according to the machine-learning algorithms used, the type of input data used, the impact factor of the journal in which the article had been published and the attributes and/or factors used in the studies. Extensive review of the identified literature was then conducted to extract the required knowledge in summary form.

2 LITERATURE REVIEW

2.1 Employability

Employability refers to an acceptable set of skills and competencies that enable a new graduate to compete in the labor market and attain employment irrespective of any change in society or in need by employers. Employability includes, in addition to the skills in demand, a set of attributes and experiences that are amassed during learning in a higher-education institution [8]. Bridgstock [9] has categorized employability into two perspectives. The first is the conventional view that focuses on generic and domain-specific skills and the initial employment outcomes. The second is the broader contemporary view that focuses on a holistic approach that incorporates personal characteristics and places work into the context of an individual's life and the current demands of the job market [10].

Rich [11] defines employability as a set of attributes that increase the potential of a graduate being employed in a private or public company. The level of employability is thus based on the adequacy of knowledge accumulated in a learning institution with the needs of the labor market. Employability is not employment; it can be summarized by knowledge, skills, and social capital. Suleman [12] defined graduate employability as the individual skills, knowledge, and complex attributes that a person obtains from higher education that help him/her to secure and maintain a job. Recently, employability has been defined as an ever-changing process that requires continuous revision of academic education. This means that employability can be increased through development of new skills fostered through university-industry partnerships, the strengthening of quality assurance systems, and the alignment of university education with a country's development plans to provide a human capital that is endowed with employable intelligence, ability, and capacity [13].

Employment has become critical in recent times owing to the deteriorating global economy that resulted from the COVID-19 pandemic. The pandemic crisis has impacted on employment levels, the skills of graduates, and the stability of organizations offering employment. This has made employability less achievable and caused scholars to review certain attributes that are impacting job retention [14], [15]. Other factors affecting employability include internal conflicts in regions such as Nigeria, Somalia, and the DRC [16], where investors have shied from venturing their capital and jobs are difficult to find. As a result, in the DRC for example, about 70% of the youth are unemployed, and only about 100 out of every 9,000 graduates find employment every year [17]. Other socio-political factors such as tribalism, favoritism, and politicization of employment also affect employability in many regions of Africa [18].

2.2 Machine learning

ML is at the heart of scientific invention and artificial intelligence (AI). This branch of AI is experiencing unimaginable success. Researchers purport that the success of ML is due to the nature of the data-driven epistemological approach. This approach enables learning from a large number of similar situations [19]. Unlike older programming systems, a ML system is not explicitly programmed [20]. Instead, it is trained using numerous examples that are relevant to the given task [21]. During training, the system attempts to discover a statistical structure in the examples. As a result, the system develops rules to automate the task after processing a series of data, creating a generalizable model, and validating the obtained model [22].

ML has become indispensable in the day-to-day activities. It offers a variety of applications ranging from simple web applications to complex systems such as speech recognition, computer vision, natural language processing, and facial recognition [19]. ML algorithms have been summarized by Sha et al. [23] as belonging to three major categories: Supervised Learning, Unsupervised Learning, and Reinforcement Learning. Reinforcement Learning is a sequential framework for decision-making, and Deep Learning is an associated framework for representation learning. More recently, Reinforcement Learning has been coupled with Deep Learning to create Deep Reinforcement Learning (DRL). This represents a new area in AI that aims to solve complex tasks with sparse available knowledge from the environment through its ability to learn from the data at different levels of abstraction [24].

2.3 Data mining

Data mining is a set of techniques for exploring massive data. This approach uses various ML algorithms. Through unsupervised learning, for example, data mining can apply cluster analysis to differentiate objects that possess a certain set of features. Subsequently, it can divide these objects into categories with no need to train data or employ factor analysis [25]. Data mining uses the classification technique to observe the features of a new object under analysis and assign to it a specific class. Common analyses from data mining include decision trees, Naive Bayes or K-Nearest Neighbor, which highlight characteristics of data [25], or regression techniques, which identify intrinsic dependencies among variables [25].

Data-mining techniques can be generally categorized into Descriptive Mining Techniques and Predictive Mining Techniques. Descriptive techniques use unsupervised learning of the available data to generate patterns that humans can interpret to discover knowledge. Clustering, which automatically classifies objects according to common trends, is the most widely applied descriptive technique [26]. On the other hand, predictive techniques attempt to predict the behavior of future data using the current patterns. Often, supervised ML models are deployed as predictive techniques. Such techniques can, for instance, predict whether or not a graduate will belong to the category of employed individuals, given some features. In predictive data mining, classification is the most popularly applied data-mining technique [27].

3 RESEARCH METHODOLOGY

This research combined quantitative and qualitative methods to achieve the study objectives. Descriptive analytics were used to assess results from the past literature that has applied data mining in employability as well as to estimate the commonly used ML algorithms. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) technique was used to quantitatively determine which studies from the existing literature to include in the current systematic review [28]. Additionally, the Strengths, Weaknesses, Opportunities, and Threats (SWOT) qualitative analysis [29] was used to examine the strengths of previous research on the application of data mining in the field of employability, to detect the conceptual and knowledge gaps so as to identify opportunities that the existing literature offers in modeling employability.

Systematic review conducted on the PRISMA standard comprised four stages: identification, screening, eligibility, and inclusion as advised by the literature [30]. The identification process considered authoritative scientific articles on employability prediction that were published between 2008 and 2021. Justification for the temporal scope is that 2008 and later years saw a near-global economic crisis that impacted negatively on employability, with results still being felt to date [31]. Similarly, in 2020 and 2021, the COVID-19 pandemic and its associated preventive measures caused the closure of many businesses, loss of employment, a rise in the cost of living, and deterioration of most socio-economic aspects. As such, the pandemic may have revealed to mankind a new paradigm that calls for the construction of a new economic order, and the principle that needs, wishes, and resources are limited. The resultant shift of several economic activities into the virtual world affected employability since it caused companies to significantly reduce the number of employees [32].

The SLR process adopted the following set of steps:

1. The authors searched the ScienceDirect, Springer, Wiley, IEEE Xplore, and Taylor and Francis databases and collected relevant research articles, i.e., those dealing with the application of data-mining techniques in employability prediction.
2. An examination of the collected items was conducted to establish that articles met the predefined criteria outlined in Table 1.
3. The articles selected from step 2 above were analyzed and results synthesized.
4. A report was compiled from result of the identified research papers in order to make reasonable inference and recommendations for further research in the field of employability prediction and prescription using data-mining techniques.

Table 1. Criteria for including and excluding papers

Exclusion	Inclusion
Papers with qualitative research	Papers containing empirical research that applies descriptive, predictive, and prescriptive analyses
Book chapters	Full-length papers published in international conference proceedings and international colloquia
Workshops/Notes/Essay Papers published in a predatory journal	Papers published in peer-review journals with an impact factor of 1 or greater
Theoretical papers on predictive models	Papers presenting quantitative results
Predictive data-mining-technique papers without a value of accuracy of the model used	Research articles applying predictive data-mining techniques, papers on prescriptive models, and other unsupervised techniques (PCA, text mining, clustering)

Applying the exclusion and inclusion rules in Table 1, a total of 20 papers were identified. These included one international colloquium article, ten papers presented at international conferences and published in high-impact-factor journals, and nine articles published in peer-review journals. It was found that limited work has been done to assess graduate employability using predictive data-mining techniques. The majority of identified literature had adopted a theoretical perspective to assess employability using statistical methods [33].

The authors also used the unsupervised ML technique to model the topics. This involved the use of latent Dirichlet allocation (LDA), a natural-language processing approach for discovering common topics across the articles under analysis and, thus, the most frequent words on employability and data mining. A geographical heat map was used to plot the number of articles as grouped by author and nationality. Figure 1 illustrates the stages adopted to create the database of current research based on the PRISMA standard.

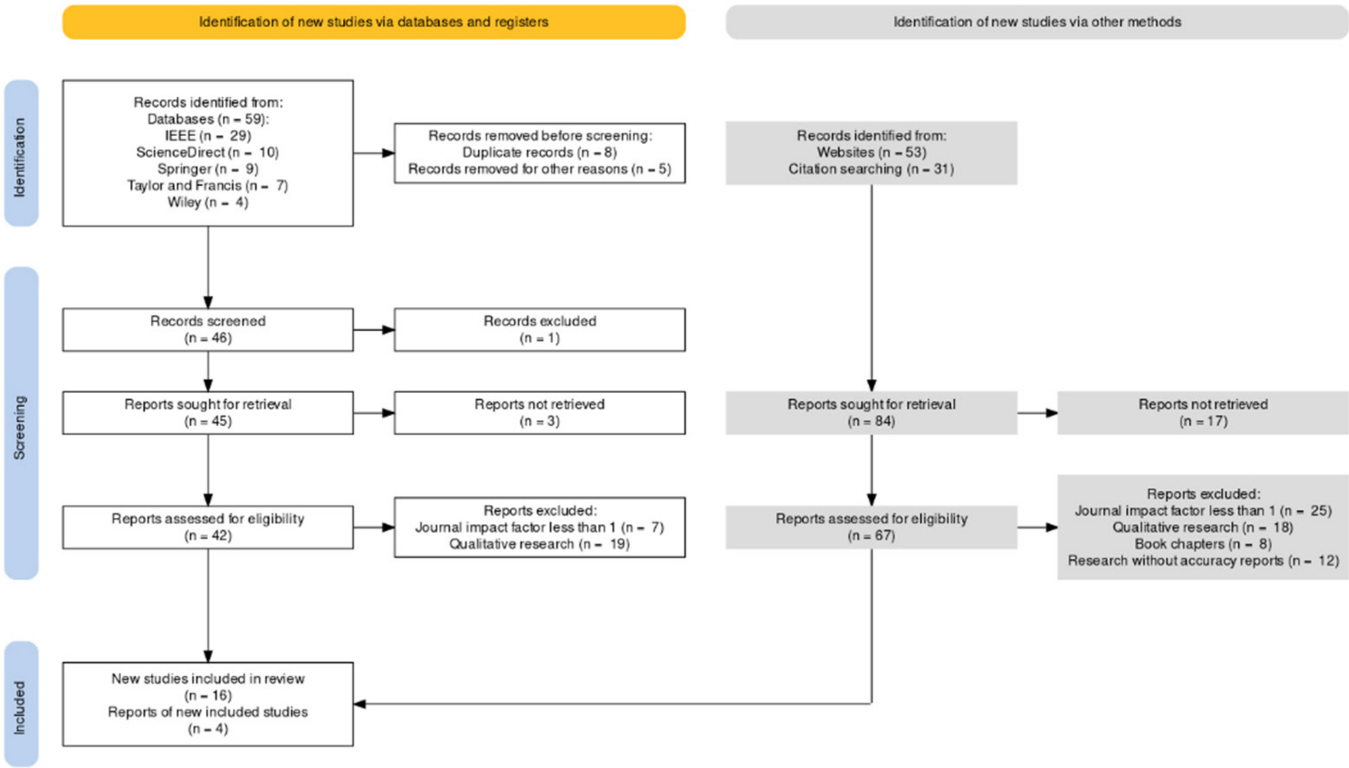

Fig. 1. Flowchart of the selection process

4 RESULTS

4.1 Number of publications by year

Out of the identified research articles, the following descriptive results regarding the number of yearly published papers were identified (Table 2). 2020 had the largest number of relevant published articles, while 2013 had the smallest.

Table 2. Statistics of articles on prediction and prescription of employability

Year	Number of Published Papers	Frequency
2013	1	5%
2016	2	10%
2017	2	10%
2018	4	20%
2019	4	20%
2020	5	25%
2021	2	10%
Total	**20**	**100%**

4.2 Statistics by paper type, publisher, and country of publication

Results from exploration of reviewed paper type (i.e., whether journal article or conference proceeding), publisher, country where the research was conducted, author, and the type of used model (descriptive, predictive, or prescriptive) are presented in Table 3. Most of the reviewed articles were focused on predictive techniques.

Table 3. Type of the paper, publisher, authors and countries where the research was conducted

Year	Type	Publisher	Country	Authors	Model
2013	Conference proceedings	Elsevier	Kingdom of Bahrain	Alsultanny [34]	Predictive
2016	Conference proceedings	IEEE	Philippines	Piad et al. [35]	Predictive
2016	Conference proceedings	AIP Publishing	Malaysia	Aziz and Yusof [5]	Predictive
2017	Article	Taylor & Taylor Group	England	Forsythe [36]	Descriptive
2017	Article	Kassel University Press	Spain	Bañeres and Conesa [37]	Prescriptive
2018	Article	INSIGHT—Indonesian Society for Knowledge and Human Development	Malaysia	Othman et al. [38]	Predictive
2018	Conference proceedings	IEEE	India	Sawla et al. [39]	Predictive
2018	Conference proceedings	IEEE	Saudi Arabia	Almutairi and Hasanat [40]	Predictive
2018	Article	Universidad Internacional de La Rioja	Spain	García-Peñalvo et al. [41]	Predictive
2019	Conference proceedings	IEEE	Philippines	Casuat and Festijo [42]	Predictive
2019	Conference proceedings	IEEE	USA	Dubey and Mani [43]	Predictive
2019	Conference proceedings	IEEE	India	Vashisht and Grover [44]	Predictive
2019	Conference proceedings	IEEE	Thailand	Watthananon and Chintanaporn [45]	Predictive
2020	Article	IEEE	England	Montañez and Hurst [46]	Predictive
2020	Article	Springer	China	Wang [47]	Predictive
2020	Article	IEEE	England	Sobnath et al. [48]	Predictive
2020	International colloquium	IEEE	Philippines	Casuat and Festijo [49]	Predictive
2020	Article	IEEE	Palestine	Qamhieh et al. [50]	Prescriptive
2021	Article	Lambert Publications	India	Suryawanshi et al. [51]	Prescriptive
2021	Conference proceedings	Springer	India	Purkar et al. [52]	Prescriptive

4.3 Analysis of topics covered in the reviewed literature

47 topics were identified in the corpus of texts reviewed in this SLR. The recurrent words [53] found in these topics can be seen in the word cloud in Figure 2:

Fig. 2. Word cloud of recurrent words in the existing literature

Table 4 illustrates the contents of all the 47 topics.

Table 4. Topics and information contents

Topics	Possible Content
1	Students' employability in high schools of business in Philippines by applying Radom Forest model
2	Attributes influencing graduates' employability in Philippines using Decision Tree
3	Skills of students as predictors of graduates' recruitment in Saudi Arabian Industry
4	Application of Decision Tree, Naive Bayes, and Neural Networks to predict employability of disabled persons in Thailand
5	Recommender system for career path guidance of engineering discipline using skills and competencies of graduates
6	Graduates' skill to predict employment in Malaysia using Decision Tree
7	Demographic attributes to predict employability of engineering students in India
8	A case study of predictive model of employability in Spain
9	An international conference paper on computational human interaction to predict labor market using Naive Bayes
10	Recommender system to provide services to small or part-time workers
11	Smart-metering infrastructure to collect data

(Continued)

Table 4. Topics and information contents *(Continued)*

Topics	Possible Content
12	Information about disabled students, disability type, and education
13	Computing framework and human-computer interaction according to Ting Wang
14	Information on IEEE publisher and international conference
15	Competence of graduates in Spain and their careers and employability
16–23, 25, 27–29, 42, 43	Human-computer interaction is integrated to construct a user-predictive model and person disability in Thailand
24	Application of Decision Tree and Naive Bayes as classifiers to predict occupation of disabled persons and students in Thailand
26	Employability of engineering students; data mining; and gender, age, performance, and skill of graduates in Philippines and India
30–32, 33, 35–39, 44–46	Nonverbal inference, intelligence and employment-rate prediction based on neural computing framework and human-computer interaction
34, 40	Saudi Arabian industry and recruitment of graduates. Challenge, impact, achievement, and employability
41	Employability of students in technological institute and higher education in Philippines
47	Data mining to predict employability, neural network, and linear model

4.4 Heat map of reviewed papers by country

Figure 3 below presents the number of retained reviewed papers by countries of publication using a heat map.

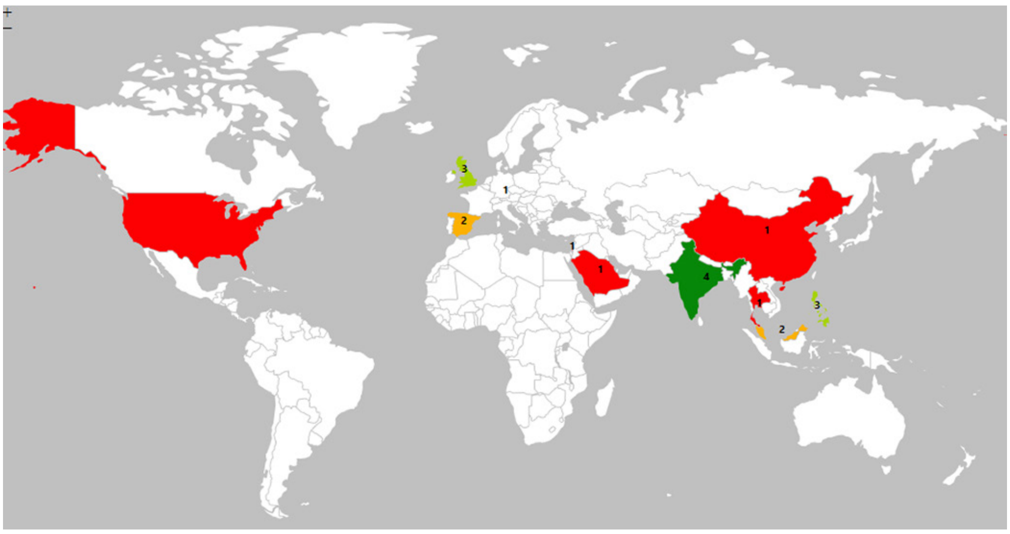

Fig. 3. Total number of reviewed papers by country

4.5 Applied ML algorithms for employability prediction

This subsection presents all ML algorithms that have been used in data mining to predict employability. It includes also, out of all the applied ML algorithms in each review paper, the best algorithm and its accuracy score.

Table 5 below presents the most applied ML algorithms in the existing literature.

Table 5. ML algorithms applied to predict graduates' employability

Approach	Year	Authors	Algorithms	Best Algorithms	Accuracy
Supervised	2013	Alsultanny [34]	Naive Bayes Classifier, Decision Rules, Decision Tree	Decision Tree	100%
	2016	Piad et al. [35]	Logistic Regression, CHAID Algorithm, Naive Bayes Classifier, Decision Tree (J48), SimpleCart	Logistic Regression	78.4%
	2016	Aziz and Yusof [5]	Naive Bayes Classifier, Logistic Regression, K-NN, Multilayer Perceptron, Decision Tree (J48)	Logistic Regression	92.47%
	2018	Othman et al. [38]	SVM (SMO), Decision Tree (J48), Multilayer Perceptron	Decision Tree (J48)	66.18%
	2018	Sawla et al. [39]	Random Forest, SVM, ANN	Random Forest	70%
	2018	Almutairi & Hasanat [40]	K-NN, Naive Bayes	Naive Bayes	69%
	2018	García-Peñalvo et al. [41]	Random Forest	Random Forest	70.2%
	2019	Casuat and Festijo [42]	SVM	SVM	91.22%
	2019	Dubey and Mani [43]	Logistic Regression, Random Forest, SVM, K-NN, Decision Tree	SVM, Random Forest	93%
	2019	Watthananon and Chintanaporn [45]	Decision Tree, Neural Networks, Naive Bayes	Naive Bayes	86.59%
	2020	Casuat and Festijo [49]	PCA applied in SVM	PCA applied in SVM	93%
	2020	Montañez and Hurst [50]	Multilayer Perceptron	Multilayer Perceptron	69%
	2020	Wang [47]	Neural Networks	Neural Networks	98%
	2020	Sobnath et al. [48]	Logistic Regression, Linear Discriminant Analysis Decision Tree, Gaussian NB, Decision Tree Classifier	Decision Tree Classifier and Logistic Regression	96%
Unsupervised	2017	Forsythe [36]	PCA	PCA	
	2019	Vashisht and Grover [44]	Ward's method	Ward's method	

4.6 Applied recommender system techniques for prescribing employability

This subsection presents different recommender systems techniques applied in the existing literature to prescribe either job or employability. It also identifies several knowledge gaps, such as sample limitations and the lack of consideration of white-collar jobs when modeling employability, in line with the sustainable-development goals. The commonly applied recommender system techniques are listed in Table 6 as follows.

Table 6. Recommender system techniques applied to prescribe graduates' employability

Year	Author	Techniques	Knowledge Gap
2017	Bañeres and Conesa [37]	Not clear from the literature	Limited sample comprising a single university.
2020	Qamhieh, Sammaneh, and Demaidi [50]	Fuzzy intelligence	Proposed a recommender system for higher education in Palestine, but socio-economic drivers were not considered in the study
2021	Suryawanshi, Patil, and Choudhari [51]	The study proposed to use ANN	The study did not implement the recommender system. It proposed only a graphical framework
2021	Purkar, et al. [52]	Content based	The model was limited to informal jobs.

4.7 Data mining tools and techniques applied in previous studies

Out of the total articles selected based on the criteria outlined in Table 1, six papers applied the Waikato Environment for Knowledge Analysis (WEKA) tool for the data analysis. One paper used libraries from the R programming language, and two papers applied the Python Scikit Learn. Three of the reviewed papers used SPSS, and nine papers did not specify the tools applied for the analysis. The distribution is illustrated in Table 7.

Table 7. Description of the tools applied to model employability

Authors	Year	Tools Used			
		WEKA	R	Python/ Scikit Learn	SPSS
Alsultanny [34]	2013	Yes	No	No	No
Piad, et al. [35]	2016	Yes	No	No	Yes
Aziz and Yusof [5]	2016	Yes	No	No	No
Forsythe [36]	2017	No	No	No	No
Bañeres and Conesa [37]	2017	No	No	No	No
Othman et al. [38]	2018	Yes	No	No	No
Sawla et al. [39]	2018	No	Yes	No	No
Almutairi and Hasanat [40]	2018	No	No	No	No
García-Peñalvo et al. [41]	2018	No	No	Yes	No
Festijo and Casuat [42]	2019	Yes	No	No	No
Dubey and Mani [43]	2019	No	No	Yes	No
Vashisht and Grover [44]	2019	No	No	No	Yes
Watthananon and Chintanaporn [45]	2019	Yes	No	No	No
Montañez and Hurst [46]	2020	No	No	No	No
Wang [47]	2020	No	No	No	No
Sobnath et al. [48]	2020	No	No	No	No
Casuat and Festijo [49]	2020	No	No	No	Yes
Qamhieh et al. [50]	2020	No	No	No	No
Suryawanshi et al. [51]	2021	No	No	No	No
Purkar et al. [52]	2021	No	No	No	No

4.8 Assessment of criteria used to determine employability in previous studies

Several criteria have been considered in the previous literature for determining employability. Table 8 outlines the criteria used in the reviewed studies.

Table 8. Criteria for employability determination

Description of Criteria Considered	Source	Year
Forecast of the labor market in Kingdom of Bahrain	Alsultanny [34]	2013
IT employability in Philippines	Piad et al. [35]	2016
IT graduates employed in MARA Professional College (Kolej Profesional MARA)	Aziz and Yusof [5]	2016
Perceptions of employability in higher education through student survey	Forsythe [36]	2017
Development of recommender system as a prototype for knowledge and skills matching for university courses	Bañeres and Conesa [37]	2017
Graduates' employability prediction in Malaysia	Othman et al. [38]	2018
Employability measurement based on the candidate's vocal analysis during job interviews	Sawla et al. [39]	2018
Suitability analysis of information system students' skills	Almutairi and Hasanat [40]	2018
Graduate employability factors in Spain	García-Peñalvo, et al. [41]	2018
Analysis of students' mock job interview results	Casuat and Festijo [42]	2019
Assessment of employability for high school students in Loudoun County (Virginia, USA)	Dubey and Mani [43]	2019
Profile of engineering students in India through clustering	Vashisht and Grover [44]	2019
Assessment of persons with disabilities in IT occupation in Thailand	Watthananon and Chintanaporn [45]	2019
Smart metering infrastructure to detect unemployment in England	Montañez and Hurst [46]	2020
Prediction of the employment rate via an intelligent model based on neural computing	Wang [47]	2020
Employment assessment for UK disabled students engaged in post-higher education	Sobnath et al. [48]	2020
Employability assessment for undergraduate students in the Philippines	Casuat and Festijo [49]	2020
Recommender system for skills matching into engineering programs among high school students	Qamhieh, Sammaneh, and Demaidi [50]	2020
Post-graduate student employability assessment and skills matching	Suryawanshi et al. [51]	2021
Recommender system for skills matching of recruitment candidates	Purkar et al. [52]	2021

The first question of this research that was addressed was (RQ1): Which ML algorithms were the most used in data mining to predict and/or prescribe employability? Six ML algorithms in the reviewed research were most used in data mining to predict employability. The decision tree with different types of algorithms such as J48 was the most used. Eight studies were identified that predicted employability by applying this algorithm. The best score for these different works was 100% accuracy. The decision tree has several advantages. This algorithm identifies different sets of predictors and different interactions between them for different subgroups in a dataset [54]. Being a nonparametric algorithm [55], it spares researchers from making distributional and metric assumptions about the data, especially since collected data are often correlated and skewed [56]. Its ease of representation and interpretation allows the researcher to represent results in a non-statistically intensive and comprehensive framework for novices in data science [57].

The next-most applied algorithm is the Multilayer Perceptron neural network (MLP). Six researches applied this algorithm to predict employability. The highest score of accuracy was 98%. This ML algorithm is a feed-forward model [58], in which learning is through two phases: forward and backward [59]. Given the loss function $E(x, y, \theta)$, with x the inputs, θ parameters, y the target, and w_{ij}^{h} as the weight of the connection from neuron i of layer h-1 to neuron j, the MLP algorithm uses gradient descent theory or its variants to adjust weights by applying partial derivation of the loss function regard to each parameter [60] to get a new weight, as follows:

$$w_{ij}^{h+} = w_{ij}^{h} - \eta * \frac{\partial E(x, y, \theta)}{\partial w_{ij}^{h}} \tag{1}$$

η represents the learning rate, which is very important. It is a hyper-parameter to speed up the training of the Artificial Neural Network (ANN) and is optimized based on the error of the validation set after a large number of neuron weight updates [61]. Five reviewed studies have used SVM to predict employability. The SVM algorithm offers robustness in learning biases in data [62]. It is simple and best suited to small sample cases [63]. Dubey and Mani [43] applied SVM to predict high school student employability in the US and got the an accuracy score of 93%. The same score was obtained by Casuat and Festijo [49] when coupled PCA and SVM were used to predict undergraduates' employability In the Philippines, while identifying the most predictive variables among the employability signals of students, they used a dataset of 3000 engineering students.

Besides SVM, Naive Bayes has also been applied five times in the surveyed research. Based on the theorem of Bayes, it is known for calculating probabilities and conditional probabilities. The Naive Bayes is a faster algorithm and offers accurate models [64]. The Naive Bayesian classifier algorithm was used by Alsultanny [34] to create a dataset table in order to classify a series of unknown information for employment to handle the constant change of labor market demands in the Kingdom of Bahrain. He got an accuracy of 83.33%. Naive Bayes performed better than Decision tree and Neural network when it was applied by Watthananon and Chintanaporn [45] to predict IT Occupation of persons with disabilities in Thailand. It presented an accuracy of 86.59%, while Decision Tree presented 82.03% and Neural Network presented 66.10 %. The results of the two researchers confirm the statement of Mansour et al. [65] and Jackins et al. [66], who say that despite the fact that Naive Bayes is among the simplest ML algorithms, it is widely used to solve practical problems. In the reviewed papers, the Naive Bayes classifier algorithm was used, but this algorithm also offers the possibility of performing regression. The Naive Bayes classifier is a probabilistic ML model based on Bayes' theorem [67] and is written as follows:

$$P(c|A) = \frac{P(A|c)P(c)}{P(A)} \tag{2}$$

In the equation (2), c represents the class of variables that are supposed to be predicted and A can be considered as parameters or features [68]. Features are also called predictors in Bayes' theorem and are independent variables. Thus, the presence of one does not affect the other. That is why they are called naive [69]. Logistic regression, a statistical ML algorithm [69] that fits a regression surface to data in the case of dichotomous dependent variables [70], was applied in four different projects, and the highest accuracy was 96%, as shown in Table 5. Sobnath et al. [48] applied this algorithm to a dataset of 270,934 records of students with a known disability in UK schools since logistic regression requires large sample sizes to achieve an adequate level of stability [71]. However, in the case of larger data sets and large features, the inference of logistic models can become computationally prohibitive and problematic [72]. Random Forest was used three times in the reviewed research. This ML algorithm is flexible and gives a great result, even without hyper-parameter tuning [73]. It can be used for both classification and regression tasks [74]. Its simplicity makes it popular. Random Forest is an ensemble learning model which combines multiple trees to overcome the overfitting of individual decision trees [75] by building each tree randomly and drawing samples with replacements from the original data [76]. Dubey and Mani [43] got an accuracy of 93% when applying the Random Forest classifier to predict high school students' employability with local businesses for part-time jobs in Loudoun County in Virginia (USA). In sum, the Random Forest does not always perform better in various practical cases. Thus, several researchers have been able to propose methods adapted to practical problems that improve different aspects of the Random Forest [77].

The literature review revealed that only a few studies have been conducted to prescribe employability, and out of four reviewed prescriptive works, the authors conclude that TF-IDF was the algorithm most used to vectorize data, making it suitable for similarity calculation [52], and Cosine similarity was the recurrent model used to compute similarities.

4.9 The used factors/variables to predict and/or prescribe employability

This section discusses research question **(RQ2):** Which are the factors identified in the previous literature for predicting graduate employability? In the reviewed papers, there is no one who conducted an exploratory factor analysis (EFA) to capture contextual factors that affect employability. Nevertheless, the authors noticed that in about eight papers, some factors/attributes were retained as predictors of employability in different cases. According to Piad et al. [35], IT Core, IT Professionally, and Gender constitute the factors of predicting employability of graduates in IT. For Aziz and Yusof [5], Gender, program of courses, semester (number of semesters taken to complete the program), Co-curriculum, and Academic performance were conditions affecting employability at MARA Professional College, in Malaysia.

Forsythe [36], in her study on the predictors of employability perceptions in higher education, was able to identify four factors influencing employability: Adaptability, Goal orientation, Family support, and Spirituality. She concluded that mindset is the most important factor influencing students'' self-perceived employability. She thus proposed that future research should include in-depth work on students' attitudes towards their employability, as these attitudes change throughout their course of study, and the role that goal orientation plays over time. On the other hand, García-Peñalvo et al. [41] estimate that there are six factors that affect employability: university where the graduate studied, graduate gender, degree studied by the graduate, graduate's opinion on the relevance of choice of a job (which depends on the conditions related to the prestige of the employer and the position which he/she will occupy with the company), graduate's opinion about the relevance of a chosen job (which depends on the conditions related to the personal context), and graduate's reasons to live abroad while obtaining the degree.

Othman et al. [38] retained age, faculty, field of study, co-curriculum, marital status, industrial internship, and English skills as factors affecting graduate employability in Malaysia. Casuat and Festijo [49] said that there are some attributes among employability signals of undergraduate students. Hence, mental alertness and manner of speaking are the highest predictors of employability. Dubey and Mani [43] concluded that in Loudoun County (Virginia, USA), small-business employers focus on the students' GPA, school where the student studied, grade level, gender, previous work experience, number of courses taken by the student, social and task skills, available days to work, and transportation abilities when they are employing high school students. Sobnath et al. [48] said that Age, Disability type, and Institution where the person studied are the most common attributes affecting disabled students' engagement six months after their graduation in UK.

Regarding reviewed prescriptive studies, all the papers insist on the use of skills as main factors that prescribe employability. However, Qamhieh et al. [50] suggested that modeling employability in developing countries requires the use of economics and socio-political factors, as they are the reliable factors that influence concrete employability in such countries.

RQ3: What is the applicability of existing models for predicting graduate employability in unstable socio-economic and political backdrops?

This study focused on reviewing previous research works on building models to predict or to prescribe graduates' employability from 2008 to 2021. The authors developed a systematic approach to the review work. This approach included, first, the creation of rules for selecting papers in scientific databases and, second, applying a PRISMA method checklist, which guided them on how to develop this systematic review and what to include when they were writing this review [78]. This approach supported the main objective of this research, which was to capture relevant factors that can be used to better predict or prescribe employability in unstable developing countries such as the DRC.

Several ML algorithms in these reviewed papers have shown different performances, and it is not clear whether logistic regression is the best or whether the best algorithm is Decision Tree or even Naive Bayes or SVM. These algorithms performed well, depending on whether the research was modeling the employability of disabled people or graduates, and even depending on the country. Thus, there is a need to try to use an ensemble predictor such as stacking by including more factors to see if this ensemble method improves the accuracy in predicting employability [79]. Moreover, as factors are changing according to the country and degree of graduates, internal factors should be considered also, which can influence employability in an unstable developing country, such as the DRC, where many factors were pointed out by some qualitative researches.

On the side of prescriptive analysis, traditional recommender systems (Content based, Fuzzy intelligence based, etc.) seem to be the most used for employability. Researchers should consider a huge of set of factors in order to predict employability. Globalization; different crises; social issues such as tribalism, nepotism, and favoritism; and the impact of technology greatly influence recruitment practices in Africa. Employees should have today's skills and performance capable of adaptation in a changing world and a complex social background [80]. Moreover, nowadays, deep-learning technique have been revealed to be the best practice if one wants to optimize recommender system models [81], [82], [83]. SWOT analysis results for previous studies are shown in Table 9 below.

Table 9. SWOT of application of data-mining techniques on employability research

Strengths	Weaknesses
Predicting graduates' employability using skills as factors helps universities to improve strategies of their formations Use of huge datasets allows obtaining good models Giving new trends and applications of applying data-mining techniques for students' and graduates' employability	Lack of research on this area in Africa, especially in the DRC, an unstable developing country No factor analysis research that identifies contextual factors that predict employability Fewer retained attributes to predict employability Insisting only on the skills of students/graduates Not taking into account influence of some social considerations, such as racial and religion discriminations and the need of employers
Opportunities	**Threats**
Apply a comparative study of the most used ML models Use of Stacking algorithm as ensemble model to combine the results of Decision Tree, MLP, Naive Bayes, Logistic Regression, SVM, and Random Forest for a better prediction Surveying graduates themselves to get primary data Conducting a factor analysis to capture contextual factors that predict employability in unstable developing countries, particularly in the DRC Predict employability of IT graduates in the DRC, a country with manifold socio-politic realities, by including a large set of attributes taking into account the impact of COVID-19 and socio-political situation of the DRC Using prescriptive analytics to build a recommender system by using primary data and capturing contextual factors using EFA	Use of secondary data cannot sometimes reflect the real world Many papers on employability are published in non-impact journals and others in predatory journals. Hence, the risk of the quality of the research Skill is pointed by many studies as a main factor that predict and prescribe employability. However, in unstable developing countries, some inner factors predict and prescribe employability [84]. Not using deep-learning technique can make recommender systems generate recommendations that are not highly accurate

5 DISCUSSION

In this SLR, the authors found that, looking at the results presented in the subsection 4.9, there is inconsistency in factors that predict employability. Moreover, the outcomes of the **RQ1** show that it is not clear which algorithm is appropriate to build an efficient classifier to predict employability. In addition, the existing data-mining models developed to mitigate the employability issues were implemented mostly in developed countries. This SLR estimates that it is important to first conduct an EFA study to capture contextual factors that predict employability as long as the existing studies present inconsistent factors. Secondly, instead of developing only predictive models, it is imperative to use current techniques such as deep learning to build prescriptive solutions in order to mitigate the unemployment problems in unstable developing countries, especially in the DRC, where the rate of youth unemployment has reached a worrying threshold [85]. Therefore, this study should contribute to the body of knowledge by providing an overview of existing data-mining models used to either predict or prescribe employability. The review proposes some recommendations for future studies, which may lead to adequate IT/IS solutions to address the unemployment problem in unstable developing countries. These solutions will be relevant only if researchers can identify consistent contextual factors that influence employability and use them as features to construct performance-predictive and -prescriptive models.

6 LIMITATIONS OF THE STUDY

This study had some limitations, which influenced some of its results. In fact, many of the papers obtained from databases (ScienceDirect, Springer, Wiley, IEEE Xplore, and Taylor and Francis, etc.) were published in journals with an impact factor <1, and other papers obtained from ResearchGate were published in predatory journals. As result, the authors reviewed only 20 papers according to their rules (see Table 1). Moreover, the Kingdom of Bahrain was not recognized in the Excel heat map as a country (see Figure 3).

7 CONCLUSION

The research adopted the PRISMA method to determine existing studies to include as items for this SLR. Out of 143 papers related to modeling employability using data mining, only 20 predicting and/or prescribing employability were retained as adequate studies for conducting this study. Moreover, the study used SWOT as a qualitative approach to analyze the strengths of existing studies on the application of data mining in employability and to detect the conceptual and knowledge gaps in order to identify opportunities that the existing literature offers in the modeling of employability.

The result showed that six ML algorithms were the most used in data mining to predict employability: Decision tree (J48), MLP, SVM, Naive Bayes, Logistic regression, and Random Forest. From the 20 papers, J48 decision tree algorithm was identified in eight different papers, and the highest accuracy in those papers was 100%. MLP was identified in six reviewed studies, and the best score was 98%. SVM and Naive Bayes were repeated in five different papers and had the best accuracies: 93% and 86.59%, respectively. Logistic regression, presented in four papers, had 93% as highest accuracy. Random Forest was used in three papers, and the highest accuracy was 93%. Regarding studies on prescribing employability, the authors discovered that collaborative filtering and content-based techniques were the recommender systems most used to prescribe employability. The study found that no paper has conducted EFA to identify contextual factors that predict and/or prescribe employability. All the used variables or factors in the existing literature to predict and/or prescribe employability were either extracted using feature selection or feature engineering. In addition, none of the studies have addressed the recommendation of employability in unstable developing countries by applying data mining.

8 FUTURE WORK

Future work will involve the arduous task of identifying contextual factors that predict graduates' employability in a socially and politically unstable developing country, such as the DRC, in addition to skills [86], demographics, and graduate performance. The identification of those factors may help the authors in the future to conduct research on the prediction of graduate employability. Moreover, there is also a need for building a recommender system based on the socio-political and economic backgrounds of students, which can help either students or new graduates to update their competencies to maximize their chance of being employable in order to achieve the sustainable development goal target of 8.6 in such countries.

9 ACKNOWLEDGMENT

The authors express their gratitude to the Assumptionist Community of Kijenge, Arusha/Tanzania, for financially supporting the publication of this research.

10 REFERENCES

[1] L.D. Pool and P. Qualter, "Emotional Self-Efficacy, Graduate Employability, and Career Satisfaction: Testing the Associations," *Australian Journal of Psychology*, vol. 65, pp. 214–223, 2013. https://doi.org/10.1111/ajpy.12023

[2] A.M. Ahmed, A. Rizaner, and A.H. Ulusoy, "Using Data Mining to Predict Instructor Performance," *12th International Conference on Application of Fuzzy Systems and Soft Computing*, Vienna: Procedia Computer Science, pp. 137–142, 2016. https://doi.org/10.1016/j.procs.2016.09.380

[3] A. Jha, M. Dave, and A. Madan, "A Review on the Study and Analysis of Big Data using Data Mining Techniques," *International Journal of Latest Trends in Engineering and Technology*, vol. 6, no. 3, pp. 94–102, 2016.

[4] A. Moumen, E.H. Bouchama, and Y.E. Idirissi, "Data Mining Techniques for Employability: Systematic Literature Review," *IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science*, pp. 1–5, 2020. https://doi.org/10.1109/ICECOCS50124.2020.9314555

[5] M.T. Aziz and Y. Yusof, "Graduates Employment Classification using Data Mining Approach," *AIP Conference Proceedings*, 1761, 2016. https://doi.org/10.1063/1.4960842

[6] G. Piccoli and B. Ives, "Review: IT-Dependent Strategic Initiatives and Sustained Competitive Advantage: A Review and Synthesis of the Literature," *MIS Quarterly*, vol. 29, no. 4, pp. 747–776, 2005. https://doi.org/10.2307/25148708

[7] V. Venkatesh, F.D. Davis, and M.G. Morris, "Dead or Alive? The Development, Trajectory and Future of Technology Adoption Research," *Journal of the Association for Information Systems*, vol. 8, no. 4, pp. 267–286, 2007. https://doi.org/10.17705/1jais.00120

[8] G.M. Lundberg, B.R. Krogstie, and J. Krogstie, "From Employable to Fully Operational: The Need for Training of Computer Science Graduates," *International Journal of Engineering Pedagogy (iJEP)*, vol. 11, no. 3, pp. 21–38, 2021. https://doi.org/10.3991/ijep.v11i3.19995

[9] R. Bridgstock, "The Graduate Attributes we've Overlooked: Enhancing Graduate Employability through Career Management Skills," *Higher Education Research and Development*, vol. 28, no. 1, pp. 31–44, 2009. https://doi.org/10.1080/07294360802444347

[10] D.S. Rychen and L.H. Salganik, Key Competencies for A Successful Life and Well-functioning Society, Boston: Hogrefe and Huber, 2003.

[11] J. Rich, (2015). "We Need to Talk About Employability, not Employment. Times Higher Education." timeshighereducation.com. https://www.timeshighereducation.com/comment/we-need-to-talk-about-employability-not-employment (accessed March 15, 2022).

[12] F. Suleman, "The Employability Skills of Higher Education Graduates: Insights into Conceptual Frameworks and Methodological Options," *Higher Education*, vol. 76, no. 2, pp. 263–278, 2018. https://doi.org/10.1007/s10734-017-0207-0

[13] S.J. Mgaiwa, "Fostering Graduate Employability: Rethinking Tanzania's University Practices," *SAGE Open*, pp. 1–14, 2021. https://doi.org/10.1177/21582440211006709

[14] ILO, COVID-19: Public Employment Services and Labour Market Policy Responses, Geneva: International Labour Organization, 2020.

[15] W. Nuankaew and P. Nuankaew, "Educational Engineering for Models of Academic Success in Thai Universities During the COVID-19 Pandemic: Learning Strategies for Lifelong Learning," *International Journal of Engineering Pedagogy (iJEP)*, vol. 11, no. 4, pp. 96–114, 2021. https://doi.org/10.3991/ijep.v11i4.20691
[16] I.T. Terwase, et al., "Conflict, War, and Peace Building in Africa as Emerging Economy," *Journal of Humanities*, vol. 1, no. 1, pp. 1–12, 2019.
[17] AfDB, African Economic Outlook 2012: Promoting Youth Employment, African Development Bank, 2012.
[18] R. Etshim, "Collaboration between Higher Education and Labor Market in Kinshasa, DR Congo," *Journal of International Education and Leadership*, vol. 7, no. 1, pp. 1–17, 2017.
[19] L.V. Rueden, et al., (2021). "Informed Machine Learning – A Taxonomy and Survey of Integrating Prior Knowledge into Learning Systems," *IEEE Transactions on Knowledge and Data Engineering*, 2021. https://doi.org/10.1109/TKDE.2021.3079836
[20] F. Chollet, Deep Learning with Python, Shelter Island: Manning Publications Co., 2018.
[21] B. Abbasi and D.M. Goldenholz, "Machine Learning Applications in Epilepsy," *Epilepsia*, pp. 2037–2047, 2019.
[22] J. Wei, et al., "Machine Learning in Materials Science," *InfoMat*, pp. 338–358, 2019. https://doi.org/10.1002/inf2.12028
[23] W. Sha, et al., "Machine Learning in Polymer Informatics," *InfoMat*, vol. 3, no. 2, pp. 353–361, 2021. https://doi.org/10.1002/inf2.12167
[24] V. François-Lavet, et al., "An Introduction to Deep Reinforcement Learning," *Foundations and Trends® in Machine Learning*, vol. 11, no. 3–4, pp. 219–354, 2018. https://doi.org/10.1561/2200000071
[25] M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," *Mobile Netw Appl*, pp. 171–209, 2014. https://doi.org/10.1007/s11036-013-0489-0
[26] R. Tamilselvi and S. Kalaiselvi, "An Overview of Data Mining Techniques and Applications," *International Journal of Science and Research*, vol. 2, no. 2, pp. 506–509, 2013.
[27] P.P. Sondwale, "Overview of Predictive and Descriptive Data Mining Techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 4, pp. 262–265, 2015.
[28] M.J. Page, et al., "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews," *Systematic Reviews*, vol. 10, no. 89, 2021. https://doi.org/10.1186/s13643-021-01626-4
[29] E. Gürel and M. Tat, "SWOT Analysis: A Theoretical Review," *The Journal of International Social Research*, vol. 10, no. 51, pp. 994–1006, 2017.
[30] A. Liberati, et al., "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Health Care Interventions: Explanation and Elaboration," *Journal of Clinical Epidemiology*, vol. 62, pp. e1–e34, 2009. https://doi.org/10.1016/j.jclinepi.2009.06.006
[31] P. Ramskogler, "Tracing the Origins of the Financial Crisis," *OECD Journal*, pp. 47–61, 2015. https://doi.org/10.1787/fmt-2014-5js3dqmsl4br
[32] M. Toprak, Y. Bayraktar, and A. Özyilmaz, A. (2020). "The COVID-19 Pandemic and the Digital Transformation in Turkish Higher Education: An Evaluation from the Perspective of Industry 4.0 and Society 5.0," in *The COVID-19 Pandemic and Its Economic, Social, and Political Impacts,* D. Demirbas, V. Bozkurt, and S. Yorgun, Eds. Istanbul: Istanbul University Press, 2020, pp. 167–185.
[33] Y. Bharambe, et al., "Assessing Employability of Students using Data Mining Techniques," *International Conference on Advances in Computing, Communications and Informatics,* IEEE, pp. 2110–2114, 2017. https://doi.org/10.1109/ICACCI.2017.8126157
[34] Y.A. Alsultanny, "Labor Market Forecasting by Using Data Mining," *International Conference on Computational Science*, pp. 1700–1709, 2013. https://doi.org/10.1016/j.procs.2013.05.338

[35] K.C. Piad, et al., "Predicting IT Employability Using Data Mining Techniques," *The Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications*, IEEE, pp. 26–30, 2016. https://doi.org/10.1109/DIPDMWC.2016.7529358

[36] A. Forsythe, "I Doubt Very Seriously whether Anyone will Hire Me; Factors Predicting Employability Perceptions in Higher Education," *Cogent OA*, vol. 4, no. 1, pp. 1–13, 2017. https://doi.org/10.1080/23311908.2017.1385131

[37] D. Bañeres and J. Conesa, "A Life-long Learning Recommender System to Promote Employability," *International Journal of Emerging Technologies in Learning*, vol. 12, no. 6, pp. 77–93, 2017. https://doi.org/10.3991/ijet.v12i06.7166

[38] Z. Othman, et al., "Classification Techniques for Predicting Graduate Employability," *International Journal on Advanced Science Engineering Information Technology*, vol. 8, no. 4–2, pp. 1712–1720, 2018. https://doi.org/10.18517/ijaseit.8.4-2.6832

[39] H. Sawla, et al., "Computational Inference of Candidates' Oratory Performance in Employment Interviews based on Candidates' Vocal Analysis," *Proceedings of the Second International Conference on Intelligent Computing and Control Systems*, IEEE, pp. 164–169, 2018. https://doi.org/10.1109/ICCONS.2018.8662991

[40] M.M. Almutairi and M.H. Hasanat, "Predicting the Suitability of IS students' Skills for the Recruitment in Saudi Arabian Industry," *21st Saudi Computer Society National Computer Conference*, IEEE, pp. 1–6, 2018. https://doi.org/10.1109/NCG.2018.8593016

[41] F.J. García-Peñalvo, et al., "Proposing a Machine Learning Approach to Analyze and Predict Employment and its Factors," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, pp. 39–45, 2018. https://doi.org/10.9781/ijimai.2018.02.002

[42] C.D. Casuat and E.D. Festijo, "Predicting Students' Employability using Machine Learning Approach," *6th IEEE International Conference on Engineering Technologies and Applied Sciences*, Kuala Lumpur: IEEE, pp. 1–5, 2019. https://doi.org/10.1109/ICETAS48360.2019.9117338

[43] A. Dubey and M. Mani, "Using Machine Learning to Predict High School Student Employability – A Case Study," *EEE International Conference on Data Science and Advanced Analytics*, pp. 604–605, 2019. https://doi.org/10.1109/DSAA.2019.00078

[44] M.G. Vashisht and R. Grover, "Employability Profiling of Engineering Students Using Clustering," *5th IEEE International Conference on Signal Processing, Computing and Control*, IEEE, pp. 191–195, 2019. https://doi.org/10.1109/ISPCC48220.2019.8988505

[45] J. Watthananon and P. Chintanaporn, "Utilizing Data Mining Techniques to Predicting IT Occupation of Persons with Disabilities in Thailand," *34th International Technical Conference on Circuits/Systems, Computers and Communications*, IEEE, pp. 1–4, 2019. https://doi.org/10.1109/ITC-CSCC.2019.8793319

[46] C.A. Montañez and W. Hurst, "A Machine Learning Approach for Detecting Unemployment Using the Smart Metering Infrastructure," *IEEE Access*, vol. 8, pp. 22525–22536, 2020. https://doi.org/10.1109/ACCESS.2020.2969468

[47] T. Wang, "Intelligent Employment Rate Prediction Model based on a Neural Computing Framework and Human–Computer Interaction Platform," *Neural Computing and Applications*, pp. 16413–16426, 2020. https://doi.org/10.1007/s00521-019-04019-w

[48] D. Sobnath, et al., "Feature Selection for UK Disabled Students' Engagement Post Higher Education: A Machine Learning Approach for a Predictive Employment Model," *IEEE Access*, vol. 8, pp. 159530–159541, 2020. https://doi.org/10.1109/ACCESS.2020.3018663

[49] C.D. Casuat and E.D. Festijo, "Identifying the Most Predictive Attributes Among Employability Signals of Undergraduate Students," *16th IEEE International Colloquium on Signal Processing & Its Applications*, IEEE, pp. 203–206, 2020. https://doi.org/10.1109/CSPA48992.2020.9068681

[50] M. Qamhieh, H. Sammaneh, and M.N. Demaidi, "PCRS: Personalized Career-Path Recommender System for Engineering Students," *IEEE Access*, vol. 8, pp. 214039–214049, 2020. https://doi.org/10.1109/ACCESS.2020.3040338

[51] K. Suryawanshi, A. Patil, and S. Choudhari, "Proposed Recommended System for Employability Skilling using Neural Network," *International Journal of Advanced Research in Science, Communication and Technology*, vol. 4, no. 3, pp. 111–117, 2021. https://doi.org/10.48175/IJARSCT-V4-I3-017

[52] M. Purkar, et al., "Recommendation System for Workers & Customers for Informal Jobs," *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*, 2021. https://doi.org/10.2139/ssrn.3833762

[53] Valentina Ibrahim, et al., "A Word Cloud Model based on Hate Speech in an Online Social Media Environment," *Baghdad Science Journal*, vol. 18, no. 2, pp. 937–946, 2021. https://doi.org/10.21123/bsj.2021.18.2(Suppl.).0937

[54] E. Park and J. Dooris, "Predicting Student Evaluations of Teaching using Decision Tree Analysis," *Assessment & Evaluation in Higher Education*, vol. 45, no. 5, pp. 776–793, 2020. https://doi.org/10.1080/02602938.2019.1697798

[55] Y. Song and Y. Lu, "Decision Tree Methods: Applications for Classification and Prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, pp. 130–135, 2015.

[56] M.F. Gadi, A.P. Lago, and J. Mehnen, "Data Mining with Skewed Data," in *New Advances in Machine Learning*, Y. Zhang, Ed. London: InTech, 2010, pp. 173–187. https://doi.org/10.5772/9382

[57] L. Rokach and O. Maimon, "Decision Trees," in *Data Mining and Knowledge Discovery Handbook*, L. Rokach and O. Maimon, Eds. Boston: Springer, 2005, pp. 165–192. https://doi.org/10.1007/0-387-25465-X_9

[58] H.N. Mpia and N.B. Inipaivudu, "Gradient Back-Propagation Algorithm in the Multi Layer Perceptron: Foundations and Case Study," *International Journal of Innovation and Applied Studies*, vol. 32, no. 2, pp. 271–290, 2021.

[59] H. Liu, et al., "Unmanned Aerial Vehicle Acoustic Localization Using Multilayer Perceptron," *Applied Artificial Intelligence*, vol. 35, no. 7, pp. 537–548, 2021. https://doi.org/10.1080/08839514.2021.1922849

[60] I.K. Omurlu, et al., "An Empirical Study on Performances of Multilayer Perceptron, Logistic Regression, ANFIS, KNN and Bagging CART," *Journal of Statistics and Management Systems*, vol. 23, no. 4, pp. 827–841, 2020. https://doi.org/10.1080/09720510.2019.1696924

[61] J. Park, D. Yi, and S. Ji, "A Novel Learning Rate Schedule in Optimization for Neural Networks and It's Convergence," *Symmetry*, vol. 12, no. 4, 2020. https://doi.org/10.3390/sym12040660

[62] S. Begum, et al., "Identification of Biomarker on Biological and Gene Expression Data using Fuzzy Preference Based Rough Set," *Journal of Intelligent Systems*, vol. 30, pp. 130–141, 2021. https://doi.org/10.1515/jisys-2019-0034

[63] S. Du and Y. Li, "A Novel Deformation Forecasting Method Utilizing Comprehensive Observation Data," *Advances in Mechanical Engineering*, vol. 10, no. 9, pp. 1–9, 2018.

[64] M. Mayo and E. Frank, "Improving Naive Bayes for Regression with Optimized Artificial Surrogate Data," *Applied Artificial Intelligence*, vol. 34, no. 6, pp. 484–514, 2020. https://doi.org/10.1080/08839514.2020.1726615

[65] N. Mansour, et al., "Accurate Detection of Covid-19 Patients based on Feature Correlated Naive Bayes (FCNB) Classification Strategy," *Journal of Ambient Intelligence and Humanized Computing*, 2021. https://doi.org/10.1007/s12652-020-02883-2

[66] V. Jackins, et al., "AI-Based Smart Prediction of Clinical Disease using Random Forest Classifier and Naive Bayes," *The Journal of Supercomputing*, vol. 77, pp. 5198–5219, 2021. https://doi.org/10.1007/s11227-020-03481-x

[67] J.C. Lamirel and D. Reymond, "Automatic Websites Classification and Retrieval using Websites Communication Signatures," *COLLNET Journal of Scientometrics and Information Management*, vol. 82, pp. 293–310, 2014. https://doi.org/10.1080/09737766.2014.954855
[68] Z. Hussain and M.D. Borah, "Nutritional Status Prediction in Neonate Using Machine Learning Techniques: A Comparative Study," in *Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science*, A. Bhattacharjee, S. Borgohain, B. Soni, G. Verma, and X. Gao, Eds. Singapore: Springer, Vol. 1241, 2020, pp. 69–83. https://doi.org/10.1007/978-981-15-6318-8_7
[69] H. Aldabbas, et al., "Google Play Content Scraping and Knowledge Engineering using Natural Language Processing Techniques with the Analysis of User Review," *Journal of Intelligent Systems*, vol. 30, pp. 192–208, 2021. https://doi.org/10.1515/jisys-2019-0197
[70] S. Sperandei, "Understanding Logistic Regression Analysis," *Biochemia Medica*, vol. 24, no. 1, pp. 12–18, 2014. https://doi.org/10.11613/BM.2014.003
[71] K.C. Asanya, et al., "Robust Bayesian Approach to Logistic Regression Modeling in Small Sample Size Utilizing a Weakly Informative Student's T Prior Distribution," *Communications in Statistics – Theory and Methods*, 2021. https://doi.org/10.1080/03610926.2021.1912767
[72] T. Whitaker, B. Beranger, and S.A. Sisson, "Logistic Regression Models for Aggregated Data," *Journal of Computational and Graphical Statistic*, 2021. https://doi.org/10.1080/10618600.2021.1895816
[73] B.M. Khammas, "Ransomware Detection using Random Forest Technique," *ICT Express*, vol. 6, pp. 325–331, 2020. https://doi.org/10.1016/j.icte.2020.11.001
[74] L. Weinholda, et al., "A Random Forest Approach for Bounded Outcome Variables," *Journal of Computational and Graphical Statistics*, vol. 29, no. 3, pp. 639–658, 2020. https://doi.org/10.1080/10618600.2019.1705310
[75] A. Upadhyay, M. Singh, and V.K. Yadav, "Improvised Number Identification using SVM and Random Forest Classifiers," *Journal of Information and Optimization Sciences*, vol. 41, no. 2, pp. 387–394, 2020. https://doi.org/10.1080/02522667.2020.1723934
[76] C.Y. Lin, "Forward Stepwise Random Forest Analysis for Experimental Designs," *Journal of Quality Technology*, 2021. https://doi.org/10.1080/00224065.2020.1865853
[77] H. Wang and G. Wang, "Improving Random Forest Algorithm by Lasso Method," *Journal of Statistical Computation and Simulation*, vol. 91, no. 2, pp. 353–367, 2020. https://doi.org/10.1080/00949655.2020.1814776
[78] A.A. Selçuk, "A Guide for Systematic Reviews: PRISMA," *Turk Arch Otorhinolaryngol*, vol. 57, no. 1, pp. 57–58, 2019. https://doi.org/10.5152/tao.2019.4058
[79] Z. Ma, et al., "Ensemble of Machine Learning Algorithms Using the Stacked Generalization Approach to Estimate the Warfarin Dose," *PLOS ONE*, pp. 1–12, 2018. https://doi.org/10.1371/journal.pone.0205872
[80] K.S. Shultz and D.A. Olson, "The Changing Nature of Work and Retirement," in *The Oxford Handbook of Retirement*, M. Wang, Ed. New York: Oxford University Press, 2012, pp. 543–558. https://doi.org/10.1093/oxfordhb/9780199746521.013.0162
[81] X. Fang, "Making Recommendations Using Transfer Learning," *Neural Computing and Applications*, vol. 33, pp. 9663–9676, 2021. https://doi.org/10.1007/s00521-021-05730-3
[82] G. Yang, "Research on Music Content Recognition and Recommendation Technology Based on Deep Learning," *Security and Communication Networks*, vol. 2022, 2022. https://doi.org/10.1155/2022/7696840
[83] H. Ko, et al., "A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields," *Electronics*, vol. 11, no. 1, 141, 2022. https://doi.org/10.3390/electronics11010141

[84] W. Mainga, R. Daniel, and L. Alamil, "Perceptions of Employability Skills of Undergraduate Business Students in a Developing Country: An Exploratory Study," *Higher Learning Research Communications*, vol. 12, no. 1, pp. 28–63, 2022. https://doi.org/10.18870/hlrc.2022.v12i1.1257

[85] H.N. Mpia, S.N. Mwendia, L.W. Mburu. "Predicting Employability of Congolese Information Technology Graduates Using Contextual Factors: Towards Sustainable Employability," *Sustainability,* vol. 14, no. 20, 13001, 2022. https://doi.org/10.3390/su142013001

[86] A.A. AlMunifi and A.Y. Aleryani, "Knowledge and Skills Level of Graduate Civil Engineers Employers and Graduates' Perceptions," *International Journal of Engineering Pedagogy (iJEP)*, vol. 9, no. 1, pp. 84–101, 2019. https://doi.org/10.3991/ijep.v9i1.9744

11 AUTHORS

Héritier Nsenge Mpia is currently a Ph.D. candidate in Information Systems at KCA University, Nairobi, Kenya. His main research interests include Artificial Intelligence, Deep Learning, Data Science, and application of Artificial Intelligence in Education. He has a BSc in Mathematics and Computing and MSc in Information Systems Design. (email: 2004988@students.kcau.ac.ke). ORCID: https://orcid.org/0000-0001-7428-8092.

Lucy Waruguru Mburu is Senior Lecturer and the Chair of the Department of Networks and Applied Computing, School of Technology, KCA University. She has a Ph.D. in Geoinformatics from the University of Heidelberg, Germany. Her main research interests include Geo-computation, Information management, Geographic Information Systems, Inter-disciplinary research (email: mburul@kcau.ac.ke). See https://sot.kcau.ac.ke/personnel/dr-lucy-w-mburu.

Simon Nyaga Mwendia is Senior Lecturer and the Dean of the School of Technology at KCA University, Nairobi, Kenya. He has a Ph.D. in Computer Science from the University of Nairobi. His main research interests include Ambient learning in Higher-learning institutions, Applications of Data Science, and Mobile applications in Education. (email: smwendia@kcau.ac.ke). See https://sot.kcau.ac.ke/personnel/dr-simon-nyaga-mwendia.