International Journal on Advanced Science Engineering Information Technology

Vol.14 (2024) No. 1
ISSN: 2088-5334

Classification Techniques Using Machine Learning for Graduate Student Employability Predictions

Radiah Haque [a], Albert Quek [a], Choo-Yee Ting [a], Hui-Ngo Goh [a,*], Md Rakibul Hasan [a]

[a] *Faculty of Computing and Informatics, Multimedia University, Cyberjaya, 63100, Malaysia*

[*] *Corresponding author: hngoh@mmu.edu.my*

***Abstract*—The issue of employability has gained significant importance, not only for graduate students but also for higher educational institutions. In this regard, employability prediction models using machine learning have emerged as crucial techniques for assessing students' potential to secure employment after graduation. Enhancing university graduate employability is critical because student unemployment is a global concern that has widespread negative effects on both individuals and institutions. Therefore, focusing on graduate employability predictions using machine learning techniques is considered essential in addressing this issue. Traditionally, demographic and academic attributes, such as CGPA, have been considered key factors in determining student employment status. However, research suggests that various other factors, such as student satisfaction, might influence employability. This study employs machine learning techniques to identify the factors that affect graduate student employability. The objective is to investigate the features significantly influencing students' ability to secure employment. Data was collected from Malaysia's Ministry of Education's graduate tracer study (SKPG). Several classification algorithms were applied, including Logistic Regression, Random Forest, Naïve Bayes, Support Vector Machine, Extreme Gradient Boosting, and Artificial Neural Networks (ANN). The results show that ANN achieved the highest accuracy, with around 80%. The findings also revealed that student demographic and academic features and student satisfaction level with the university facilities (e.g., library and counseling service) are considered significant for graduate student employability predictions. Consequently, the empirical results can help higher educational institutions enhance facilities and prepare students with the necessary skills for future employability.**

***Keywords*— Student employability prediction; feature selection; classification; machine learning.**

Manuscript received 15 Jan. 2023; revised 12 Sep. 2023; accepted 12 Dec. 2023. Date of publication 29 Feb. 2024.

I. INTRODUCTION

Higher education institutions are crucial in boosting a nation's economy as they form an industry and support other industries by providing a skilled workforce [1]. Previously, most universities were primarily concerned with declining student success rates, low student retention, high student attrition to competing universities, and inadequate counseling in subject selection [2]. However, as education becomes increasingly focused on employment, the employability of graduates has become a significant factor in shaping an institution's reputation and a central area of concern. In this regard, it is believed that a high percentage of unemployed graduates can have negative consequences for society, resulting in economic costs. This includes a decrease in productivity and an unsatisfactory return on the investments made in higher education institutions by individuals and the government [3].

In recent years, the issue of unemployment, particularly among university graduates, has gained significant attention and is widely regarded as a persistent crisis [4]. In Malaysia, there has been a significant rise in the unemployment rate among recent graduates, reaching 25% in 2020. This upward trend in graduate unemployment has been observed consistently, primarily due to the challenging economic conditions in the country [5]. The Malaysia Ministry of Education has implemented various measures to address these challenges and enhance employability. These include curriculum revisions, the promotion of entrepreneurship courses, and a focus on developing essential skills and competencies such as English language proficiency, teamwork, and analytical skills [6]. Additionally, fostering successful partnerships between universities, industries, and the government can effectively showcase graduates' skills to potential employers in the industry, thereby benefiting the graduates [7].

As such, graduate student employability holds great importance for universities as it serves as a critical indicator of their effectiveness. However, as the job market is undergoing significant changes driven by globalization, automation, and advancements in artificial intelligence, it is essential to identify the crucial factors that influence employability and understand the evolving requirements of the job market to benefit all parties involved. Students can plan their career paths more effectively by recognizing their strengths and weaknesses. Meanwhile, universities can focus on enhancing the relevant facilities to cultivate new competencies, ensuring that current and future students can learn and acquire the required skill sets that align with the rapidly changing labor markets [8]. Through these collective efforts, the overall employability of university graduate students can be significantly enhanced. More specifically, predicting graduate student employability status can enable educational organizations to identify students who may be aware of the challenges in finding employment. This knowledge can serve as a means for universities to devise strategies and interventions to motivate students to improve their career prospects [4].

Previously, data-driven and machine-learning techniques have been widely applied in different areas of educational data mining. Over the last decade, researchers have been exploring the use of machine learning and data mining techniques to predict employability. The majority of the existing studies exhibit considerable variation in terms of the data utilized, the methods employed, and even the research questions posed [9]. Such questions on student employability predictions include: what type of data can be collected and mined for student employability prediction? What machine learning algorithms that are effective for student employability predictions can be used? Overall, research in employability prediction is still in its infancy. Existing research primarily focuses on identifying the skills and attributes sought by employers, typically obtained through questionnaires and interviews [4], [10], [11]. Most studies have employed statistical methods that are descriptive rather than predictive in nature. It is comprehensible that limited progress has been made in employability prediction, largely due to a lack of authentic and comprehensive data. The concept of graduate employability prediction presents challenges in defining and measuring it accurately, leading to a paucity of studies examining its predictors and outcomes. Consequently, the work in this paper on graduate student employability prediction and classification model development will significantly contribute to the field of educational machine learning.

In machine learning, predicting student employability is considered an iterative process that includes gathering relevant data, cleaning and preparing the data, constructing models, validating them, and deploying the models for prediction [12]. Rather than considering all available student attributes as features, the prediction model focuses on selecting an optimal set (or a combination sets) of features that contribute to improving prediction performance in terms of accuracy [4]. Generally, educational institutions gather and store vast amounts of data, including students' academic records, personal profiles, behavioral observations, and faculty profiles. This data represents a valuable source of information that must be explored to gain a strategic advantage among educational organizations. Predicting student employability with relevant attributes can help identify students at risk of unemployment, enabling universities to intervene promptly and take necessary steps to enhance their performance. In this regard, three primary explanations for unemployment among fresh graduates have been identified: 64% are poor in academic performance, 60% are poor in interpersonal skills, and 59% are poor in emotional competence and satisfaction [13].

In most studies, researchers have tried to establish links between demographic profiles and students' academic performance and skills with their employability [2,14,4,15]. Moreover, previous research has suggested that emotional competence plays a significant role in enhancing graduate employability. Psychological studies have demonstrated that a students' emotional and satisfaction levels are influential factors in predicting their employability status. However, there is a lack of extensive research in this area to validate existing knowledge or generate new insights regarding emotional skills and student satisfaction levels as predictors of performance. One primary reason for this limited exploration is the scarcity of authentic data [16]. Collecting authentic primary data encompassing relevant factors such as demographic profile, academic excellence, and student satisfaction level is crucial.

The unique contribution of this paper is that, apart from the student demographic and academic parameters, it also explores the link between student satisfaction level and facilities provided by the universities (e.g., lecture room, library, and counseling service) to predict employability using machine learning techniques. This paper focuses on identifying the factors that affect university graduate students' employability and comparing the classification techniques. The rest of the paper is organized as follows: The background study examines previous research in the field. The method section outlines the data collection process, data preprocessing, and the development of the dataset. This is followed by the results and discussion section, which compares the performance of various classifiers in predicting employability. The final section concludes the paper by summarizing key findings and suggesting potential opportunities for future research.

A. Background of the Study

Machine learning is a scientific field that uses statistical models and algorithms to perform tasks by systems without explicit instructions, relying on inference and patterns instead [17]. Consequently, machine learning operates within systems capable of identifying and comprehending data patterns, utilizing them to make autonomous decisions [18]. Since the last couple of decades, machine learning techniques have been majorly used for student employability predictions. For example, a three-scale categorization of output classes was employed in an early study by Sapaat [19], which focused on constructing a graduate employability model using machine learning classification techniques. The study utilized data obtained from the tracer study, a web-based survey system administered by the Ministry of Higher Education, Malaysia, in 2009. Information gain was employed to rank attributes, revealing that the job sector, job status, and reasons for unemployment were the top three attributes directly impacting employability. The study employed Naïve Bayes (NB) classifier, Decision Tree (DT), and Random Forest (RF) to classify graduates into three classes: employed, unemployed, or in an undetermined situation. The results demonstrated that a variant of the DT algorithm achieved the highest accuracy, outperforming the other classifiers. The current study also employs a dataset obtained from the tracer study in Malaysia, which was conducted in 2021, and includes additional relevant attributes, such as student satisfaction.

In another early study by Jantawan and Tsai [20], a prediction model was developed to forecast the employability of graduates using a three-scale categorization: employable, unemployable, and unpredictable. The model was validated using real data collected from graduate profiles at Maejo University in Thailand, comprising 11,853 instances over three academic years. The study aimed to build a graduate employment model utilizing a classification technique and compare various machine learning algorithms, including NB and RF. In the same year, Hugo [21] conducted a study to determine how undergraduate student academic and experiential employability attributes, such as major, General Point Average (GPA), co-curricular activities, and internships, can predict whether a student secures full-time employment prior to graduation. The study employed widely recognized and advanced machine learning models to predict employment before graduation, including RF, Artificial Neural Network (ANN), and Logistic Regression (LR). The results showed that employment before graduation can be predicted with 73% accuracy, with the ANN model yielding the highest accuracy. Additionally, a sensitivity analysis revealed that linguistic abilities and trimester GPAs were statistically significant variables in predicting employment upon graduation. Both attributes are considered in the current study for graduate student employability prediction.

In a work conducted by Piad et al. [22], the employability of IT graduates was predicted by considering various demographic factors, student performance, and professional variables. The data for this study was collected from the five-year profiles of 515 students randomly selected from the placement office tracer study. Different classification algorithms were applied, where LR achieved an accuracy of 78%. Meanwhile, Mishra et al. [2] studied student employability prediction using various classification models, including DT, RF, NB classifier, and ANN. They found that RF algorithm was the most suitable for predictions of employability, with an accuracy of around 71%. The latter study mainly identified the factors that impact employability, where they found that certain emotional skill attributes, such as empathy and stress management skills, significantly affect job placement. However, it should be noted that these emotional attributes do not measure student satisfaction levels.

The study by Aziz and Yusof [23] proposed a machine learning technique for classifying graduate employability, specifically focusing on determining whether graduates were employed or unemployed based on data obtained from MARA Professional College Malaysia. They employed five different classification models, namely LR, NB, RF, and ANN to predict the employability status of graduates. The results demonstrated that LR was the preferred choice for accurate graduate employability classification. Another study conducted in Malaysia [13] focused on proposing a classification model for predicting and assessing the attributes of student datasets to meet the industry's selection criteria for graduates in the academic field. The study investigated different bachelor's degree programs, gender, and Cumulative GPA (CGPA) of Malaysian university students. Various supervised machine learning algorithms, including NB, RF, LR, SVM, and ANN, were employed, and it was revealed that the latter algorithm achieved the highest accuracy score. The proposed model aimed to assist university management in developing long-term plans for producing graduates who possess the necessary skills and knowledge required by the industry. Nevertheless, additional attributes, such as grades for the common subjects taken during the study period, need to be assessed to determine their impact on student employability.

Another study was conducted by Thakar et al. [24] to develop an employability prediction model using a dataset of master's in computer applications students from various colleges of a State University in Delhi, India. To build the predictive model, the researchers integrated clustering and classification techniques. At the preprocessing stage, a two-level clustering approach (k-means kernel) with Chi-square analysis was applied to select relevant attributes automatically. Subsequently, an ensemble vote classification technique was employed using RF. This ensemble model was used to predict students' employability. The study found that including supplementary attributes such as personal, social, cognitive, and environmental variables can improve the prediction of students' employability.

Nonetheless, the previous study focused on academic attributes, while psychometric attributes were considered only secondary. Applying machine learning and data mining techniques to determine the factors that influence graduate employability, a study [7] employed a seven-year dataset (2011-2017). The dataset was collected from the Malaysia Ministry of Education tracer study, consisting of a total of 43,863 data instances for developing the employability classification model. The study identified several factors that affect graduate employability, including faculty, field of study, CGPA, marital status, and English language skills, all of which are considered academic attributes. Three classification algorithms, namely DT, SVM, and ANN were employed and compared to identify the best models. The results indicate that the proposed tree-based prediction model achieved an accuracy of 66%, which is comparatively a lower accuracy rate than the other existing predictive models in the literature.

A study [25] collected data from students attending various Engineering colleges in Hyderabad to achieve high employability prediction accuracy. Supervised machine learning algorithms, including DT, SVM, and NB, were applied to predict the students' employability and identify the factors influencing it. The results revealed that SVM outperformed the other classifiers in predicting student employability, yielding 98% accuracy. Meanwhile, Casuat [26] conducted a study aiming to utilize machine learning approaches for predicting students' employability. The researcher did a case study involving 9 features, which include mock job interview evaluation results, On-the-Job

Training (OJT), student performance rating, and GPA of students enrolled in the OJT course for three years. The study employed six learning algorithms: RF, LR, NB, SVM, ANN, and Extreme Gradient Boosting (XGB) to gain insights into students' employability. The performance of these algorithms was evaluated using accuracy measures, precision and recall measures, f1-score, and support measures. In the experiments, SVM achieved an accuracy measure of 91%, which was significantly higher than the other algorithms, with DT 85% and RF 84%. It is worth noting that the datasets used in these studies are comparatively small and consist of a limited number of attributes. However, the dataset used in the current study contains different types of attributes (e.g., demographic attributes, academic attributes, and student satisfaction level), where identifying the most suitable features (or subset of features) is critical for achieving a high prediction accuracy.

The study conducted by Bai and Hira [4] introduced a hybrid model using ANN and Softmax regression for predicting student employability. To improve the accuracy of the prediction model, a feature selection model based on the crow search algorithm was utilized. This feature selection model helped to identify the optimal subset of features from the original set, which significantly contributes to the prediction of student employability. The optimal features selected were selected as inputs for the ANN model, which enabled the learning of intrinsic features to capture high-level representations. Subsequently, the Softmax regression was employed to predict whether students would be employed or unemployed. A statistical simulation analysis for the proposed prediction model was conducted in MATLAB using a dataset collected through questionnaires, which includes basic academic attributes (e.g., CGPA) and intellectual attributes of students (e.g., programming skill).

In 2021, a study [27] constructed a detailed academic dataset that included student performance data and subject test scores. These data were then used for classification purposes, specifically employing a DT classifier to predict the employability of students across various disciplines. A more recent study conducted in 2022 [28] aimed to develop a supervised machine learning (SML) model for predicting the employability of graduates based on their academic scores and fields of study. The study utilized the DT algorithm to construct the SML model. With 65% accuracy, the model successfully predicted the likelihood of placement based on students' academic scores and fields of study. In a similar vein, the study by PS and Abraham [29] focused on how academic and co-academic abilities of students, faculty characteristics, and teaching practices contribute to student learning. The dataset used in the study was collected from the same institution for which placement predictions were to be made. The study formulated a sequential event prediction problem and employed ANN and deep learning algorithms. The proposed model utilized a dataset with 18 attributes to assess the performance of lower-level and higher-order skills, providing methods for enhancing a student's chances of securing full-time employment. The findings of this study revealed that, as universities face increasing accountability for students' career outcomes and job competition intensifies, it is crucial for institutions to understand which students are more likely to be employed upon graduation and the factors influencing their employability.

Maaliw et al. [30] compared multiple classification algorithms to create an ensemble prediction model for forecasting graduates' employability using machine learning techniques. The evaluation of various metrics determined that an ensemble model consisting of RF, SVM, and NB achieved the highest cross-validated accuracy score. This indicates the effectiveness of the ensemble model in accurately predicting graduates' employability. Another recent work by Saidani et al. [31] presents an effective approach for predicting student employability by employing XGB classifier. The main contribution of this work involved leveraging the capabilities of gradient boosting algorithms to conduct context-aware predictions of employability status. This is achieved by considering both the student context, which includes student features, and the internship context, which encompasses internship-related features. The results indicate that employing XGB with the internship context yielded the best performance compared to the student context. This suggests that the employability of graduates can be effectively predicted by considering the information derived from non-academic context, such as previous employment experience of graduate students. Table 1 summarizes the studies reviewed in this section along with the machine learning classification models applied in each study.

TABLE I
SUMMARY OF THE CLASSIFICATION MODELS

Ref.	DT	RF	NB	LR	SVM	ANN	XGB
[2]	X	X	X			X	
[4]						X	
[7]	X				X	X	
[13]		X	X	X	X	X	X
[19]	X	X	X				
[20]		X	X				
[21]		X		X		X	
[22]		X	X	X	X	X	
[23]		X	X	X		X	
[24]		X					
[25]	X		X		X		
[26]		X	X	X	X	X	X
[27]	X						
[28]	X						
[29]						X	
[30]		X	X		X		
[31]							X

Overall, the key limitation of existing research on university graduate student employability prediction is concerned with data relevance and limited scope of the machine learning models. The proposed employability models mostly focus on a specific set of attributes or factors influencing employability. This limited scope may not account for all relevant aspects of increasing the employability of university graduate students. Furthermore, the accuracy and completeness of the data used for constructing the employability model can significantly impact its validity. In this case, if the data used is outdated or contains incomplete or irrelevant features, it may affect the reliability of the model's predictions and generalizability. Having said, not all attributes would have the same level of impact in terms of increasing the accuracy of the models. As a result, predictive modelling can be affected by the selection of optimal features (or subset of features) used in the training process. If important attributes that influence employability are omitted or if irrelevant features are used, the resulting model may not accurately capture the complexities of graduate student employability predictions. In this regard, while academic performance has traditionally been an important criterion for assessing students' efforts and dedication, many companies now consider other factors such as emotional attributes and student satisfaction when considering potential employees, especially fresh graduates who have inadequate knowledge and required technical skills [2]. Therefore, apart from the student demographic and academic parameters, the current study in this paper also explores the link between student satisfaction level with university facilities (e.g., lecture room, library, and counseling service) using common machine learning techniques for student employability prediction.

II. Material and Method

The research in this paper consists of three phases. The first phase is concerned with collecting data from relevant resources. The second phase includes data preprocessing and constructing structured datasets for data analyses and pattern recognition. In the final phase of this research, various supervised machine-learning algorithms are employed for binary classification.

A. Data Collection and Preprocessing

This study collected data from the Ministry of Education's graduate tracer study in Malaysia (SKPG) conducted in 2021. In particular, data obtained from the graduates of Multimedia University (MMU) was employed as a case study. Each graduate student was required to conduct a survey near their convocation day where they answered multiple choice questions as a form of structured questionnaire and provided their student ID. This ID was used to extract their demographic and academic information stored in the university database. However, each student ID was masked and replaced with a mock ID before the acquisition of full dataset for experimental work and predictive analytics.

In this work, **data integration** was the first step during data preprocessing, a technique used to combine sets of data and information from different sources. In effect, data integration has been carried out as the data was collected and obtained from different datasets. A sample of 3000 student data was gathered and stored in a Comma Separated Variables (CSV) format file, which contains several attributes that have been categorized into five types: (i) demographic attributes, (ii) academic attributes, (iii) GPA attributes, (iv) subject grade attributes, and (v) SKPG attributes.

Demographic attributes include students' demographic information such as address, date of birth, nationality, and gender. Academic attributes include information about their faculty and study program at the university, academic performance (e.g., *CGPA*), and if they have Graduated on Time (*GOT*). It should be noted that, in this study, GPAs achieved in each trimester are separated from the academic attributes and categorized under GPA attributes. A similar approach is taken with the grade achieved in Sijil Pelajaran Malaysia (SPM) and language subjects and categorized under subject grade attributes. This was important to identify if these attributes contribute to student employability. Meanwhile, SKPG attributes include the questionnaire data, such as student employability status, sponsorship status, and their satisfaction level with university facilities (e.g., *lecture_room_facility*, *library_facility*, and *counselling_service*).

Data cleaning was the second step of data preprocessing performed on the student dataset to remove irrelevant and null values to achieve data consistency. During the data cleaning process, the first task was to remove or modify data errors, inconsistent data, missing data, and overlapping records. The second task was to remove irrelevant records. In this study, only undergraduate student records were kept, while data obtained from other levels of studies such as Diploma, Ph.D., and Masters have been removed. This process reduced the dataset to 2375 records. The third step of data preprocessing was to perform **data transformation**. Some attributes have been transformed into categories, such as *CGPA* and *GPA*. Originally, these attributes were numerical (e.g., *3.6*), but for the work in this paper, they were transformed into categorical ranges (e.g., *CGPA = 3.6* was replaced with a grade range *3.50-4.00*). Moreover, some attribute values, such as nationality, were modified to reduce data imbalance. Originally, nationality attributes included different country names. However, for this work, only two values were kept (*Malaysian* and *Non-Malaysian*).

The fourth step of data preprocessing was to **eliminate unnecessary attributes**, such as *Student_mock_ID*. From student demographic category, attributes including *DoB*, *Degree_code*, and *Graduation_year* were removed. From the subject grade category, attributes including *Subject_ID* and *Subject_code* were eliminated. Additionally, only core subjects such as mathematics, physics, and chemistry were kept in the dataset. Originally in GPA category, data until trimester 15 was obtained. Nevertheless, data for trimesters 1 to 12 was kept for this study because most programs required 12 trimesters. Originally the attribute *Employment_status* from SKPG category consisted of multiple values, including employed, unemployed, further study, and own/family business. In this work, the latter two values were considered as employed. Consequently, *Employment_status* attribute was considered as the target output with two classes: *Employed*, which denotes to students who are currently employed (or employed after graduation) in a company, doing further studies, or have business, and *Unemployed*, which denotes to students who did not get employed at the time of data collection. The final dataset consists of 62 attributes (61 input features and 1 target output), as can be seen in Table 2.

TABLE II

DATASET ATTRIBUTES

	Attribute	Value	Description
		Demographic Attributes	
1	State	Kuala Lumpur, Selangor, Johor, Melaka, and the rest of the states	Permanent address state of student in Malaysia
2	Nationality	Malaysian, Non-Malaysian	Nationality of student
3	Race	Malay, Chinese, Indian, and others	Race of student
4	Gender	Male, Female	Gender of student
5	Disability	Yes, No	If student has any disability
6	Status	Single, Married, Others	Marital status of student
		Academic Attributes	
1	Program	B.S.C, B.B.A., L.L.B., B.I.T., B.E., and the rest of the programs	Study program of students
2	Faculty	FCI, FOB, FOM, FET, FOL, …etc.	Study program faculty
3	GOT	Yes, No	If student graduated on time
4	CGPA	2.00-2.49, 2.50-2.99, 3.00-3.49, 3.50-4.00	Cumulative GPA
5	Class_of_ Honours	First class, Second class (Upper), Second class (Lower), Third class	Class of honors degree
6	Credit_ transfer	Yes, No	If student transferred any credit
7	BM_Score	A+, A, A-, B+, B, C+, C, D, E, Not applicable	Score for Malay language test
8	BI_Score		Score for English language proficiency
9	Muet_Score	Band 3, Band 4, Band 5, Band 6, Not applicable	Score for Malaysian university English language test
		GPA Attributes	
1	T1_GPA		GPA for trimester 1
2	T2_GPA		GPA for trimester 2
3	T3_GPA		GPA for trimester 3
4	T4_GPA		GPA for trimester 4
5	T5_GPA	0-1.99, 2.00-2.49, 2.50-2.99, 3.00-3.49, 3.50-4.00, Not applicable	GPA for trimester 5
6	T6_GPA		GPA for trimester 6
7	T7_GPA		GPA for trimester 7
8	T8_GPA	0-1.99, 2.00-2.49, 2.50-2.99, 3.00-3.49, 3.50-4.00, Not applicable	GPA for trimester 8
9	T9_GPA		GPA for trimester 9
10	T10_GPA		GPA for trimester 10
11	T11_GPA		GPA for trimester 11
12	T12_GPA		GPA for trimester 12
		Subject Grade Attributes	
1	Matematik		Score for Mathematic course
2	Prin_Akaun		Score for accounting
3	Sains		Score for science
4	Sejarah		Score for history
5	Bahas_Cina		Score for Mandarin language
6	Matematik_Tambahan	A+, A, A-, B+, B, C+, C, D, E, Not applicable	Score for additional mathematics
7	Pendidikan_Moral		Score for moral studies
8	Biologi		Score for biology
9	Fizik		Score for physics
10	Kimia		Score for chemistry
		SKPG Attributes	
1	Entry_ eligibility	IPTS, STPM, Diploma, Others	University entry eligibility certificate
2	Sponsor_ category	Scholarship, Loan, Self-sponsored	If student had scholarship or loan for their study
3	Worked_ before	Yes, No	If student worked before their study
4	Part_time_ job		If student had part time job during study
5	Enter_job_market		Student satisfaction level with entering job market
6	Internship_ job		Student satisfaction level regarding if doing internship helps finding job
7	Interactive_ learning		Student satisfaction interactive learning
8	Innovative_ learning	Extremely satisfied, Satisfied, Moderate, Unsatisfied, Extremely unsatisfied	Student satisfaction innovative learning
9	Online_ interaction		Student satisfaction with online learning
10	Library_ facility		Student satisfaction level with library facility in university
11	Lab_facility		Student satisfaction level with lab facility
12	Lecture_		Student satisfaction with lecture room

	room		
13	Sport_ facility	Extremely satisfied, Satisfied, Moderate, Unsatisfied, Extremely unsatisfied	Student satisfaction with sport facility
14	Cafeteria_ facility		Student satisfaction with cafeteria facility
15	Accom_ facility		Student satisfaction with accommodation
16	Transport_ facility		Student satisfaction with transportation
17	Clinic_ facility		Student satisfaction with clinic facility
18	Parking_ facility		Student satisfaction with parking facility
19	Security_ facility		Student satisfaction with security facility
20	Counselling_service		Student satisfaction counselling service
21	Met_ counsellor	Yes, No	If student met counsellor
22	Counsellor_meeting	No service required, Shy on getting the service, Unsure the role of the unit/center	Reason for not meeting the counsellor in university
23	Positive_ change	Very high, High, Moderate, Low, Very low	If university is giving a positive change to student
24	University_ reputation	Very good, Good, Moderate, Poor, Very poor	Student perception of university reputation
25	Employed_ status	Employed, Unemployed	Target output for classification

In this work, the dataset was divided into 8 experimental datasets for predictive modelling and classification, as can be seen in Table 3.

TABLE III
EXPERIMENTAL DATASETS

Dataset	Description (Set of Attributes)	No. of Attributes
D_1	Contains student demographic and academic attributes	16
D_2	Contains demographic, academic, and GPA attributes	28
D_3	Contains demographic, academic, and subject grade attributes	26
D_4	Contains demographic, academic, GPA, and subject grade attributes	38
D_5	Contains demographic, academic, and SKPG attributes	40
D_6	Contains demographic, academic, SKPG, and GPA attributes	52
D_7	Contains demographic, academic, SKPG, and subject grade attributes	50
D_8	Contains demographic, academic, SKPG, GPA, and subject grade	62

The goal is to identify which attributes contribute the most to student employment prediction. From the literature, it was recognized that most student employability prediction models employ datasets containing student demography and academic attributes. Therefore, the first dataset D_1 consists of both demographic and academic attributes. The second dataset D_2 is a combination of D_1 and GPA attributes. D_2 was constructed to investigate if adding individual trimester results can enhance the performance of the predictive models applied in this study. Similarly, the third dataset D_3 was constructed using D1 attributes and subject grade attributes. The fourth dataset D_4 is a combination of the previous three datasets. The first four datasets did not include SKPG attributes. The fifth dataset D_5 consists of SKPG attributes, the original demographic, and academic attributes. The sixth D_6 and seventh D_7 datasets contain D_5 attributes, GPA attributes, and subject grade attributes, respectively. The final dataset D_8 consists of all 62 attributes.

B. Classification Models

The main objective of the work in this paper is to apply classification models using machine learning algorithm. The goal is to employ experimental datasets to predict the employment status of the graduate students. The classification models in this study are built to classify the target output employed/unemployed based on various combinations of the input features. Several classification techniques are available, each having its own advantages and disadvantages. Attributes in the dataset are nominal thus machine learning algorithms that can handle nominal data were considered in this study, including Extreme Gradient Boosting (XGB), Logistic Regression (LR), Naïve Bayes (NB), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANN).

Before making predictions, it is crucial to determine the most suitable algorithm for the problem. This necessitates a comparison of the mentioned algorithms using specific metrics. As such, in this paper, the prediction performance of the classification algorithms was evaluated and compared based on accuracy, precision, recall, and F1 Score. Moreover, Area under Receiver Operating Characteristic (ROC) score was calculated for each classification model. Typically, these measures are considered as the most comprehensive that can evaluate the classifiers' performance fairly. To calculate each metric score, it was important to map out the confusion matrix for each model, which is generated to summarize the prediction performance made by each classifier. This is typically done by identifying the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) for each prediction class [32]. From the confusion matrix, equations 1-4 are used for calculating the evaluation metrics [33]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

III. Results and Discussion

This section analyzes and compares prediction performance results for the six classification models. Each classification model was trained eight times using experimental datasets for training and testing. 64 experimental models were built using the selected machine learning algorithms. The prediction performance of the classification models was evaluated based on accuracy, precision, recall, F1, and ROC scores. Normally, F1 score is calculated based on macro, micro, and weighted averages. In this study, F1-macro is considered due to the imbalance of output classes, as can be seen in Figure 1. In this case, macro scores can return an objective measure of predictive model performance when the output classes are imbalanced [34].

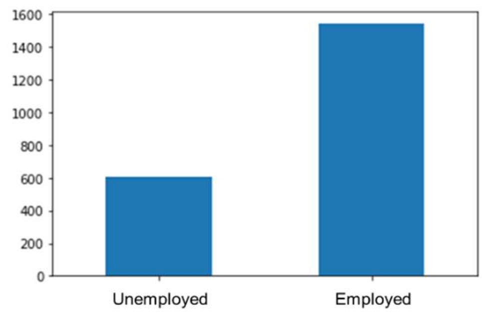

Fig. 1 Target output class distribution

Table 4 summarizes the performance results of XGB classification model. On average, the accuracy of the classifier is around 76%. However, while evaluating the performance of the model against each dataset, it is observable that the model achieved the highest accuracy on D_6 with an accuracy rate of 78%, precision = 0.8064, recall = 0.8959, F1-macro = 0.8426, and ROC = 0.7959. Similar prediction performance can be seen with LR in Table 5, where, on average, the accuracy of the model is around 76%. However, the model performed the best when applied on D_6 with an accuracy rate of 78%, precision = 0.7738, recall = 0.9432, F1-macro = 0.8458, and ROC = 0.7723. It should be noted that, in LR, recall is significantly higher than precision, which indicates that the model was able to predict most of the positive class values correctly. In this study, getting a high recall value is crucial to avoid misclassification. For instance, if employed students are classified as unemployed, it will be difficult to identify the relevant features affecting their employability. Besides, the findings from this paper will be used to build a recommender system for unemployed students. In this case, if unemployed students are classified as employed, it will be difficult for the system to recommend potential jobs and companies for them, resulting in a less effective system for student employability predictions.

NB also performed well in terms of recall, as can be seen in Table 6. Similar to LR, the average accuracy of NB classifier is around 76%. Again, the model achieved the highest accuracy on D_6 with an accuracy rate of 77%, precision = 0.8119, recall = 0.9905, F1-macro = 0.8509, and ROC = 0.6668. The ROC value of the latter model is significantly lower than the former models, tending to strong misclassification of classes with NB classifier. This issue of misclassification for the NB classifier was previously addressed [35] because the NB algorithm assumes independence by using Bayes theorem to calculate the probabilities of classes. In other words, when one output class has relatively small number of samples, the probability prediction for this minor class would be imprecise.

Table 7, 8, and 9 show that the average accuracy of RF, SVM, and ANN increased with 1% compared to the previous three classifiers. Once again, all these classifiers achieved the highest accuracy when trained with D_6. RF scored 79% accuracy, with precision = 0.8201, recall = 0.9874, F1-macro = 0.8599, and ROC = 0.7932. SVM scored 79% accuracy, with precision = 0.8171, recall = 0.9842, F1 score = 0.8635, and ROC = 0.7916. Meanwhile, ANN achieved a higher accuracy score than the previous models 80%, with precision = 0.8263, recall = 0.9761, and F1-macro = 0.8653. Interestingly, in ANN, both accuracy and ROC scores are exactly the same (i.e., 0.8021), indicating that the model performed well in predicting the output class and effectively handled imbalanced data. This result complies with the literature where it was recognized that neural networks are considered a popular method for classifying imbalanced data [36].

TABLE IV
Prediction Performance Results of XGB Model

Extreme Gradient Boosting (XGB)					
Dataset	Accuracy	Precision	Recall	F1	ROC
D_1	0.7321	0.7697	0.8328	0.8000	0.6928
D_2	0.7358	0.7600	0.8391	0.7976	0.7127
D_3	0.7337	0.7654	0.8233	0.7933	0.7127
D_4	0.7337	0.7623	0.8297	0.7946	0.7038
D_5	0.7747	0.7933	0.8801	0.8416	0.7684
<u>D_6</u>	<u>0.7789</u>	<u>0.8064</u>	<u>0.8959</u>	<u>0.8426</u>	<u>0.7959</u>
D_7	0.7748	0.8000	0.8707	0.8338	0.7819
D_8	0.7769	0.8029	0.8864	0.8415	0.7943
Average	0.7551	0.7825	0.8573	0.8181	0.7453

TABLE V
Prediction Performance Results of LR Model

Logistic Regression (LR)					
Dataset	Accuracy	Precision	Recall	F1	ROC
D_1	0.7453	0.7513	0.9243	0.8289	0.6934
D_2	0.7495	0.7605	0.9117	0.8293	0.7070
D_3	0.7389	0.7533	0.9054	0.8223	0.6857
D_4	0.7411	0.7608	0.8927	0.8215	0.7014
D_5	0.7684	0.7717	0.9274	0.8424	0.7681
<u>D_6</u>	<u>0.7705</u>	<u>0.7738</u>	<u>0.9432</u>	<u>0.8458</u>	<u>0.7723</u>
D_7	0.7621	0.7684	0.9180	0.8374	0.7557
D_8	0.7645	0.7698	0.9211	0.8379	0.7626
Average	0.7550	0.7637	0.9180	0.8332	0.7308

Figure 2 illustrates the performance of the classifiers against the experimental datasets in terms of prediction accuracy. It can be seen that all classification models have achieved lower accuracy when applied to D_1, D_2, D_3, and D_4. Meanwhile, the accuracy rate of each model has dramatically increased with D_5, D_6, D_7, and D_8. It should be noted that the latter four datasets consist of SKPG data, whereas the former datasets only include demographic and academic-based data. This indicates that SKPG attributes have helped increase the classification models' accuracy.

TABLE VI
PREDICTION PERFORMANCE RESULTS OF NB MODEL

Naïve Bayes (NB)					
Dataset	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC**
D_1	0.7453	0.7552	0.9148	0.8274	0.6643
D_2	0.7453	0.7552	0.9148	0.8274	0.6646
D_3	0.7495	0.7552	0.9243	0.8312	0.6615
D_4	0.7516	0.7558	0.9274	0.8329	0.6644
D_5	0.7663	0.8104	0.9851	0.8498	0.6548
D_6	0.7684	0.8119	0.9905	0.8509	0.6668
D_7	0.7663	0.8104	0.9851	0.8498	0.6565
D_8	0.7672	0.8119	0.9905	0.8509	0.6643
Average	0.7575	0.7833	0.9541	0.8400	0.6622

TABLE VII
PREDICTION PERFORMANCE RESULTS OF RF MODEL

Random Forest (RF)					
Dataset	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC**
D1	0.7558	0.7571	0.9401	0.8371	0.7153
D2	0.7537	0.7578	0.9464	0.8368	0.7252
D3	0.7558	0.7571	0.9401	0.8371	0.7004
D4	0.7537	0.7578	0.9464	0.8368	0.7159
D5	0.7754	0.8121	0.9765	0.8496	0.7914
D6	0.7853	0.8201	0.9874	0.8599	0.7932
D7	0.7754	0.8121	0.9765	0.8496	0.7867
D8	0.7811	0.8201	0.9874	0.8599	0.7953
Average	0.7670	0.7868	0.9626	0.8459	0.7529

TABLE VIII
PREDICTION PERFORMANCE RESULTS OF SVM MODEL

Extreme Gradient Boosting (SVM)					
Dataset	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC**
D1	0.7474	0.7494	0.9338	0.8315	0.7075
D2	0.7474	0.7526	0.9495	0.8338	0.7133
D3	0.7537	0.7538	0.9369	0.8354	0.7058
D4	0.7516	0.7590	0.9527	0.8366	0.7134
D5	0.7832	0.8149	0.9779	0.8583	0.7680
D6	0.7937	0.8171	0.9842	0.8635	0.7916
D7	0.7832	0.8109	0.9748	0.8579	0.7575
D8	0.7895	0.8115	0.9811	0.8607	0.7758
Average	0.7687	0.7837	0.9614	0.8472	0.7416

Table 10 shows that, among these four datasets (i.e., $D_5 \ldots D_8$), the models provided prediction with the lowest average accuracy rate with D_7, which consists of demographic, academic, subject grade, and SKPG features. The next lowest average accuracy rate was attained with D8, which consists of demographic, academic, GPA, subject grade, and SKPG features. Both of these datasets contain subject grade attributes. This revealed that these features are causing a decreased classification accuracy rate. On the other hand, all classifiers have performed the best with D_6, including demographic, general academic, GPA, and SKPG attributes. The next best accuracy was achieved by the classifiers on D_5, which consists of similar features except for GPA features.

TABLE IX
PREDICTION PERFORMANCE RESULTS OF ANN MODEL

Extreme Gradient Boosting (ANN)					
Dataset	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC**
D_1	0.7489	0.7494	0.9148	0.8239	0.7115
D_2	0.7432	0.7614	0.8959	0.8232	0.7292
D_3	0.7468	0.7595	0.8864	0.8180	0.6972
D_4	0.7489	0.7533	0.9054	0.8223	0.7140
D_5	0.7897	0.8056	0.9590	0.8551	0.7795
D_6	0.8021	0.8263	0.9761	0.8653	0.8021
D_7	0.7832	0.8116	0.9558	0.8547	0.7300
D_8	0.7865	0.8201	0.9653	0.8547	0.7766
Average	0.7687	0.7859	0.9323	0.8397	0.7425

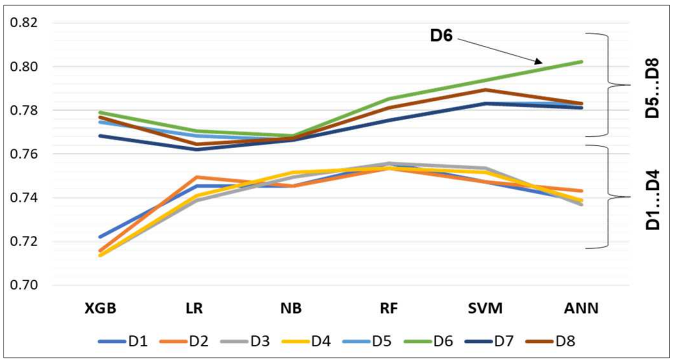

Fig. 2 Comparison of accuracy for all experimental datasets

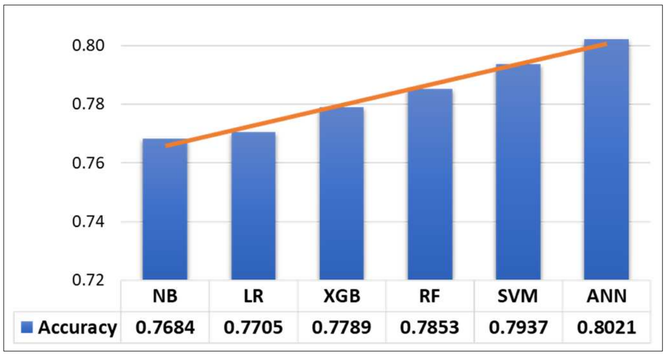

Fig. 3 Comparison of accuracy for D_6

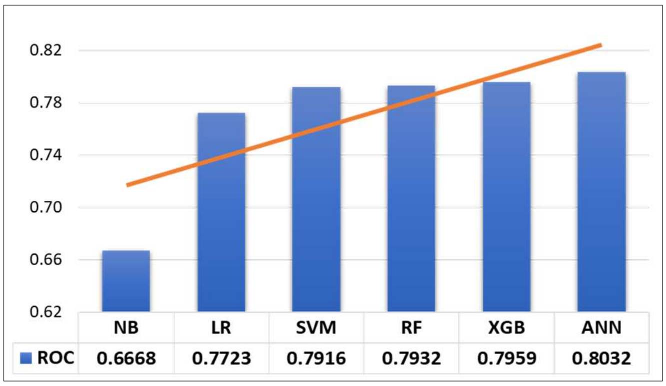

Fig. 4 Comparison of ROC for D_6

Subsequently, the prediction performance of the classification models applied on D_6 were compared. Figure 3 shows that ANN has achieved the highest accuracy rate (80%) among all classifiers applied in this work, followed by SVM and RF (79%), and XGB (78%). Meanwhile, LR and NB classifiers have provided predictions with the lowest accuracy rate (77%). Regarding ROC, ANN performed best, followed by XGB, RF, and SVM. Again, LR and NB classifiers gave the lowest ROC scores, as seen in Figures 4 and 5. Overall, the ANN algorithm obtained the best performance results for classifying student employability status. This result is comparable to the work conducted by Hugo [21], where they achieved the highest classification accuracy using ANN (73%). However, the model in the current study has achieved a higher accuracy rate, where Stochastic Gradient Descent (SGD) with momentum is used as the activation function.

TABLE X
AVERAGE ACCURACY OF ALL CLASSIFIERS

	XGB	LR	NB	RF	SVM	ANN
D_1	0.7321	0.7453	0.7453	0.7558	0.7474	0.7489
D_2	0.7358	0.7495	0.7453	0.7537	0.7474	0.7432
D_3	0.7337	0.7389	0.7495	0.7558	0.7537	0.7468
D_4	0.7337	0.7411	0.7516	0.7537	0.7516	0.7489
D_5	0.7747	0.7684	0.7663	0.7754	0.7832	0.7897
D_6	0.7789	0.7705	0.7684	0.7853	0.7937	<u>0.8021</u>
D_7	0.7748	0.7621	0.7663	0.7754	0.7832	0.7832
D_8	0.7769	0.7645	0.7672	0.7811	0.7895	0.7865

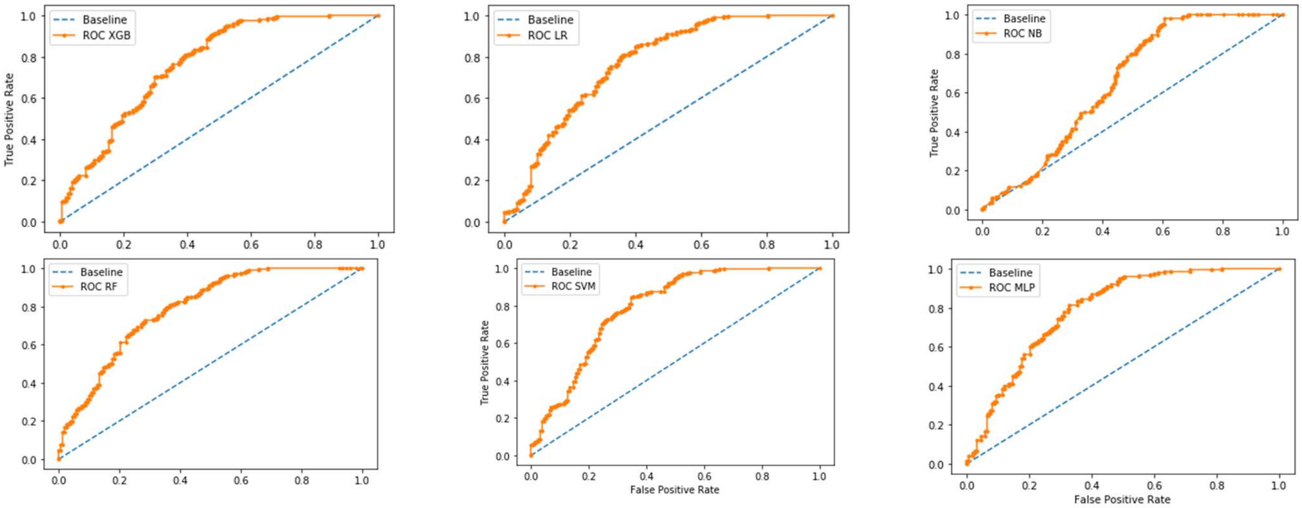

Fig. 5 ROC curve of all machine learning classifiers

This research's key finding is identifying relevant features that contribute to student employability and unemployability. Here, it was revealed that student demographic information, academic information, and student satisfaction level are important features for student employability predictions. It should be noted that along with the overall academic performance of students (i.e., CGPA), specific academic performance in each trimester (i.e., GPA) is considered critical for predicting their employability status. This finding complies with the literature where almost all studies revealed that the most commonly used predictors of student employability are demographic profile and the academic result of the students [2], [14], [4], [15]. Nonetheless, the findings from the current study reveal that individual subject grades might not be highly significant for student employability predictions. This is probably because the selected subjects in this study were generally taught subjects (e.g., mathematics, physics, and chemistry), that might not contribute to student employability in their specific fields. This case could differ for course-based subjects; thus, this result cannot be generalized on all taught subject grade attributes.

Meanwhile, previous studies have conducted empirical analysis to investigate the impact of employability on student life satisfaction [37] or to study the relationship between satisfaction level at the workplace and student employability [38]. However, few studies have examined the impact of student satisfaction levels with university facilities on their employability. Therefore, this is considered the main contribution of this research, where the findings revealed that student satisfaction with university facilities (e.g., library and counseling services) is important for their future employability. This experimental result can be justified by the study's findings conducted by [39], who revealed a relationship between student satisfaction with university facilities and class attendance. More specifically, in their study, optimism was found to predict whether students would be interested in starting a business after graduation, while networking on campus was found to predict whether students wished to pursue a management career. This indicates that when students are satisfied with the facilities provided on campus, their attendance increases, resulting in enhanced practical, teamwork, and communication skills. All these skills have proven effective in increasing employability among students in the literature [38,40,41,42]. Having said that, there is a need to conduct further studies on feature selection to determine which facilities have a higher impact that can increase student employability during or after graduation. Such study can help universities to enhance their facilities and, thus, increase the employability rate of their graduates.

IV. Conclusion

The primary objective of this paper was to apply various classifiers to determine student employability and develop an employability model based on the most suitable classifier. While academic performance has traditionally been considered a significant feature for student employability predictions along with demographic attributes of students, it is important to identify other factors that might contribute to student employability. In this study, student satisfaction with university facilities (e.g., lecture room, library, and counseling service) was considered an important factor that can enhance university graduate student employability. The work in this study utilized machine learning techniques to predict the status of graduate students employed/unemployed. Several classification algorithms were applied and employed for comparative analysis, including LR, NB, RF, SVM, XGB, and ANN. The results indicate that ANN outperformed the other classifiers, achieving an accuracy rate of 80%. Furthermore, from the findings of this study, it was recognized that student demographic and academic performance attributes are insufficient for student employability predictions. Interestingly, all the applied models provided predictions with higher accuracy when SKPG data was included, indicating the importance of measuring student satisfaction during university life.

Further research is necessary to validate the effectiveness of the identified features and student employability predictors through feature selection methods. This can help to identify the specific factors that can enhance the future employability of fresh graduates. A critical limitation of this study is the relatively small size of the dataset. More data should be collected from different universities to validate the results for future work. Additionally, ensemble machine learning approaches can ensure higher accuracy in employability predictions.

Acknowledgment

The TM R&D Fund from Telekom Malaysia supports this project.

References

[1] M. H. Baffa, M. A. Miyim, and A. S. D. Dauda, "Machine learning for predicting students' employability," *UMYU Sci.*, vol. 2, no. 1, 2023, doi:10.56919/usci.2123_001.

[2] T. Mishra, D. Kumar, and S. Gupta, "Students' employability prediction model through data mining," *Int. J. Appl. Eng. Res.*, vol. 11, no. 4, 2017, doi: 10.17485/ijst/2017/v10i24/110791.

[3] R. Awang-Hashim, L. H. Eam, A. Zubairi, H. Yon, and O. Osman, "Estimating a prediction model for the early identification of low employability graduates in Malaysia," *Singapore Econ. Rev.*, vol. 60, no. 4, 2015, doi: 10.1142/S0217590815500277.

[4] A. Bai and S. Hira, "An intelligent hybrid deep belief network model for predicting students employability," *Soft Comput.*, vol. 25, no. 14, 2021, doi: 10.1007/s00500-021-05850-x.

[5] G. S. Nor Azilah Husin, Salina Rasli, Ganesh Kumar, "Unemployment crisis among fresh graduates," *Am. Int. J. Soc. Sci. Res.*, vol. 10, no. 1, 2021, doi: 10.46281/aijssr.v10i1.1461.

[6] N. Rahmat, A. R. Ayub, and Y. Buntat, "Employability skills constructs as job performance predictors for Malaysian polytechnic graduates : A qualitative study," *Malaysian J. Soc. Sp. 12*, vol. 12, no. 3, 2016, [Online]. Available: http://journalarticle.ukm.my/9908/.

[7] Z. Othman, S. W. Shan, I. Yusoff, and C. P. Kee, "Classification techniques for predicting graduate employability," *Int. J. Adv. Sci. Eng. Inf. Technology*, vol. 8, no. 4–2, 2018, doi: 10.18517/ijaseit.8.4-2.6832.

[8] N. Mezhoudi, R. Alghamdi, R. Aljunaid, G. Krichna, and D. Düştegör, "Employability prediction: A survey of current approaches, research challenges and applications," *J. Ambient Intell. Humaniz. Comp.*, vol. 14, no. 3, 2023, doi: 10.1007/s12652-021-03276-9.

[9] I. Römgens, R. Scoupe, and S. Beausaert, "Unraveling the concept of employability, bringing together research on employability in higher education and the workplace," *Stud. High. Educ.*, vol. 45, no. 12, 2020, doi: 10.1080/03075079.2019.1623770.

[10] A. Dubey and M. Mani, "Using machine learning to predict high school student employability: A case study," In *IEEE Int. Conf. on Data Science and Advanced Analytics (DSAA) Proceedings*, 2019, doi:10.1109/DSAA.2019.00078.

[11] M. M. Almutairi and M. H. A. Hasanat, "Predicting the suitability of IS students' skills for the recruitment in Saudi Arabian industry," In *Saudi Computer Society National Computer Conf. (NCC) Proceedings*, 2018, doi: 10.1109/NCG.2018.8593016.

[12] R. Haque, S. B. Ho, I. Chai, and A. Abdullah, "Intelligent health informatics with personalisation in weather-based healthcare using machine learning," *Innovative Systems for Intelligent Health Informatics*, vol. 72. 2021, doi: 10.1007/978-3-030-70713-2_4.

[13] N. A. B. A. Rahman, K. L. Tan, and C. K. Lim, "Supervised and unsupervised learning in data mining for employment prediction of fresh graduate students," *J. Telecommun. Electron. Comput. Eng.*, vol. 9, no. 2–12, 2017, [Online]. Available: https://jtec.utem.edu.my/jtec/article/view/2787.

[14] D. Sobnath, T. Kaduk, I. U. Rehman, and O. Isiaq, "Feature selection for uk disabled students' engagement post higher education: A machine learning approach for a predictive employment model," *IEEE Access*, vol. 8, 2020, doi: 10.1109/access.2020.3018663.

[15] M. E. Cruz and R. E. Encarnacion, "Analysis and prediction of students' academic performance and employability using data mining techniques: A research travelogue," *Eurasia Proc. Sci. Technol. Eng. Math.*, vol. 16, 2021, doi: 10.55549/epstem.1068566.

[16] L. D. Pool and P. Qualter, "Emotional self-efficacy, graduate employability, and career satisfaction: Testing the associations," *Aust. J. Psychol.*, vol. 65, no. 4, 2013, doi: 10.1111/ajpy.12023.

[17] R. Haque, S. B. Ho, I. Chai, and A. Abdullah, "Parameter and hyperparameter optimisation of deep neural network model for personalised predictions of asthma," *J. Adv. Inf. Technol.*, vol. 13, no. 5, 2022, doi: 10.12720/jait.13.5.512-517.

[18] R. Haque C. Y. Ting, Y. K. Ee, K. H. Ng, M. N. Shaaban, C. Y. Pee, L. K. Wong, and D. B. Raja, "Geospatial features influencing the formation of COVID-19 clusters," *J. Syst. Manag. Sci.*, vol. 12, no. 5, 2022, doi: 10.33168/JSMS.2022.0501.

[19] M. A. Sapaat, A. Mustapha, J. Ahmad, and K. Chamili, "A data mining approach to construct graduates employability model in Malaysia," *Int. J. New Comput. Archit. Their Appl.*, vol. 1, no. 4, 2011, [Online]. Available: https://www.academia.edu/6838650/.

[20] B. Jantawan and C. F. Tsai, "The application of data mining to build classification model for predicting graduate employment," *Int. J. of Comp. Science and Information Security*, vol. 11, no. 10, 2013, doi:10.48550/arXiv.1312.7123.

[21] L. S. Hugo, "A comparison of machine learning models predicting student employment," *J. of Chemical Information and Modeling*, vol. 53, no. 9. 2018, [Online]. Available: http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1544127100472053

[22] K. C. Piad, M. Dumlao, M. A. Ballera, and S. C. Ambat, "Predicting IT employability using data mining techniques," In *3rd Int. Conf. Digit. Inf. Process, Data Mining, Wirel. Commun. Proceedings*, pp. 26–30, 2016, doi: 10.1109/DIPDMWC.2016.7529358.

[23] M. T. Aziz and Y. Yusof, "Graduates employment classification using data mining approach," in *AIP Conference Proceedings*, 2016, vol. 1761. doi: 10.1063/1.4960842.

[24] P. Thakar, A. Mehta, and Manisha, "A unified model of clustering and classification to improve students' employability prediction," *Int. J. Intell. Syst. Appl.*, vol. 9, no. 9, 2017, doi: 10.5815/ijisa.2017.09.02.

[25] M. S. Kumar and G. P. Babu, "Comparative study of various supervised machine learning algorithms for an early effective prediciton of the employability of students," *Journal of Engineering Sciences*, vol. 10, no. 10, pp. 240–251, 2019, [Online]. Available: https://jespublication.com/upload/2019-V10-I10-37.pdf.

[26] C. D. Casuat, "Predicting students' employability using support vector machine: A SMOTE-optimized machine learning system," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 5, 2020, doi:10.30534/ijeter/2020/102852020.

[27] C. Patro and I. Pan, "Decision tree-based classification model to predict student employability," *Advances in Intelligent Systems and Computing*, vol. 1355, 2021, doi: 10.1007/978-981-16-1543-6_32.

[28] S. Srivastava, A. Mendiratta, P. Pankaj, R. Misra, and R. Mendiratta, "Happiness at work through spiritual leadership: A self-determination perspective," *Empl. Relations*, vol. 44, no. 4, 2022, doi: 10.1108/ER-08-2021-0342.

[29] A. P S and B. Abraham, "A predictive model for student employability using deep learning techniques," *ECS Trans.*, vol. 107, no. 1, 2022, doi: 10.1149/10701.10149ecst.

[30] R. R. Maaliw, K. A. Quing, A. C. Lagman, B. H. Ugalde, M. A. Ballera, and M. A. Ligayo, "Employability prediction of engineering graduates using ensemble classification modeling," In *IEEE 12th Annual Comp. and Communication Workshop and Conf. (CCWC) Proceedings*, 2022, doi: 10.1109/CCWC54503.2022.9720783.

[31] O. Saidani, L. J. Menzli, A. Ksibi, N. Alturki, and A. S. Alluhaidan, "Predicting student employability through the internship context using gradient boosting models," *IEEE Access*, vol. 10, 2022, doi:10.1109/access.2022.3170421.

[32] R. Haque, S. B. Ho, I. Chai, and A. Abdullah, "Investigating the impacts of weather and personalisation on asthma exacerbations using machine learning," In *Int. Conf. on Software and Comp. App. Proceedings*, pp. 191–197, 2022, doi: 10.1145/3524304.3524332.

[33] R. Haque, S. B. Ho, I. Chai, and A. Abdullah, "Optimised deep neural network model to predict asthma exacerbation based on personalised weather triggers," *F1000Research*, vol. 10, 2021, doi:10.12688/f1000research.73026.1.

[34] F. Younas, M. Usman, and W. Q. Yan, "An ensemble framework of deep neural networks for colorectal polyp classification," *Multimed. Tools Appl.*, 2022, doi: 10.1007/s11042-022-14177-0.

[35] R. Blanquero, E. Carrizosa, P. Ramírez-Cobo, and M. R. Sillero-Denamiel, "Constrained Naïve Bayes with application to unbalanced data classification," *Cent. Eur. J. Oper. Res.*, vol. 30, no. 4, 2022, doi:10.1007/s10100-021-00782-1.

[36] S. Sen, K. P. Singh, and P. Chakraborty, "Dealing with imbalanced regression problem for large dataset using scalable artificial neural network," *New Astron.*, vol. 99, 2023, doi:10.1016/j.newast.2022.101959.

[37] B. Güneş, "To what extent does perceived employability affect life satisfaction ? Findings from Munzur University students," *Journal of Yasar University*, vol. 17, no. 65, pp. 117–133, 2022, [Online]. Available: https://dergipark.org.tr/en/download/article-file/1901336.

[38] I. Otache and J. E. Edopkolor, "Work placement learning and student employability: Do student satisfaction, commitment and achievement matter?," *Ind. High. Educ.*, vol. 36, no. 6, 2022, doi:10.1177/09504222221091989.

[39] O. N. Bordean and A. Sonea, "Student satisfaction and perceived skills: Any link to employability?," *Entrep. Sustain. Issues*, vol. 6, no. 1, 2018, doi: 10.9770/jesi.2018.6.1(22).

[40] M. D. Laddha, V. T. Lokare, A. W. Kiwelekar, and L. D. Netak, "Performance analysis of the impact of technical skills on employability," *Int. J. Performability Eng.*, vol. 17, no. 4, 2021, doi:10.23940/ijpe.21.04.p5.371378.

[41] A. S. Deshmukh and A. B. Raut, "Student's employability concern for colleges and universities," *Smart Innovation, Systems and Technologies*, 2023, vol. 311, doi: 10.1007/978-981-19-3571-8_28.

[42] S. Y. Ong, C. Y. Ting, H. N. Goh, A. Quek, and C. L. Cham, "Workplace preference analytics among graduates," *J. of Informatics and Web Eng.*, vol. 2, no. 2, 2023, doi: 10.33093/jiwe.2023.2.2.17.