Analysis of IT Graduates Employment Alignment Using C4.5 and Naïve Bayes Algorithm

Pamela Grace L. Denila[*], Allemar Jhone P. Delima[2], Ramcis N. Vilchez[3]
[1,2] Professional Schools, University of Mindanao, Matina, Davao City, Philippines
[3] College of Computer Education, University of Mindanao, Matina, Davao City, Philippines
[1]pameladenila@umindanao.edu.ph, [2]allemardelima@umindanao.edu.ph, [3]ramcis_vilchez@umindanao.edu.ph

ABSTRACT

There is a significant increase in the number of graduates produced by higher education institutions every year and the number of graduates is growing immensely compared to job opportunities in the market. Data mining techniques serve as the foundation of various researches in generating knowledge from the dataset obtained from schools, universities, and companies. In this paper, the application of classification mining with the use of the C4.5 and Naïve Bayes algorithms, analyzed the dataset gathered from the Office of the Registrar of Davao del Norte State College. This study aims to determine the correlation between students' grades in their major subjects and their employment right after completing a degree in Information Technology. The experimental results showed that the Naive Bayes algorithm garnered 67.22% accuracy using 10-fold cross-validation scheme and 55.91% accuracy with the 70% training and 30% testing percentage split. Further, the C4.5 algorithm accumulated 95.52% and 92.13% accuracies using the same cross-validation method and data composition percentage split, respectively. It goes to show that the best algorithm in analyzing the employment alignment of the IT students is the C4.5. This research would help the students and the university to improve various aspects of educating the students and see to it that they would produce quality graduates in the following years. The use other algorithms to generate more significant models, as well as adding more attributes to identify other factors affecting graduate's employment is recommended.

Key words: BSIT, Data Mining, Employment Alignment, Tracer Study

1. INTRODUCTION

The higher education institutions (HEI) are producing more graduates every year. The significant increase in the number of graduates does not equate to the job opportunities available in the market [1]. Today, graduates' employment is one of the problems in different institutions around the world [2]. In Malaysia, there are more than 180,000 graduates each year, but, unfortunately, based on the survey conducted by the Malaysian government, it showed that 60,000 graduates were unemployed [1]. With this, in-depth investigations should be done to determine the factors that affect graduates' employability moreover to apply efficient strategies in solving the problem [3].

In previous years, data mining has been a trend in analyzing massive amounts of data [4]; this includes clustering, regression, association, and classification mining [1]. Data mining techniques serve as the foundation of various researches from schools, universities, and companies in achieving better knowledge extraction from the available dataset [5].

The application of classification mining such as C4.5 and Naïve Bayes Algorithm in this research will analyze the data set gathered from the Office of the Registrar of Davao del Norte State College. The data set comprises the grades in their major subjects of 2018, 2017 and 2016 graduates and their first employment based on the Tracer Study administered by the Institute of Information Technology. This research aims to determine the correlation between students' grades in their major subjects and their employment right after completing a degree in Information Technology. This research would also analyze the students' educational records based on the grades provided to prove how different subjects affect the future employment of the graduates. Analyzing such would help the students and the university to improve various aspects in educating the students and see to it that they would produce quality graduates in the following years.

2. RELATED LITERATURE

The application of data mining techniques to a vast amount of educational data has helped different schools in the field of research and other aspects [6].

A study made by [7] implemented three data mining techniques: Decision Tree, Logistic Regression and Artificial Neural Network to extract information from the Malaysian graduates' employment data set. This study is conducted to determine the best data mining technique in predicting graduates' employment, whether in the public or private sectors. The results showed that the Artificial Neural Network with 81.52% accuracy is the most reliable in predicting graduates' employment.

In [8], the authors assessed the employability of students using five data mining techniques. The students were first evaluated based on the number of failed subjects, and their assessment test scores composed of soft skills, problem-solving skills and technical skills. The students' records were filtered, and out of 89,077 student records, only 2,100 student records were used as data sets eligible for testing using a classification algorithm. After predicting the students' employability, their strengths and weaknesses were analyzed based on the scores they have obtained in the assessment test. After comparing the accuracies of the five classification algorithms, the Random Forest Algorithm garnered the highest accuracy of 99% which implies that Random Forest Algorithm is best suited for assessing students' employability based on the given data sets.

The study made by [1] used six hundred thirty-three (633) data sets from numerous sources: examination unit, graduation information and curriculum unit in classifying graduates' employment. The variables used in this research include gender, program, co-curriculum, academic, CGPA, employment status and the number of semesters the student has spent in the college. The data set from excel were converted to comma-separated values file (.csv) for the data to load in the WEKA software. The data are interpreted to classification models wherein it showed that the Logistic Regression Model obtained the highest accuracy of 92.47%, the study concluded that the Logistic Regression Algorithm is the best in classifying the presented dataset.

Another study concerning graduates employability used Naive Bayes, C4.5, Simple Chart, Logistic Regression and Chaid Algorithm in generating data models. These models were used to determine the correlation between significant variables and the employability of Information Technology graduates. The data set comprises of the students' primary information such as location and gender; and grades in the subjects IT Core, IT Electives, IT Professional, Free Electives, Language and Humanities, Mathematics, and Social Science. By applying the said algorithms. It was found out that the Logistic Regression gained the highest accuracy of 78.4%. It revealed that the variables Gender, IT Core, and IT Professional had significantly affected the prediction, this indicates that the higher the grade they get from these two subjects, the higher the possibility of landing to a job aligned to Information Technology [2].

A comparative analysis conducted by [6] concluded that J48 is the best algorithm in predicting graduates' employability. The data set comprises of the demographic profile, academic integration, and emotional skill parameters of 1400 students. The data set were then loaded to WEKA. Upon loading the data set, the model revealed that J48 garnered 70.19% accuracy compared to other algorithm used in the research .

Data mining can be applied in different fields; one study used Naive Bayes Algorithm in predicting employee performance. Twenty-eight variables used in the data set, which includes gender, marital status, employee status, performance score, position and termination status. The result showed that the algorithm accumulated 96.77%, which implies that the Naive Bayes Algorithm is best in predicting employee performance.

The Table 1 below shows the summary of the related literature that contains the research authors, and the problem studied, as well as the data mining techniques used and the percentage of accuracy the algorithm accumulated.

Table 1: Related Works

Research Authors	Problem Studied	Used Data Mining Techniques	Accuracy
Nur Iman Natasha Binti A'rifian, Nur Sakinah Amirah Binti Mohd Daud, Athirah Faiz Binti Muhamad Romzi, and Nur Huda Nabihan Binti Md Shahri	A Comparative Study on Graduates' Employment in Malaysia by using Data Mining	• Decision Tree • Logistic Regression • Artificial Neural Network	• 81.34% • 80.82% • 81.52%
Yogesh Bharambe, Nikita More, Manisha Mulchandani, Dr Radha Shankarmani, and Sameer Ganesh Shinde	Assessing Employability of Students using Data Mining Techniques	• Random Forest • Decision Tree • KNN • Multi-class Ada • QDA	• 99% • 97 % • 91% • 89% • 72%
Mohd Tajul Rizal Ab Aziz and Yuhanis Yusof	Graduates Employment Classification using Data Mining Approach	• Naïve Bayes • Logistic Regression • Multilayer Perceptron • K-Nearest Neighbor • J48	• 79.40% • 92.47% • 68.91% • 70.09% • 66.53
Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, and Shaneth C. Ambat	Predicting IT Employability Using Data Mining Techniques	• Naive Bayes • J48 • Simple Cart • Logistic Regression • Chaid	• 75.33% • 74.95% • 73.01% • 78.4% • 76.3%

| Tripti Mishra, Dharminder Kumar, Sangeeta Gupta | Students' Employability Prediction Model through Data Mining | • J48
• Random forest
• Random Tree
• SMO
•   Mult ilaye r
Perceptron
•   Naiv e Baye s | • 70. 19%
• 71. 304%
• 63. 35%
• 63. 7%
• 70. 64%
• 62. 87% |
| Riyanto Jayadi, Hafizh M. Firmantyo, Muhammad T. J. Dzaka, Muhammad F. Suaidy, and Alfitra M. Putra | Employee Performance Prediction using Naïve Bayes | •   Naïv e Baye s | •   96 % |

3. METHODOLOGY

Predicting the employment alignment of a student requires several factors to be tested. Factors include the grades in their major subjects, and their first employment, these variables are essential in effectively analyzing the employment prospect of the students.

3.1 Data Preparation and Pre-processing

2018, 2017, and 2016 Bachelor of Science in Information Technology (BSIT) graduates academic records, and their first employment were the variables in this research. With the permission of the graduates and the College President, the College Registrar of Davao del Norte State College, including the Institute of Information Technology, provided the pertinent data needed in this research. There were a total of 423 graduates from 2016 to 2018; 128 graduates from 2018, 153 graduates from 2017 and 142 graduates from 2017.

3.2 Data Selection and Transformation

This section presents the exact data that were used in this research. The grades were obtained from the composite record of grades given by the Office of the Registrar, and the Institute of Information Technology gave the employment information of the graduates. To ensure that the data gathered were precise and complete, the primary author individually selected the grades and employment information from the composite record and tracer study respectively and compiled it to another excel file.

The Table 2 shows the summary of the variables with their description and possible values.

Table 2: Student Variables

Variables	Variable Description	Possible Values
grades.IT2	Student grade in the subject IT2 – Program Logic Formulation	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT1	Student grade in the subject IT 1 – IT Fundamentals	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 1- 5.0- 75 below}
grades.IT9	Student grade in the subject IT9 – Computer Programming 2	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 1- 5.0- 75 below}
grades.IT8	Student grade in the subject IT8 – PC Maintenance and Management 1	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT7	Student grade in the subject IT7 – Computer Systems Organization	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 1- 5.0- 75 below}
grades.IT6	Student grade in the subject IT6 – Discrete Mathematics	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 1- 5.0- 75 below}
grades.IT10	Student grade in the subject IT10 – Data Communications and Networking Technology	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 1- 5.0- 75 below}
grades.IT15	Student grade in the subject IT15 – Network	{1.0 – 100, 97, 98, 1.25- 97, 96, 95,

	Installations and Administration	1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 1. 5.0- 75 below}
grades.IT14	Student grade in the subject IT14 – Data Structures with File Organization	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT13	Student grade in the subject IT13 – Computer Programming 3	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
gardes.IT12	Student grade in the subject IT12 – PC Maintenance and Management 2	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT11	Student grade in the subject IT11 – Operating Systems	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}}
grades.ITEELEC1	Student grade in the subject ITEELEC1 – Project Management	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT21	Student grade in the subject IT21 - Accounting	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75,
grades.IT19	Student grade in the subject IT19 – Information Resource Management	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80,
grades.IT18	Student grade in the subject IT18 - Multimedia	2.75- 79, 78, 77, 3.0- 76, 75, {1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT17	Student grade in the subject IT17 – Object Oriented Programming	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT16	Student grade in the subject IT16 – Database Management System	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.FELEC1	Student grade in the subject ITFELEC1 – E-commerce	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.ITELEC2	Student grade in the subject ITELEC2 - Entrepreneurship	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT3	Student grade in the subject IT3 – Professional Ethics for Filipino IT Professionals	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT24	Student grade in the subject IT24 – Fieldtrip and Seminar	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT23	Student grade in the subject IT23 – Software	{1.0 – 100, 97, 98, 1.25- 97, 96, 95,

	Engineering	1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT22	Student grade in the subject IT22 – Systems Analysis and Design	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT20	Student grade in the subject IT20- Principles of Information Security	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.ITEELEC4	Student grade in the subject ITEELEC4 – Business Enterprise and Processes	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.ITEELEC3	Student grade in the subject ITEELEC3 – Human Resource Management	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.FELEC3	Student grade in the subject FELEC3 – Linux Networking	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.FELEC2	Student grade in the subject FELEC2 – Web Programming	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.CAPSTONE	Student grade in the subject IT 422 – Capstone Project 2	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83,

		2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
grades.IT25	Student grade in the subject IT 421 - Practicum	{1.0 – 100, 97, 98, 1.25- 97, 96, 95, 1.50- 94, 93, 92, 1.75- 91, 90, 89, 2.0- 88, 87, 86, 2.25- 85, 84, 83, 2.50- 82, 81, 80, 2.75- 79, 78, 77, 3.0- 76, 75, 5.0- 75 below}
IT Related Job	Students' first job after completing the degree	YES NO

3.3 Implementation of Data Mining

The classifiers were created by utilizing the Waikato Environment for Knowledge Analysis (WEKA) software. This software can analyze the data set easily by converting the excel file to Attribute-Relation File Format (ARFF) file type. In addition, several formats like CSV file, CSV: Comma-Separated Values were also applicable [3]. The data loaded to WEKA is classified by using the Naive Bayes and C4.5 Algorithm to assess its accuracy and identify the significant variables that affect graduates' employment.

3.4 Naïve Bayes Classifier

Naïve Bayes Classifier (NBC) is used to predict graduates' employment based on their grades and first employment. NBC is a simple algorithm that counts the frequency and combinations of values in a given dataset to calculate probabilities. This algorithm uses the Bayes theorem, and this treats all attributes as an independent variable [9].

3.5 C4.5 Classifier

An implementation of the C4.5 decision tree algorithm is the J48 algorithm in which the tree structure presents a flowchart of decision trees. Each internal node examines the condition of the attributes, and the outcome of the study is represented by every branch of the tree [9].

4. RESULTS AND DISCUSSION

The C4.5 and Naive Bayes algorithms were used in this research. The Table 3 below exhibits the comparative results between the two algorithms applied to the given dataset. It showed that the Naive Bayes algorithm obtained 67.22% accuracy using the 10-fold cross-validation scheme and 55.91% accuracy using 70% training and 30% testing percentage split. Further, the C4.5 algorithm accumulated 95.52% and 92.13% accuracies using the same cross-validation scheme and percentage split, respectively, which proves that the best algorithm in analyzing the employment alignment of the IT students is the C4.5 Algorithm. In determining the most suitable algorithm to examine the graduate's employment alignment, the accuracy, precision for yes, precision for no, recall, and f-measure

which was generated by the WEKA software were instumental.

Table 3: Results of C4.5 and Naïve Bayes Algorithm

Criteria	Cross-validation (10 folds)		Percentage Split (70%)	
	C4.5	NBC	C4.5	NBC
Accuracy	95.52 %	67.22%	92.13%	55.91%
Precision for Yes	0.988	0.637	0.903	0.548
Precision for No	0.932	0.698	0.938	0.565
Recall	0.955	0.672	0.921	0.559
F-Measure	0.955	0.671	0.921	0.546

The Figure 1 below displays the accuracy of the algorithms in graphical representation.
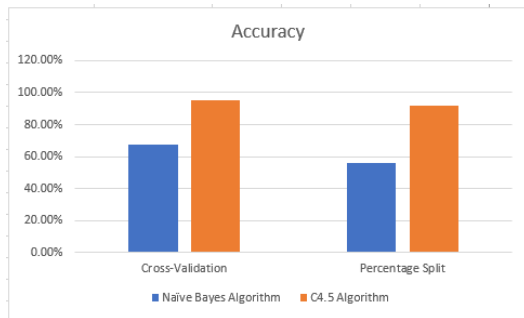
Figure 1: Accuracy percentage of the C4.5 and Naïve algorithms

The area under Receiver Operating Characteristic (ROC) curve can measure the quality of a probabilistic classifier (ROC Curve, Lift Chart and Calibration Plot). The primary purpose of ROC curves is to determine the statistical methods discrimination capability that links traces, results and other data for prediction. [10] stated that the curve should be close to one on the y-axis for it to be considered as ideally classified. However, in Figures 2 and 3, it can be seen that the curve is far from one; this indicates that the Naive Bayes Algorithm is not ideal for classifying the given data set. Meanwhile, Figures 4 and 5 shows that the curve is close to one, which implies that C4.5 is excellent in classifying the given data set.
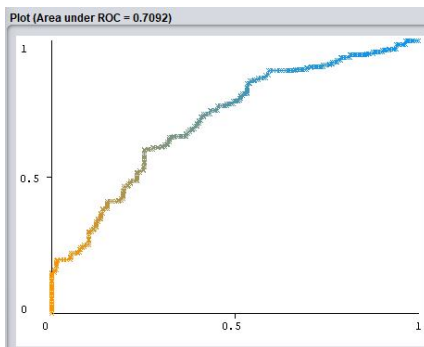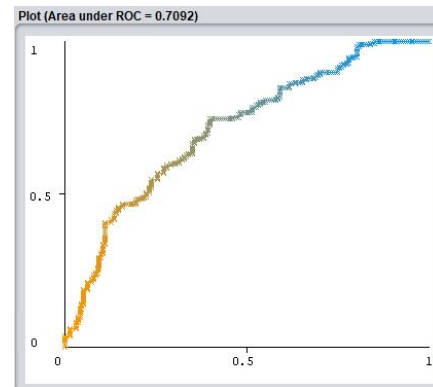

Figure 2: Naïve Bayes ROC for YES.

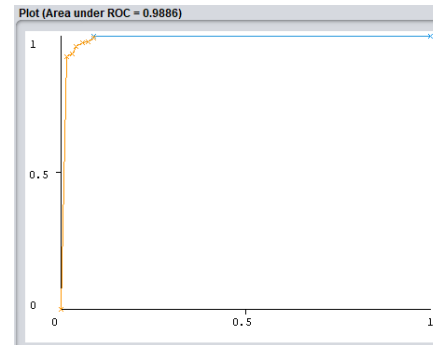**Figure 3:** Naïve Bayes ROC for NO.
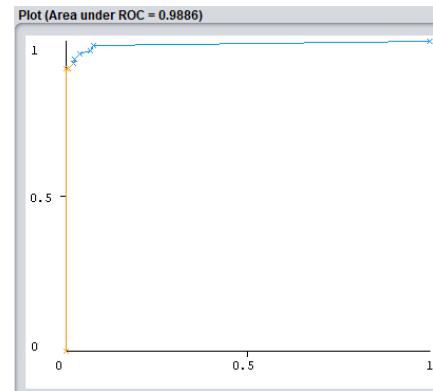

Figure 4: C4.5 ROC for YES.

Figure 5: C4.5 ROC for NO.

Figure 6 and Figure 7 exhibit the classifier errors obtained after loading the data set into the WEKA software. Classifier errors are typically placed in a confusion matrix where correct predictions can be found on the diagonal, while the prediction errors are located in off-diagonal cells. Associating the off-diagonal elements of two confusion matrices would enable comparing the error patterns of two different conditions [11]. Figure 6 illustrates that there are 175 incorrect predictions which is much higher than the number of correct predictions using Naive Bayes Algorithm. Figure 7, on the other hand, shows that there were only a few inaccurate predictions obtained by using the C4.5 classification.
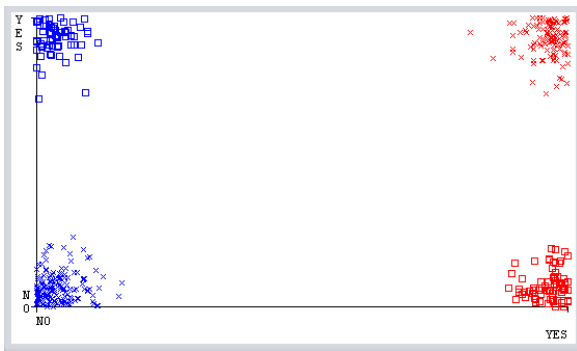
Figure 6: Naïve Bayes Classifier Errors.

Figure 7: C4.5 Classifier Errors.

The gain ratio value determines the factors that affect graduates' employment. Seventeen attributes significantly affected the graduates' work after completing a degree in Information Technology, as shown in Table 4. We can infer that these subjects would significantly affect the result of determining the employment of a student.

Table 4: Gain Ratio Attribute Evaluation

Attribute	Value
IT13	0.1664
IT 19	0.1617
IT18	0.1313
IT17	0.1286
IT7	0.1196
IT9	0.1196
IT24	0.1042
ITEELEC3	0.0801
IT3	0.0542
IT6	0.0447
FELEC1	0.0436
ITEELEC1	0.0417
IT11	0.0386
IT22	0.0365
ITEELEC4	0.0331
IT15	0.0266
IT14	0.0234

5. CONCLUSION AND RECOMMENDATION

This study has shown that universities can apply data mining techniques in determining the correlation between students' grades in their major subjects and their employment right after completing a degree in Information Technology.

Simulation results revealed that the grades accumulated by the graduates in the primary subjects of the Bachelor of Science in Information Technology would significantly affect the employment alignment of the graduates. Institutions should focus more on their student's performance to make sure that they would land to a job related to the course they have completed.

This paper also aims to determine the best algorithm in analyzing the graduates' data. The experimental results showed that C4.5 classification with the accuracy percentage of 95.52 % is most suitable in analyzing the given data. The use of other algorithms to generate more significant models, as well as the addition of more attributes to identify other factors affecting graduate's employment is recommended.

REFERENCES

[1]   M. T. R. A. Aziz and Y. Yusof, "Graduates employment classification using data mining approach," *AIP Conf. Proc.*, vol. 1761, 2016, DOI: 10.1063/1.4960842.

[2]   K. C. Piad, M. Dumlao, M. A. Ballera, and S. C. Ambat, "Predicting IT employability using data mining techniques," *2016 3rd Int. Conf. Digit. Inf. Process. Data Mining, Wirel. Commun. DIPDMWC, 2016*, no. January 2014, pp. 26–30, 2016, DOI: 10.1109/DIPDMWC.2016.7529358.

[3]   B. Jantawan and C. Tsai, "A Classification Model on Graduate Employability Using Bayesian Approaches: A Comparison," *Int. J. Innov. Res. Comput. Commun. Eng. (An ISO Certif. Organ.*, vol. 3297, no. 6, pp. 4584–4588, 2007.

[4]   M. A. Yehuala, "Application Of Data Mining Techniques For Student Success And Failure Prediction The Case Of Debre Markos University," *Int. J. Sci. Technol. Res.*, vol. 4, no. 4, pp. 91–94, 2015.

[5]   N. Premalatha and S. Sujatha, "A comparative study on students placement performance using data mining algorithms," *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 1806–1812, 2019.

[6]   T. Mishra, D. Kumar, and S. Gupta, "Students' Performance and Employability Prediction through Data Mining: A Survey," *Indian J. Sci. Technol.*, vol. 10, no. 24, pp. 1–6, 2017, DOI: 10.17485/ijst/2017/v10i24/110791.

[7]   N. I. N. Binti A'Rifian, N. S. A. Binti Mohd Daud, A. F. B. Muhamad Romzi, and N. H. N. Binti Md Shahri, "A comparative Study on Graduates' Employment in Malaysia by using Data Mining," *J. Phys. Conf. Ser.*, vol. 1366, no. 1, 2019, DOI: 10.1088/1742-6596/1366/1/012120.

[8]   Y. Bharambe, N. More, M. Mulchandani, R. Shankarmani, and S. G. Shinde, "Assessing employability of students using data mining techniques," *2017 Int. Conf. Adv. Comput. Commun.*

Informatics, ICACCI, 2017, vol. 2017-January, pp. 2110–2114, 2017, DOI: 10.1109/ICACCI.2017.8126157.

[9]   M. Suljic, "Data Mining Approach for Predicting Student Performance," *Econ. Rev. J. Econ. Bus.*, vol. X, no. 1, pp. 3–12, 2012.

[10]   M. Vuk and T. Curk, "ROC curve, lift chart and calibration plot," *Metod. Zv.*, vol. 1, no. 3, pp. 89–108, 2006.

[11]   E. Olivetti and D. B. Walther, "A Bayesian Test for Comparing Classifier Errors," *Proc. - 2015 Int. Work. Pattern Recognit. NeuroImaging, PRNI, 2015*, no. June 2015, pp. 69–72, 2015, DOI: 10.1109/PRNI.2015.11.