The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2021

Volume 16, Pages 117-131

IConTES 2021: International Conference on Technology, Engineering and Science

Analysis and Prediction of Students' Academic Performance and Employability Using Data Mining Techniques: A Research Travelogue

Maria Elisa Linda Taeza CRUZ
University of Nizwa

Riah Elcullada ENCARNACION
Oman Tourism College

Abstract: Higher education institutions (HEIs) handle tons of data to analyze and generate the most relevant information. Data mining is considered a useful tool to extract knowledge to predict future educational trends in this process. Hence, such a method is significant to HEIs to understand and predict students' employability and other critical academic elements. A comprehensive and systematic literature review was conducted to identify data mining techniques, algorithms, and the various data sets that will lead to the smart prediction and accuracy of student employability. The same method was used to determine the relationship between academic achievement and the employability of students. According to the research findings, the most frequently used data mining techniques for determining students' academic achievement and employability are Classification techniques, specifically the J48(C4.5) algorithm, the Naïve Bayes algorithm, and the CHAID Decision Tree algorithm. The most frequently used data sets or attributes for predicting students' academic performance and employability are their cumulative grade point average (CGPA), gender, technical, communication, problem-solving, analytical, critical thinking, and decision-making skills, extracurricular activities, and age, as well as psychomotor factors such as behavior and attendance and training/internship placement. Academic performance is the primary determinant of employability. The application of data mining techniques in academia has demonstrated its value in enhancing the performance of higher education institutions (HEIs). As a result, more research is urgently needed to ascertain the efficacy of the approaches, algorithms, and data sets identified as predictors of students' employability. Moreover, automated approaches should be utilized to ascertain their accuracy.

Keywords: Student employability, Data mining techniques, Datasets, Prediction, Student academic performance

Introduction

Graduates' employability is considered one of the key performance indicators for educational achievement and a measurement of the success of every higher education institution. A higher employment rate also signifies that HEIs have provided quality education and their graduates possess the necessary knowledge, skills, and values needed by industries. Much research has proven that academic qualifications and employability skills are vital in the employment process (Brown, 2002; Tholen, 2014). With this, many HEIs have put a considerable amount of effort into providing students with proper knowledge and employability skills to fulfill employment demands (Santiago et al., 2008).

However, regardless of the large number of graduates produced each year and the best pedagogies followed by HEIs, they still face the problems of high drop-out rates, low academic performance, and high unemployment among their students. According to Andrews et al. (2008), several concerns including the gap between the

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

needed graduate attributes and the industry requirements. The primary concern of industries has been that higher education institutions globally are failing to produce graduates with the necessary professional and life-long learning capabilities to prosper in the workplace (Mwita, 2018). Many studies have found that graduates must possess employability skills before immersing themselves in the workplace (Atkins, 1999; Billing, 2003; Crebert, et al., 2004a). This has strongly put pressure on HEIs to enhance graduates' employability and ensure that the knowledge, skills, and competencies they have gained will satisfy the labor market (Pitan, 2017).

Understanding and analyzing students' academic performance factors and predicting their employability using powerful tools such as data mining techniques can provide HEIs with valuable input for making policies and derive an appropriate solution. Information systems and educational databases serve as the backbone of HEIs' daily operations. The use of data mining has proven its worth in marketing, finance and banking, advertising, and recently in the academic setting. The enormous amount of students' data collected by HEIs can be a great source of relevant knowledge and the use of data mining techniques can reveal hidden patterns. Thus, these precious discoveries can help HEIs meet the industry requirements and provide them with high-quality graduates.

This paper conducted a thorough literature review of research and various sources on data mining techniques used to analyze and predict students' academic performance and employability. This research sought to answer the following questions:

1. What data mining techniques are frequently used to assess students' academic performance and employability?
2. Which data mining methods are most appropriate for predicting academic performance and employability of students?
3. Which data sets are most frequently utilized to predict academic performance and employability of students?
4. What is the relationship between the academic performance and employability of students?

Method

The purpose of this paper is to undertake a comprehensive review of relevant data mining methodologies and algorithms for predicting students' employability and academic performance based on earlier research. Furthermore, it sought to identify data sets that may be used to predict students' academic and employabilty. Finally, it reviewed research studies on the relationship between academic performance and employability of students. Thus, Table 1 illustrates how researchers previously used a variety of data mining techniques to predict students' academic performance and employability.

Table 1. Summary of the reviewed related studies on data sets, data mining techniques and algorithms used and the authors' findings

No.	Author/s	Datasets	Data Mining Techniques and Algorithms Used	Findings
1	Girase et al. (2018)	Academic Performance, Extra Curricular Activities, Personality Development, Technical, Training, Placement	Classification Techniques	• By utilizing the classification techniques, the number of students who would likely pass the examination will be successfully predicted. • The study demonstrated that the Decision Tree method outperformed the other data mining techniques tested.

No.	Author/s	Datasets	Data Mining Techniques and Algorithms Used	Findings
2	Mishra et al. (2016)	Socio-economic conditions, academic performances and a few more emotional skill factors (assertion, leadership skill empathy, decision making, drive strength, time management, self-esteem, stress management).	Bayesian methods, Multilayer Perceptrons and Sequential Minimal Optimization (SMO), Ensemble Methods and Decision Trees	For predicting employability, J48 is the most effective.
3	Thakar et al. (2017)	Personal Information, Demographic Information, Academic Records, Family Information, Social Parameters, Marketing, Cognitive Aptitude, Leadership Capabilities, Core Technical Skills Personality/Psychometric Aptitude, Finance Aptitude, English Aptitude, Quantitative Aptitude, Logical Aptitude, Mathematical Aptitude	• Clustering and classification techniques • Using chi-square analysis, clustering on two levels (k-means kernel); combination of four classifiers: k-star, random tree, simple cart, and random forest	High accuracy rates were seen with all classifier types.
4	Piad (2018)	Student's biographical information and their cumulative grade point average (CGPA) in their major and general subjects	• Logistic Regression • Chaid Algorithm (Different Classification algorithms were used to general models)	The most accurate algorithm was logistic regression, followed by the CHAID algorithm.
5	Piad et al. (2016)	Sudent's biographic information and Cumulative Grade Point Average (CGPA)	Classification techniques were employed to construct data models.	The Chaid Decision Tree has the highest accuracy, followed by Naïve Bayes and J48.
6	Othman et al. (2018)	Age, faculty, subject of study, co-curricular activities, marital status, industrial internship, and English proficiency were all taken into account. Age, industrial internship, and faculty all contain significant amounts of data and influence the final class, i.e., employability status.	• J48 • Multilayer • Perceptron (MLP) • Self-Organizing Maps (SOM)	J48 outperformed the other approaches.
7	Denila, P., et al. (2020)	Grades and First Employment	• C4.5 • Naïve Bayes	C4.5 classification is the most suitable algorithm
8	Tan et al. (2019)	Age, CGPA, Employability	• Logistic Regression • Decision Tree • Naïve Bayes • kNearest Neighbor • Support Vector Machine • Neural Network	Neural Network achieved the highest accuracy, followed by Support Vector Machine and Decision Tree

No.	Author/s	Datasets	Data Mining Techniques and Algorithms Used	Findings
9	Sapaat et al. (2011)	Graduate's gender, age, university/institution of current qualification, level of study for current qualification, field of study, CGPA, current employment status, levels of IT skills, Malay and English language proficiency, general knowledge, interpersonal communication, creative and critical thinking, analytical skills, problem solving, instillation of positive values, and teamwork acquired through the study program.	• J48 • Naïve Bayes • Information Gain	J48 performed with the highest accuracy compared to Bayes algorithms
10	Jantawan et al. (2013)	Prefix, Gender, Province, Degree, Educational Background, Faculty, Program, GPA, WorkProvince, Status, Talent, Position, Satisfaction, PeriodTimeFindWork, WorkDirectGraduate, ApplyKnowledgeWithWork, ReasonNotWork, ProblemofWork, RequirementsofStudy, LevelofStudyRequired, InstitutionNeed	• BFTree • NBTree • REPTree (J4.8 in WEKA) • Naïve Bayes classifier with five variants, including Averaged One-Dependence Estimators (AODE), BayesNet, HNB, and NaiveBayes, as well as Weightily Averaged One-Dependence Estimators (WAODE)	The method with the highest accuracy was the WAODE algorithm, a variant of the Bayes algorithm.
11	Sunday, et al. (2020)	Class Test Score, Completed Assignments, Class Laboratory Work, and Class Attendance	• ID3 • J48	J48 has a higher classification accuracy than the ID3 algorithm.
12	Al Lluhaybi, et al. (2018)	Admission, Course Related Data and Level 1 Final Grades	• C4.5 Decsion Tree • Naïve Bayes	• Naïve Bayes is better than C4.5 decision tree algorithm in predicting the students at high risk of failing. • Final grades influenced the results of the students in Level 2.
13	Funmilayo, O. and Ibukun, A. (2019)	Graduation CGPA, Biodata, Pre-admission requirements/results and performance in the university.	• Multiple linear Regression • C4.5	Multiple Linear Regression for a more effective prediction.

No.	Author/s	Datasets	Data Mining Techniques and Algorithms Used	Findings
14	Anuradha, C. and Velmurugan, T. (2016)	Student Gender, Students Branch, Students' Grade in High School, Students' Grade in High School, Medium of Instruction, Students' Residence Location whether a Student lives in a hostel or Not, Students' Family Size, Family Type, Family Annual Income, Fathers' and Mothers' Qualifications, Previous Semester Grade, Class Test Grade, Seminar Performance Assignment, Assignment, General Proficiency, Attendance, Lab Work and End Semester Marks	• Fast Boost Decision Tree algorithm	Datasets should be tried out on a set of qualifiers and the best one should be used.
15	Baradwaj et al., (2011)	Attendance, Class Test, Seminar and Assignment Marks	• Decision Tree	A decision tree was utilized to identify students who require additional attention and assistance in the upcoming semester exams.
16	Osmanbegovic et al. (2012)	Attained high school results, entrance exam scores, attitude toward learning and grades.	• Naïve Bayes • Decision Tree • Neural Network Methods	Naïve Bayes outperformed in prediction decision tree and neural network methods.
17	Shahiri et al. (2015)	Internal Assessment, CGPA, Extracurricular Activities, Student Demographics, High School Background, Scholarship, Interaction with Social Networks, Internal Assessment, Assessment, External Assessment	• Decision Tree • Artificial Neural Networks • Naïve Bayes • K-Nearest Neighbor • Support Vector Machine	Classification technique is most popular in predicting students' performance.
18	Kavyashree et al. (2016)	Gender, Education of the Father, Education of the Mother, Father's Occupation, Mother's Occupation Income, Loan, and Annual Family Income, Early Life, Instructional Medium, Performance in 10th Grade, 12th Grade, First Semester, and Second Semester, Academic Hours, Type of Graduation Degree, Gap Year Assertion, Empathy, Leadership, Drive and Management of Stress	• Modelling • Classification	Classification techniques can be used to identify weak students and take necessary action when developing a decision support system.
19	Sayana, (2015)	CGPA, Graduation Percentage, Attendance, Assignment Work and Unit Test Performance	• Decision Trees • Clustering using K-Means • Association	K- means clustering is preferred than other types.

No.	Author/s	Datasets	Data Mining Techniques and Algorithms Used	Findings
			Rule	
20	Ashraf, A., Anwer, S. & Khan, MG. (2018)	**Student Demographics** Age, Gender, Region, Place of Residence, and Information on the Guardian **Previously Obtained Results** Certificates, Scholarships and Results **Grades** Assessment Results, Quizzes, Final Exam, Cumulative Grade Point Average (CGPA) and Attendance **Social Network Details** Interaction with Social Media Websites **Extra-Curricular Activities** Games Partitions, Sports, Hobbies **Psychometric Factors** Behavior, Absence and Remarks	Decision Tree algorithms: ADTree, JRip, Ridor, logistic regression and neural network approach	The Decision Tree algorithms: JRip, Ridor ADTree, logistic regression and neural network approaches all produce accurate results when predicting and helping the educational system.
21	Thakar et al. (2015)	**Primary Attributes** Age, Secondary Percentage, Senior Secondary Percentage, Stream in Senior Secondary, Graduation Degree, Graduation %, Type (Regular/Distance), Graduating University, Post Graduate Sem 1%, Post Graduate Sem 2%, No. of Supplementaries in 1st Year, Gap in years between Graduation and PG **Secondary Attributes** Gender, Permanent Address, State, English, Quantitative Ability, Logical Ability, Attention to Detail, Computer Programming, Computer Science, Psychometric	Naïve Bayes from Bayes Category, RBF Network and Multilayer Perceptron from Functions Category, IB1 and IBk from Lazy Category, PART and DTNB from Rules Category, J48, Random Tree and Random Forest from Trees Category.	The secondary attributes played an essential role predicting accuracy of students' employability.
22	Limbu et al. (2019)	Gender, Age, Previous Year's Academic Performance (Academic), Family's Highest Level of Education, and Psychometric Factors (Have you ever achieved the highest grade in your class?, I study because I am interested in learning., I am willing to devote time to studying., My academic performance is contingent upon my efforts. Generally, my state of mind is positive and I feel good., I understand the purpose of the	Hybrid Algorithm (i.e., combination of K-means Algorithm and Support Vector Machine Algorithm)	By using an unsupervised machine learning algorithm (such as the K-means Clustering Technique) results in a high level of output accuracy, using a hybrid model results in an even higher level of output accuracy than using any other single machine learning technique.

No.	Author/s	Datasets	Data Mining Techniques and Algorithms Used	Findings
		course contents., I establish a conducive environment for effective study., I select acceptable classmates for teamwork., What makes me comfortable when I study.)		
23	Yusuf et al. (2018)	Socio-Economic, Psychometric Factors, University Entrance Examination, Senior, Secondary Graduation Examinations Performance, CGPA	Decision Tree, Artificial Neural Networks, Naïve Bayes, K-Nearest Neighbor and Support Vector Machine	The CGPA was employed by most studies and internal assessments to predict students' academic success. Classification is a frequently used method. Researchers generally employ the Decision Tree method to predict students' academic performance.
24	Shade et al. (2013)	Family Background Factors, Previous Academic Achievement, Cumulative Grade Point Age (CGPA)	10 Classification Models (Random Forest, Random Tree, J48 Decision Stump and REPTree and five rule induction algorithms – Jrip, OneR, ZeroR, PART and Decision Table)	The performance of Random Trees has been proved to be superior to that of other algorithms. The findings of the study back up Quadri and Kalyankar's (2010) and Yadav et al. (2012) findings that Decision Trees are the best algorithms for predicting student performance.
25	Kumar et al. (2016)	Students Branch, Gender, High School Grade, Senior Grade, Secondary Grade, Student's Family Size, Students Family Status, Fathers Qualification, Mother's Qualification, Father's Occupation, Mother's Occupation, Result in BTech First Year	J48, Random Forest, REPTree and LADTree	The J48 classification method is the most accurate of the many decision trees.
26	Estrera et al. (2017)	GPA, student demographics (gender), psychometric variables (study habits, interest in studies, and time spent studying), extracurricular activities, and student demographics	Classification model was used (Decision Tree, Artificial Neural Networks, Naïve Bayes, K-Nearest Neighbor and Support Vector Machine)	The Decision Tree Method is the most effective algorithm for predicting student achievement.
27	Yadav et al. (2012)	Students Branch, Students Sex, Students Category, Students Grade in High School, Students Grade in Senior Secondary, Admission Type, Medium of Teaching, Living Location of Student, Student Live in Hostel or Not, Student's Family Size, Students Family Status, Family Annual Income,	Classification Techniques (ID3, C4.5, CART)	The decision tree predicted how many students would pass, fail, or be promoted to the next year. When the results were compared, it became clear which students required further assistance to improve their performance.

No.	Author/s	Datasets	Data Mining Techniques and Algorithms Used	Findings
		Father's Qualification, Mother's Qualification, Father's Occupation, Mother's Occupation, Result in BTech First Year		
28	Kumar et al. (2019)	**Student Attributes** (Possible Values Used in all Research Papers while Implementing DM Algorithm)**, Academic Attributes** (Internal and External Assessment, Lab Marks, Sessional Marks, Attendance, CGPA, semester marks, grade, seminar performance, assignment, attendance, school marks, etc.), **Personal Attributes** (Age, Gender, height, weight, student interest, level of motivation, sports person, hobbies, etc.), **Family Attributes** (Qualification, Occupation, Income, Status, Support, Siblings, Responsibilities, etc.), **Social Attributes** (number of friends, social networking, girls/boys friends, movies, trip outings, and friend parties, etc.) (Teaching Medium, Accommodation, Infrastructure, Water & Toilet facilities, Transportation system, Class Size, School Reputation, School Status, Class Size, School Type, Teaching Methodology, etc.)	Classification Algorithms (Decision Tree, Naïve Bayes, Neural Network, K-Nearest Neighbor Algorithm, Support Vector Machine Algorithm)	Predictions are made using the CGPA, as well as internal and external assessment marks. Classification is the most commonly used technique for performance prediction.
29	Jayaprakash et al. (2018)	**Student Pre-enrollment Data** (Junior and Higher Secondary Grades, Entrance Exam Scores, Medium of Study), **Continuous Assessment** (Attendance, Team Scores, Lab Performance, Course and Lecturer Assessment, Forum Participation, Effective Use of Resources), **Psychological Data** (Personality, Social Activities), **Socio Economic Status** (Parents Education Qualification, Parents Occupation, Place of Stay/Distance Travelled, Family Size)	Naïve Bayes Algorithm, Rule-Based Classifier, SPPN (Students Performance Prediction System), Ensemble Learning	Sudents' pre-enrollment data, ongoing assessment data, and psychological data all contribute significantly to the prediction of academic performance. The Naïve Bayes Algorithm and Rule-Based Classifier can be used to perform predictive analysis.
30	Aziz et al. (2013)	Race, Gender, Hometown, Family Income, University Entry Mode and Grade Point	Naïve Bayes Classifier	The Naïve Bayes Classifier predicts with a precision of 57.5 percent.

No.	Author/s	Datasets	Data Mining Techniques and Algorithms Used	Findings
		Average		The study discovered that the income of students' families, their gender, and their locality all have an effect on their academic achievement. Additional data should be collected to improve the accuracy of the prediction model.

Scope and Limitations

This investigation focused on the application of data mining techniques to the education sectors of education. It emphasizes on the analysis and prediction of students' performance and employability. Due to the COVID19 pandemic, the study adopted the online research approach. It analyzed primary research and secondary data published from 2010 to 2020 only to ensure their timeliness and relevance to the present scenario. The complete literature review, analyzing journal articles, blogs, case studies, and primary studies, lasted for three months. It was not aimed at achieving a concluding resolution to the pressing problem but to unveil information that could enhance awareness among educational stakeholders on the usefulness of data mining techniques.

The Input-Process-Output (IPO) diagram below shows three significant boxes with their series of activities and requirements. The flow of control is shown by the arrows depicting the first stage to the third stage. The researchers were able to develop a conceptual framework suitable to the study by substituting generic model variables.

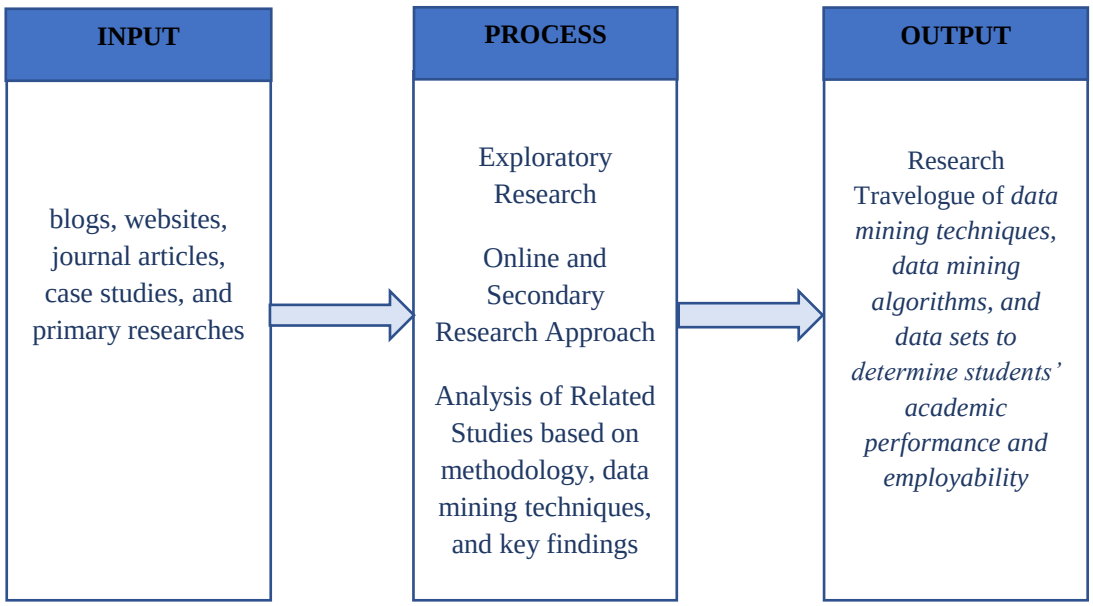

Figure 1: The Conceptual Framework of the Study

As depicted in Figure 1.0, the Input stage is the initial phase that identified the study's requirements taken from blogs, websites, journal articles, case studies, and primary researches. The Process is the second stage in the model where the inputs were analyzed, processed, and evaluated using an online research approach. Finally, the Output summarizes important discoveries such as the data mining techniques, algorithms, and data sets /variables essential to determine students' employability and students' academic performance.

Research Design

This present study was classified as exploratory research. It is considered a novel effort to comprehensively review various significant investigations in educational data mining techniques in a single study. According to Bhat (2020), exploratory research results aim to provide a hypothetical or theoretical idea for further investigation.

Data Collection and Analysis

The online research approach was used to gather the desired amount of data from the internet. The study only covers those pieces of literature that are undisputable and reliable. The secondary research approach was employed to summarize, collate, and synthesize existing primary research sources and generate the necessary information. A total of thirty (30) related studies were analyzed and synthesized. The results were then categorized according to data mining techniques/algorithms used, the variables or dimensions used and the research findings.

Results and Discussion

The results and findings of the study are found below:

The Most Common Data Mining Methods to Determine Students' Academic Performance and Employability

Generally, it was found out that among the different studies conducted to predict graduates' employability and students' academic performance, Classification techniques were the most common. A complete description was presented by the various authors, specifying the methods, the results and discussion of the results.

The application of Classification techniques in education was very useful for determining students' employability. These include predicting students' successful examinations (Girase et al., 2018), developing of emotional skills (Mishra et al., 2016), promising job offers & employment (Mishra et al., 2016; Piad, 2018; Piad et al., 2016; Denila et al., 2020; Jantawan et al., 2013), constructing a unified employability prediction model (Thakar et al., 2017; Sapaat et al., 2011), and identifying critical factors affecting employment probability (Othman et al., 2018; Tan et al., 2019).

Meanwhile, there are also many benefits of utilizing Classification techniques for analyzing students' academic performance (Sunday et al., 2020; Kumar et al., 2016; Kumar et al., 2019; Aziz et al., 2013), predicting the high risks for failure (Al Lluhaybi et al., 2018; Baradwi, 2011; ; Oladapo et al., 2019), identifying weak students (Kavyashree et al., 2016), developing a predictive system of academic performance (Funmilay et.al., 2019; Anuradha, et.al., 2016), predicting successful academic performance (Baradwi, 2011; Osmanbegovic et al., 2012; Shahiri et al., 2015; Sayana, 2015; Estrera et al., 2017), and enhancing the educational system (Ashraf et.al., 2018)

The Most Suitable Data Mining Algorithms to Predict Students' Academic Performance and Employability

The literature review revealed that J48 (C4.5), Naïve Bayes algorithms, and CHAID Decision Tree algorithms are the most appropriate data mining algorithms for predicting students' employability and determining students' academic performance.

The J48 algorithm is one of the most effective machine learning algorithms for constructing a decision tree that categorizes and constantly classifies input. It outperformed the other procedures and produced the highest accuracy result Othman et al., 2018; Denila et al., 2020; Sapaat et al., 2011; Sunday et al., 2020; Kumar et al., 2016). The J48 algorithm is a simple classifier of C4.5 algorithms in an open-source Java platform of Weka Data mining tool and developed by Ross Quinlan (Girase et al., 2018). A decision model tree is built for the classification process using this method. After the tree has been constructed, it is applied to each tuple in the database and classified accordingly.

Denila et al. (2020), Tan et al., (2019), Jantawan et al. (2013), Sapaat et al. (2011), and Aziz et al. (2013) all agreed that the Naïve Bayes agorithm is the most common algorithm for classification tasks, notably in assessing and predicting students' employability. The Naïve Bayes method is a probabilistic machine learning algorithm based on Rev. Thomas Bayes' Bayes Theorem (Al Lluhaybi, et al., 2018; Tan et al., 2019). The algorithm uses binary and multi-class classification categories to build a model (Sapaat et al., 2011) for predicting high student academic success (Osmanbegovic et al., 2012; Shahiri et al., 2015) as well as, on the other hand, for predicting high student failure rates (Al Lluhaybi, et al., 2018).

The Chi-square Automatic Interaction Detector (CHAID) Decision Tree algorithm was used to build a predictive model and identify the relationships between variables (Piad et al., 2016; Piad, 2018). It was designed by Gordon V. Kass in 1980 (Piad, 2018); the process involved continuous splitting the predictors (nominal, ordinal, and continuous data) into categories and cross tabulating each categorical predictor to derive the best result. In this technique, a classification tree is generated, showing the relationships between the split categories and the related factors within the tree.

The Most Commonly Used Data Sets or Variables for Predicting Students' Academic Performance and Employability

Twenty-seven (27) out of thirty (30) papers identified Cumulative Grade Point Average (CGPA) as the most significant data set used in predicting students' performance and employability. The reason why most of the researchers used CGPA and grades as a measure was because of its tangible value that can predict student employability and career performance (Shahiri et al., 2015; Piad, 2018; Tan et al., 2019; Durga et al., 2019) as well as for determining high academic performance (Funmilayo et al., 2019; Shahiri et al., 2015; Sayana, 2015; Ashraf et al., 2018; Estrera et al., 2017).

Other datasets that were found significant include gender (Ashraf et al., 2018; Thakar et al., 2017; Kavyashree et al., 2016; Anuradha et.al., 2016; Jantawan et al., 2013; Sapaat et al., 2011; Limbu et al., 2019; Estrera et al., 2017; Aziz et al., 2013), technical, communication, critical thinking, problem solving, analytical and decision-making skills (Mishra et al., 2016; Thakar et al., 2017; Sapaat et al., 2011; Kavyashree et al., 2016), extra-curricular activities (Girase et al., 2018; Shahiri et al., 2015; Ashraf et.al., 2018; Oathman, Z., et al., 2018; Estrera et al., 2017), age (Tan et al., 2019; Sapaat et al., 2011; Oathman, Z., et al., 2018; Ashraf et.al., 2018; Limbu et al., 2019), psychomotor factors such as behavior and attendance (Ashraf et.al., 2018; Sayana, 2015; Anuradha et.al., 2016; Kavyashree et al., 2016; Limbu et al., 2019; Jayaprakash et al., 2018) and training/internship placement (Girase et al., 2018; Oathman, Z., et al., 2018; Gault et al., 2010; Weible et al., 2011).

The Relationship Between Students' Academic Performance and Employability

The review of various studies signified that students' academic performance dramatically influences employability. Thus, the main predictor of students' employability is academic performance (Gokuladas, 2011; Hassanbeigi et al., 2011; Imose et al., 2015; Johansen, 2014; Kuncel et al., 2010; Funmilayo et al., 2019; Shahiri et al., 2015; Sayana, 2015; Ashraf et.al., 2018; Tentama et al., 2019; Pan et al., 2011; Edinyang et al., 2015; Moya Clemente et al., 2020; Kumar et al., 2019). This was also supported by other authors that revealed that it is the main predictor of employability (Funmilayo et al., 2019; Shahiri et al., 2015; Sayana, 2015; Ashraf et al., 2018; Yusuf et al., 2018; Vasan et al., 2018.

Multiple linear regression analysis was utilized by Tentama et al. (2019) to find a significant relationship between academic achievement and employability. On the other hand, Pan et al., (2011) looked into the relationship and discovered that academic performance is closely linked to employability skills such foreign language proficiency, computer literacy, application and theory to work, and stability and pressure resistance. As a result, their data revealed that employability is linked not just to the employment process but also to academic performance.

Academic achievement is a key influence in graduate employability in a specific discipline, according to studies by Edinyang et al. (2015) and Vasan et al. (2018). Furthermore, academic factors may influence a student's ability to find work in a given field. As a result, the authors advise that industry sectors collaborate with policymakers and academic curriculum developers to meet the demands of the global labor market. Moya

Clemente et al., (2020) discovered that participation in international exchanges and internships, in addition to academic performance, predict future employability.

Conclusion

In general, this in-depth study proved that data mining is an interesting field of data science and has been widely used in the educational setting for years. Moreover, various studies have shown remarkable outcomes in predicting students' employability and identifying academic performance. The Classification Techniques were predominantly used for prediction and analysis tasks, specifically the use of J48 (C4.5), Naïve Bayes algorithms, and CHAID Decision Tree algorithms, because of their high accuracy results.

The analysis yielded some common variables that were used for classification, such as CGPA, gender, technical, communication, analytical and decision-making skills, problem-solving, critical thinking, extra-curricular activities, age, psychomotor factors such as behavior and attendance and training/internship placement. In addition, the systematic review of related literature and studies manifested that academic performance greatly influences students' employability.

Recommendations

From the present study's findings, it can be deduced that data mining techniques are significantly relevant in uplifting the performance of every HEI around the world. The discovery of precious knowledge is essential for facilitate the planning of services and guide actions for improvement. Therefore, a follow-up developmental research is deemed necessary to apply the identified data mining algorithms in Omani Higher Education and the identified attributes/datasets and reveal their importance for students' employability and academic performance. Moreover, automated methods will be used to determine the efficacy of the attributes/datasets and the accuracy of the identified data mining algorithms. It is also recommended that HEIs utilize the findings of the study. The industry should also strongly collaborate with the academe to identify and set the needed skills to meet the labor market requirements.

The findings of this investigation could generate a better understanding and social awareness of the usefulness of data mining techniques from the academic perspective. Moreover, it highlighted their implications on students' academic performance and employability. Thus, academic stakeholders particularly the students, industry partners, and educational providers, are the greatest benefactors of this study. Future researchers could also use the results of this exploration as additional literature for review and further investigative study.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

References

Al Lluhaybi, M., Tucker, A. & Yousef, L. (2018). The prediction of student failure using classification methods: a case study. *Computer Science and Information Technology, 79-90.* DOI: 10.5121/csit.2018.80506.

Andrews, J., & Higson, H. (2008). Graduate employability,'soft skills' versus 'hard'business knowledge: A European study. *Higher education in Europe, 33*(4), 411-422.

Anuradha, C. & Velmurugan, T. (2016). fast boost decision tree algorithm: a novel classifier for the assessment of student performance in educational data. *Printed in Portugal*, *31*(11), 139-155.

Ashraf, A., Anwer, S., & Khan, M. G. (2018). A Comparative study of predicting student's performance by use of data mining techniques. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, *44*(1), 122-136.

Atkins, M. J. (1999). Oven- ready and self- basting: taking stock of employability skills. *Teaching in higher education*, *4*(2), 267-280.

Aziz, A. A., Ismail, N. H., Ahmad, F., & Hassan, H. (2015). A framework for students's academic performance analysis using naive bayes classifier. *Jurnal Teknologi*, *75*(3), 13-19. DOI: 10.11113/jt.v75.5037.

Baradwaj, B. & Pal, S. (2011). Mining Educational Data to Analyze Students' Performance. International *Journal of Advanced Computer Science and Applications*, *2*(6), 63-69.

Bhat, A. (2020, January 24). *Exploratory research, definition, methods, types and examples.* https://www.questionpro.com/blog/exploratory-research/#Secondary_research_methods.

Billing, D. (2003). Generic cognitive abilities in higher education: An international analysis of skills sought by stakeholders. *Compare: A Journal of Comparative and International Education*, *33*(3), 335-350.

Brown, BL. (2002). *Generic skills in career and technical education: myths and realities.* Educational Resources Information Centre.

Moya Clemente, I., Ribes Giner, G., & Sanahuja Vélez, G. (2020). Towards sustainability in university education. ımproving university graduates chances of employability by participation in a high achievement academic program. *Sustainability*, *12*(2), 1-17. doi: 10.3390/su12020680.

Crebert, G., Bates, M., Bell, B., Patrick, C. J., & Cragnolini, V. (2004). Ivory tower to concrete jungle revisited. *Journal of education and work*, *17*(1), 47-70.

Denila, P. G., Delima, A. J. P., & Vilchez, R. N. (2020). Analysis of IT graduates employment alignment using C4. 5 and Naïve Bayes algorithm. *Int J Adv Trends ComputSciEng*, *9*(1), 745-752.

Durga, V. S., & Thangakumar, J. (2019). A complete survey on predicting performance of engineering students. *Int. J. Civ. Eng. Technol*, *10*(2), 48-56.

Edinyang, S.D., Odey, C., & Joseph, G. (2015). academic factors and graduate employability in Nigeria. *Global Journal of Human Resource Management*, *3*(5), 9-17.

Estrera, P.J., Natan, P., Rivera, B.F. & Colarte, Faith, B. (2017). *International Journal of Engineering and Techniques*, 2(5), 147-154.

Funmilayo, O. & Ibukun, A. (2019). Student's performance prediction using multiple linear regression and decision tree. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *8*(7), 256-268.

Gault, J., Leach, E., Duey, M. (2010). Effects of business internships on job marketability: the employers' perspective. *Journal of Education and Training*, *52*, 76-88.

Girase, M., Lad, S., & Pachpande, P. (2018). Student's employability prediction using data mining. International *Journal of Scientific & Engineering Research*, *9*(4), 27-29.

Gokuladas, V. K. (2011). Predictors of employability of engineering graduates in campus recruitment drives of Indian software services companies. *International Journal of Selection and Assessment*, *19*(3), 313-319. https://doi.org/10.1111/j.1468-2389.2011.00560.x,

Hassanbeigi, A., Askari, J., Nakhjavani, M., Shirkhoda, S., Barzegar, K., Mozayyan, M. R., & Fallahzadeh, H. (2011). The relationship between study skills and academic performance of university students. *Procedia-Social and Behavioral Sciences*, *30*, 1416-1424.

Imose, R., & Barber, L. K. (2015). Using undergraduate grade point average as a selection tool: A synthesis of the literature. *The Psychologist-Manager Journal*, *18*(1), 1-11.

Jantawan, B. & Tsai, C. (2013). The application of data mining to build classification model for predicting graduate employment. *(IJCSIS) International Journal of Computer Science and Information Security*, *11*(10), 1-7.

Jayaprakash, S., & Jaiganesh, V. (2018). A Survey on Academic Progression of Students in Tertiary Education using Classification Algorithms. *International Journal of Engineering Technology Science and Research IJETSR*, *5*(2), 136-142.

Johansen, V. (2014). Entrepreneurship education and academic performance. *Scandinavian Journal of Educational Research*, *58*(3), 300-314.

Kavyashree, K. R., & Laksmi, D. (2016). A review on mining students' data for performance prediction. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(4), 1104-1106.

Kumar, M., & Salal, Y. K. (2019). Systematic review of predicting student's performance in academics. *Int. J. of Engineering and Advanced Technology*, *8*(3), 54-61..

Kumar, M. & Singh, A.J. (2016). Predicting students' performance using classification techniques in data mining. *International Journal of Technology and Computing (IJTC)*, 2(10), 489-494.

Kuncel, N. R., Ones, D. S., & Sackett, P. R. (2010). Individual differences as predictors of work, educational, and broad life outcomes. *Personality And İndividual Differences*, *49*(4), 331-336.

Limbu, J. & Sah, S. (2019). Prediction on Student Academic Performance Using Hybrid Clustering. *LEBF Research Journal of Science, Technology and Management*, *1*(1), 1-22.

Mishra, T., Kumar, D., & Gupta, S. (2016). Students' employability prediction model through data mining. *International Journal of Applied Engineering Research*, *11*(4), 2275-2282.

Mwita, K. M. (2018). Tanzania graduate employability: Perception of human resource management practitioners. *International Journal of Human Resource Studies*, *8*(2), 263-272.

Oladapo, K.A. & Adedokun, A. (2019). Data mining technique for early detection of at-risk students. *2nd International Conference on Information Technology in Education and Development*, Lagos, Nigeria.

Osmanbegovic, E., & Suljic, M. (2012). Data mining approach for predicting student performance. *Economic Review: Journal of Economics and Business*, *10*(1), 3-12.

Othman, Z., Shan, S. W., Yusoff, I., & Kee, C. P. (2018). Classification techniques for predicting graduate employability. *International Journal on Advanced Science, Engineering and Information Technology*, *8*(4-2), 1712-1720.

Pan, Y. J., & Lee, L. S. (2011). Academic performance and perceived employability of graduate students in business and management–an analysis of nationwide graduate destination survey. *Procedia-Social and Behavioral Sciences*, *25*, 91-103.

Pıad, K. C. (2018). Determining the dominant attributes of information technology graduates employability prediction using data mining classification techniques. *Journal of Theoretical & Applied Information Technology*, *96*(12). 3780-3790.

Piad, K. C., Dumlao, M., Ballera, M. A., & Ambat, S. C. (2016, July). Predicting IT employability using data mining techniques. In *2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC)* (pp. 26-30). IEEE.

Pitan, O. S. (2016). Towards enhancing university graduate employability in Nigeria. *Journal of Sociology and Social Anthropology*, *7*(1), 1-11. https://doi.org/10/1080/09766634.2016.11885696.

Santiago, P., Tremblay, K., Basri, E., & Arnal, E. (2008). *Tertiary education for the knowledge society* (Vol. 1). OECD. https://doi.org/10.1787/978926406351-hu/.

Sapaat, M. A., Mustapha, A., Ahmad, J., Chamili, K., & Muhamad, R. (2011, July). A classification-based graduates employability model for tracer study by MOHE. In *International Conference on Digital Information Processing and Communications* (pp. 277-287). Springer, Berlin, Heidelberg.

Sayana, TS. (2015). Prediction of students academic performance using data mining: analysis. *International Journal of Engineering Research & Technology*, *3*(20), 1-4.

Shade, K., Goga, N., Awodele, & Okolie, S. (2013). Optimal algorithm for predicting students' academic performance. *Internal Journal of Computers & Technology*, *4*(1), 63-75.

Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, *72*, 414-422. doi:10.1016/j.procs.2015.12.157, pp. 414-422.

Sunday, K., Ocheja, P., Hussain, S., Oyelere, S., Samson, B., & Agbo, F. (2020). Analyzing student performance in programming education using classification techniques. *International Journal of Emerging Technologies in Learning (iJET)*, *15*(2), 127-144.

Tan, K., Abdul Rahman, NA., Lim, C. (2019). A comparative of predictive model of employability. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *8*(8), 375-378.

Tentama, F., & Abdillah, M. H. (2019). Student Employability Examined from Academic Achievement and Self-Concept. *International Journal of Evaluation and Research in Education*, *8*(2), 243-248.

Thakar, P., Mehta, A. & Manisha, D. (2015). Role of secondary attributes to boost the prediction accuracy of students' employability via data mining. *International Journal of Advanced Computer Science and Application. 6*(11), 84-90.

Thakar, P., & Mehta, A. (2017). A unified model of clustering and classification to improve students' employability prediction. *International Journal of Intelligent Systems and Applications*, *9*(9), 10. DOI: 10.5815/ijisa.2017.09.02.

Tholen, G. (2014). Graduate employability and educational context: a comparison between Great Britain and the Netherlands. *British Educational Research Journal*, *40*(1), 1-17. DOI: 10.1002/berj.3023.

Vasan, T. M., Sharma, K. J., & Chauhan, N. C. (2018). A comparative study of student's employability prediction model using data mining technique. *International Journal of Creative Research Thoughts (IJCRT)*, *6*(1), 264-270.

Weible, R. & McClure, R. (2011). An exploration of the benefits of student internships to marketing departments. *Marketing Education Review*, *21*(3), 229-240.

Yadav, S. K. & Pal. S. (2012). Data Mining: A prediction for performance improvement of engineering students using classification. *World of Computer Science and Information Technology Journal (WCSIT)*, *2*(2), 51-56.

Yusuf, A. & Lawan, A. (2018). Prediction of students' academic performance using educational data mining technique: Literature review. https://www.academia.edu/34009565/prediction_of_students_academic_performance_using_educational_datamining_technique_literature_review.

Author Information

Maria Elisa Linda Taeza CRUZ
University of Nizwa
Birkat Al Mouz, Nizwa, Oman
Contact e-mail: *mariacruz@unizwa.edu.om*

Riah Elcullada ENCARNACION
Oman Tourism College
P.O. Box 2008, C.P.O, Postal Code 111, Muscat, Oman

To cite this article:

Cruz, M.E.L.T. & Encarnacion, R.E. (2021). Analysis and prediction of students' academic performance and employability using data mining techniques: A research travelogue. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 16*, 117-131.