International Journal of Information Sciences and Application.
ISSN 0974-2255 Volume 10, Number 1 (2018), pp. 7-21

http://www.irphouse.com

Predicting Academic Performance of Information Technology Students using C4.5 Classification Algorithm: A Model Development

Las Johansen B. Caluza and Jasten Keneth D. Trecene

Leyte Normal University, Tacloban City, Philippines

Abstract

Nowadays, large amount of data are stored in different databases and is increasing rapidly. These databases contain data that can be useful for predicting students' academic performance and will help the University for improvement. Educational Data Mining used to study the available data stored in the University database and create new knowledge out of it. C4.5 (J48) classification algorithm were applied to create a decision tree model that will predict the academic performance of Information technology students of the Leyte Normal University. The result of the decision tree predicted the possible students who will have the chance to graduate or do not have the chance based on their historic data and this will help the teacher to provide appropriate inputs to help the failing students.

Keywords: C4.5, J48, Machine Learning, Decision Tree, Student Performance, Information Technology, Philippines.

1. INTRODUCTION

The famous saying of John Naisbitt "We are drowning in data, but starving for knowledge" leads to a question "Now that we have gathered so much data, what do we do with it?". While database technology has provided us the basic tools for efficient storage and lookup of large data sets, the issue of how to help humans understand and analyze large bodies of data remains difficult. To deal with the data glut, a new generation of intelligent tools for automated data mining and knowledge discovery is needed. This concept and methods are widely used in marketing, decision making, especially in educational research.

Educational Data Mining (EDM) is a technique in data mining that is widely used for educational research, this will help identify patterns that is useful for predicting the

academic performance of the students (Borkar & Rajeswari, 2013). Students' academic performance is vital for educational institutions wherein it is used for strategic planning to improve and maintain the quality of studies of the students.

Bharadwaj and Pal (2011) conducted a study on the student performance based by selecting 300 students from the program of Bachelor of Science in Computer Application in Awadh University Faizabad, India. By using the Bayesian Classification Algorithm, it was found out that factors like student's grade in the senior secondary exam, current address, and medium of teaching, mother's qualification, student's habit, family's annual income and the student's family status were highly correlated with the student's academic performance.

In addition, a study on performance prediction of engineering students using decision trees conducted by Kabra and Bichkar (2011) shows that students past academic performance can be used to create a model using decision tree algorithms that can be used for predicting the academic performance of the student. Based on the confusion matrix, it was believed the model that were successfully identified which students were more likely to fail. Moreover, the result would be more likely to improve if more attributes were added and consider more instances.

Using two modeling methods in data mining; artificial neural network and decision tree, Baha and Emine (2012) compared the achievement of the students in the computer engineering department at Karabuk University using the criteria such as gender, age, type of high school graduated, and whether the student is studying in distance learning or regular education. Decision tree algorithm produced better prediction result using a 10 fold holdout dataset. The result also revealed that as the age of the student increases the success score decreases. Moreover, the researchers also found out that students in regular education have a higher success rate than in distance learning.

Osmanbegović and Suljić conducted a study on predicting student performance using data mining approach in which the researchers used three (3) supervised algorithms. They conducted some experiments to determine the prediction accuracy, correctly and incorrectly classified instances and the learning time. Yadav and Pal (2012) conducted a study on predicting the performance improvement of engineering students of VBS Purvanchal University, Jaunpur. It was concluded that students were likely to fail based on the students' related variables in first year engineering exam and it was observed that ID3 and C4.5 decision tree in the best algorithm. Tiwari, Singh and Vimal (2013) conducted a study on engineering students to evaluate the performance by using data mining techniques and applying 3 methods such as association, classification and clustering. The variables used are assigned, attendance, Sessional Marks, GPA, and current final grade from the database management system course. The result predicted that if a student is poor in attendance and assignment, then the grades are poor. The researchers believe that data mining is helpful in higher education, especially in engineering students wherein new knowledge is discovered.

Thus, this study will present the concept of data mining to predict the academic performance of the Information Technology students of Leyte Normal University.

2. FRAMEWORK OF THE STUDY

Conceptual Framework

Figure 1. Conceptual framework of the study

This study will utilize the Input-Process-Output approach; this will help in determining, processing and developing new information. The data of the 3rd year students specifically under the BSIT program will be taken from the database of the University. Data mining, specifically classification using C4.5 algorithm will be used to process the data to create a pattern and develop new knowledge.

3. STATEMENT OF THE PROBLEM

The researchers, aimed to predict the academic performance of Information Technology Students of Leyte Normal University through data mining techniques, specifically this study sought to answer the following questions:

3.1 What will be the students' academic performance predicted C4.5 (J48) model from the given data sets?

3.2 What features in the current available data are the strongest predictors of students' performance?

3.3 What patterns will be identified in the available data that could be useful for predicting BSIT students' performance at the university based on their previous grades.

3.4 What is the confusion matrix based on the data sets?

3.5 What is the accuracy rate of the developed model for Information Technology Students of Leyte Normal University?

4. PROCESS

4.1 Data mining

Data mining also known as Knowledge Discovery in Data (KDD) is the practice of sifting through very large amount of data in a particular database for a useful information (Data Mining Concepts, N.D.). Furthermore, it is used to discover patterns and trends that go beyond simple analysis. Data mining utilizes advanced mathematical calculations and algorithms to fragment the data and able to predict future events. Its concepts and methods can be applied in various fields especially in research. Further, Data mining will describe the dataset in more concise and summarized manner and will present interesting data into a more interpretable or understandable format.

4.2 Classification

Classification is a data mining technique that maps data into predefined groups or classes (Adhatrao, et.al., 2013). It is a supervised learning method that requires labeled training data to generate rules for classifying test data into predetermined groups or classes (Dunham, 2003). It is a two-phase process. The first phase is the learning phase, where the training data is analyzed; and classification rules are generated. The second phase is the classification phase, where test data is classified into classes according to the generated rules.

4.3 C4.5 Algorithm

The C4.5 algorithm is from Quinian Ross in 1993 and was based on ID3 known in WEKA as J48, J for Java (Dai, Zhang & Wu, 2016). When applied to a single decision tree algorithm, c4.5 has high classification accuracy rate and speed. The C4.5 algorithm will be used in this study to generate a decision tree which will be used for classifying patterns and is also referred to as statistical classifier.

4.4 Work Methodology

Predicting the academic performance of the students need lots of parameters to be considered. Data pertaining to students' grades in each subject will play a role in predicting the performance. In this study, the researcher will consider data from Information Technology students of Leyte Normal University.

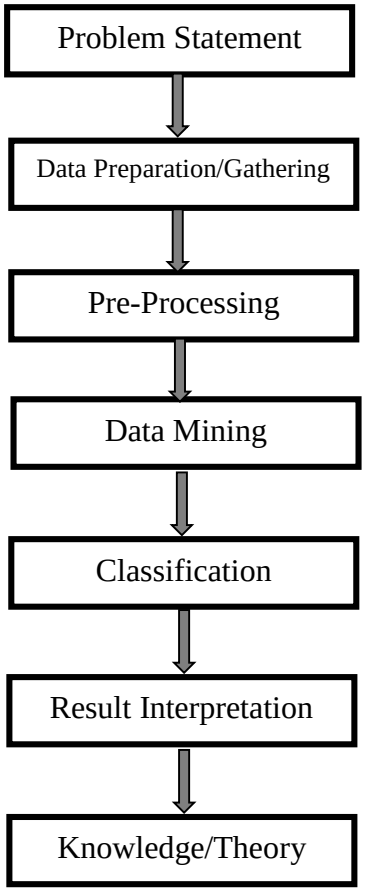

Figure 2. Work Methodology

A. Data Preparation

The data set that used in this study was obtained from the database of Leyte Normal University on the sampling method for 3rd year Information Technology Students for the past school years to carry out the simulations.

B. Data selection and transformation

After collection of data, the data set will be prepared to apply the data mining techniques. Before applying the prescribe algorithm, data preprocessing will be applied to measure the quality and suitability of the data.

C. Data Description

The data used were extracted from the database of the University from the 3rd Year BSIT students.

Br	Variable	Description	Domain	Br	Variable	Description	Domain
1	Sex	M/F	*{M, F}*	2	Location	Urban/Rural	*{Rural, Urban}*
3	G_Eng	The Mean of all the grades in English subjects of the student	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	4	G_Math	The Mean of all the grades in Mathematics subjects of the student	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
5	G_Sci	The Mean of all the grades in Science subjects of the student	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	6	G_Fil	The Mean of all the grades in Filipino subjects of the student	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
7	G_SocSci	The Mean of all the grades in Social Science subjects of the student	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	8	G_Hum	The Mean of all the grades in Humanities subjects of the student	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
9	G_PE	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	10	IT_101	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
11	IT_102	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5	*{E, VG, Good, Fair, Con,*	12	IT_103	Excellent - 10 Very Good 1.1 – 1.5	*{E, VG, Good, Fair, Con, Failure,*

		Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*Failure, WDR, DR, INC, NA, NG, NYT}*			Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*WDR, DR, INC, NA, NG, NYT}*
13	IT_104	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	14	IT_105	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
15	IT_201	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	16	IT_202	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
17	IT_203	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5	*{{E, VG, Good, Fair, Con,*	18	IT_204	Excellent - 10 Very Good 1.1 – 1.5	*{E, VG, Good, Fair, Con, Failure,*

		Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*Failure, WDR, DR, INC, NA, NG, NYT}*			Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*WDR, DR, INC, NA, NG, NYT}*
19	IT_205	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	20	IT_206	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
21	IT_207	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	22	IT_208	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
23	IT_301	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5	*{E, VG, Good, Fair, Con,*	24	IT_303	Excellent - 10 Very Good 1.1 – 1.5	*{E, VG, Good, Fair, Con, Failure,*

		Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*Failure, WDR, DR, INC, NA, NG, NYT}*			Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*WDR, DR, INC, NA, NG, NYT}*
25	IT_302	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	26	IT_305	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
27	IT_304	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	28	IT_307	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
29	IT_306	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5	*{E, VG, Good, Fair, Con,*	30	IT_309	Excellent - 10 Very Good 1.1 – 1.5	*{E, VG, Good, Fair, Con, Failure,*

		Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*Failure, WDR, DR, INC, NA, NG, NYT}*			Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*WDR, DR, INC, NA, NG, NYT}*
31	IT_308	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	32	IT_310	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*
33	IT_311	Excellent - 10 Very Good 1.1 – 1.5 Good - 1.6 – 2.5 Fair - 2.6 – 3.0 Conditioned - 4.0 Failure - 5.0 WDR - Withdrawn Subject DR – Dropped INC – Incomplete NA – No Attendance NG – No Grade NYT – Subject Not yet Taken	*{E, VG, Good, Fair, Con, Failure, WDR, DR, INC, NA, NG, NYT}*	34	CapsT	*Class Attribute* – Students who were enrolled in capstone project subject. Pass – enrolled Fail – Not enrolled	*Pass, Fail*

D. Mining Model

Weka is an open source software that implements a large collection of machine learning algorithms for data pre-processing, classification, regression, clustering and

association rules and is widely used in data mining application (Borkar & Rajeswari, 2013). It uses dataset external representation format (ARFF files).

Machine learning algorithm such as the C4.5 decision tree algorithm can learn effective predictive models from the students' data accumulated from the previous years.

5. RESULTS AND DISCUSSION

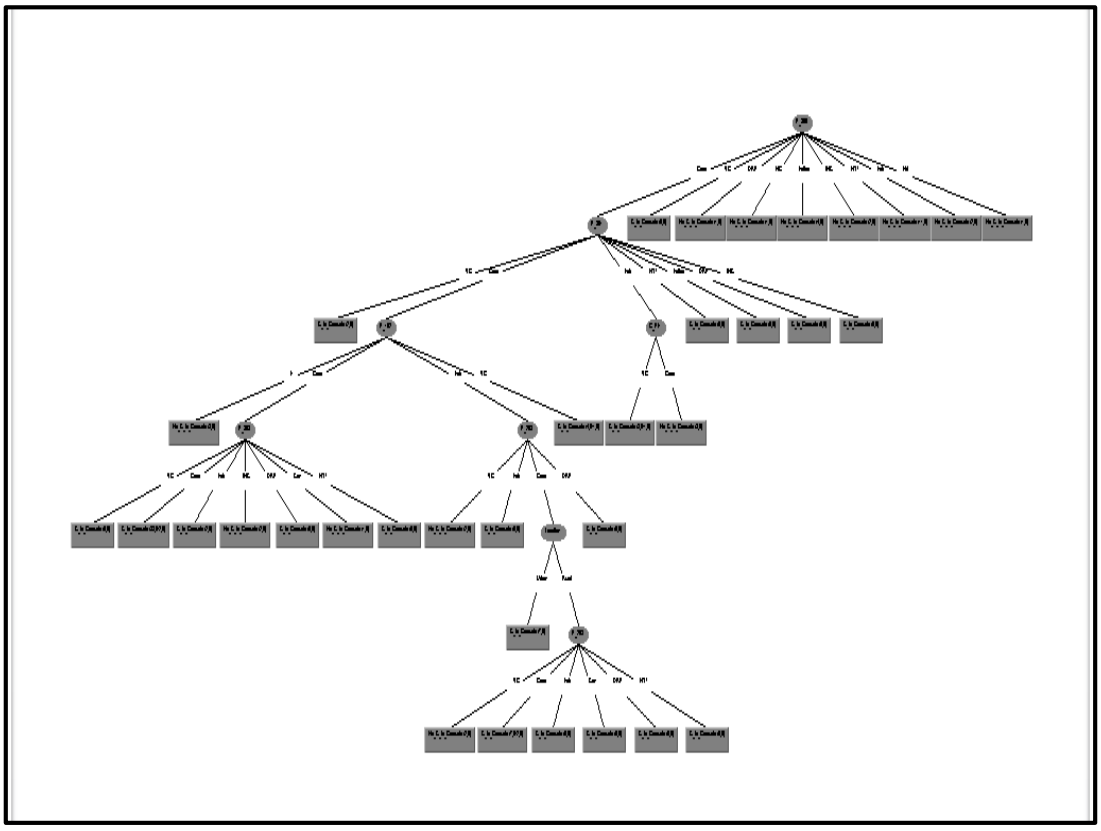

Figure 3. Decision tree model

A tree –based J48 prediction model for the student performance was constructed using C4.5 (J48) algorithm is shown in Figure 3. Attribute IT_309 is the strongest predictor in predicting student's performance. Other attributes like the sex, G_eng, G_Math, G_Sci, G_Fil, G_Socsci, G_Hum, IT_101, IT_103, IT_104, IT_105, IT_201, IT_202, IT_204, IT_205, IT_206, IT_207, IT_302, IT_305, IT_304, IT_307, IT_306, IT_310 and IT_311 are not appearing in the decision tree indicating less relevance of the prediction with such attributes.

Table 2. rules generated for the J48 pruned tree

```
IT_309 = Good
|  IT_301 = VG: C_to_Graduate (2.0)
|  IT_301 = Good
|  |  IT_102 = E: No_C_to_Graduate (3.0)
|  |  IT_102 = Good
|  |  |  IT_303 = VG: C_to_Graduate (0.0)
|  |  |  IT_303 = Good: C_to_Graduate (33.0/2.0)
|  |  |  IT_303 = Fair: C_to_Graduate (2.0)
|  |  |  IT_303 = INC: No_C_to_Graduate (2.0)
|  |  |  IT_303 = DRP: C_to_Graduate (0.0)
|  |  |  IT_303 = Con: No_C_to_Graduate (1.0)
|  |  |  IT_303 = NYT: C_to_Graduate (0.0)
|  |  IT_102 = Fair
|  |  |  IT_203 = VG: No_C_to_Graduate (2.0)
|  |  |  IT_203 = Fair: C_to_Graduate (0.0)
|  |  |  IT_203 = Good
|  |  |  |  Location = Urban: C_to_Graduate (7.0)
|  |  |  |  Location = Rural
|  |  |  |  |  IT_208 = VG: No_C_to_Graduate (2.0)
|  |  |  |  |  IT_208 = Good: C_to_Graduate (7.0/2.0)
|  |  |  |  |  IT_208 = Fair: C_to_Graduate (0.0)
|  |  |  |  |  IT_208 = Con: C_to_Graduate (0.0)
|  |  |  |  |  IT_208 = DRP: C_to_Graduate (0.0)
|  |  |  |  |  IT_208 = NYT: C_to_Graduate (0.0)
|  |  |  IT_203 = DRP: C_to_Graduate (0.0)
|  |  IT_102 = VG: C_to_Graduate (4.0/1.0)
|  IT_301 = Fair
|  |  G_PE = VG: C_to_Graduate (3.0/1.0)
|  |  G_PE = Good: No_C_to_Graduate (3.0)
|  IT_301 = NYT: C_to_Graduate (0.0)
```

```
|  IT_301 = Failed: C_to_Graduate (0.0)
|  IT_301 = DRP: C_to_Graduate (0.0)
|  IT_301 = INC: C_to_Graduate (0.0)
IT_309 = VG: C_to_Graduate (5.0)
IT_309 = DRP: No_C_to_Graduate (1.0)
IT_309 = NG: No_C_to_Graduate (1.0)
IT_309 = Failed: No_C_to_Graduate (4.0)
IT_309 = INC: No_C_to_Graduate (2.0)
IT_309 = NYT: No_C_to_Graduate (11.0)
IT_309 = Fair: No_C_to_Graduate (2.0)
IT_309 = NA: No_C_to_Graduate (1.0)
```

The table above shows the patterns identified in the available data that could be useful for predicting BSIT students' performance at the university based on their previous grades.

Table 3. Confusion matrix based on the data set

a	**b**	Classified as	
57	0	**a**	C_to_Graduate
6	35	**b**	No_C_to_Graduate

From the confusion matrix presented in Table 3, out of 57 students who have the chance to graduate, 57 were classified to have the chance and none were identified not to have the chance. Out of 41 students who don't have the chance to graduate, 35 were classified not to have the chance and 6 were classified to have the chance. This means that the algorithm was able to classify correctly all the students who have the chance.

Table 4. Summary of correctly classified and incorrectly classified instances

Correctly classified instances	**92**	**93.8776 %**
Incorrectly classified instances	6	6.1224 %

Table 2 summarizes the correctly classified and incorrectly classified instances. 92 were classified correctly and 6 or 6.1224 % instances were classified incorrectly

Table 3. Detailed accuracy by class

Class	*TP Rate*	*FP Rate*	*Precision*
Pass	1.000	0.146	0.905
Fail	0.854	0.000	1.000
Weighted Avg.	0.939	0.085	0.945

A detailed table showing the True Positive Rate (TP), False Positive Rate (FP) and the Precision.

From the classifier accuracy as shown in Table 3, it is clear that the true positive rate of the model for the students who have the chance to graduate is 1.000. In addition, the accuracy of the model is 93.8776 %, which means that the model is successfully identifying students who are likely to have the chance to graduate.

CONCLUSION

This study shows that students past academic performance was used to generate the model using C4.5 (J48) Decision tree algorithm that can be used for prediction of a student's academic performance. The result achieved from applying selected algorithms for classification on the data set reveals that the accuracy rate of the prediction is 93.8776%. Moreover, the student will not have the chance to graduate if the grade in IT_309 and IT_300 is failed and with deficiencies. However, the current location of the student does not vary in their academic performance, whether the student is living in urban or rural area.

REFERENCES

[1] Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., & Honrao, V. (2013). Predicting students' Performance using ID3 and C4. 5 classification algorithms. *ArXiv preprint arXiv:1310.2071*. Available from https://arxiv.org/ftp/arxiv/papers/1310/1310.2071.pdf

[2] Sen, B., & Ucar, E. (2012). Evaluating the achievements of computer engineering department of distance

[3] Education students with data mining methods. *Procedia Technology*, *1*, 262-267. Available at http://ac.els-cdn.com/S2212017312000540/1-s2.0-S2212017312000540-main.pdf?_tid=830f58e8-4049-11e7-b186-00000aab0f27&acdnat=1495607193_724aaf12a417650a21097f25300f9d71

[4] B.K. Bharadwaj and S. Pal. "Data Mining: A prediction for performance improvement using

[5] Classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[6] Borkar, S., & Rajeswari, K. (2013). Predicting students academic performance using education

[7] Data mining. *International Journal of Computer Science and Mobile Computing (IJCSMC)*, *2*(7), 273-279.

[8] Dai, Q. Y., Zhang, C. P., & Wu, H. (2016). Research of Decision Tree Classification Algorithm In Data Mining. *International Journal of Database Theory and Application*, *9*(5), 1-8. Available at http://www.sersc.org/journals/IJDTA/vol9_no5/1.pdf

[9] Data Mining Concepts (N.D.). Data Warehousing and Business Intelligence Retrieved from https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#DMCON002 on April 11, 2017.

[10] Dunham, M. H. (2003). Data Mining: Introductory and Advanced Topics. Pearson Education (Singapore) Pte.

[11] Kabra, R. R., & Bichkar, R. S. (2011). Performance prediction of engineering students using Decision trees. *International Journal of Computer Applications*, *36*(11), 8-12..Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.736.9&rep=rep1&type=pdf

[12] Salzberg, S. (1994). Book review: C4. 5: by j. Ross Quinlan. Inc., 1993. Programs for machine Learning, Morgan Kaufmann publishers. *Machine Learning*, *16*, 235-240.

[13] Singh, R. (2013). An Empirical Study of Applications of Data Mining Techniques for Predicting Student Performance in Higher Education. Available at http://d.researchbib.com/f/cnnJcwp21wYzAioF9xo2AmY3OupTIlpl9TMJWlqJSlrGVjZGZiIwWWZwVjZGZkZP5jMTL.pdf

[14] Yakab, 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,* 2(2), 51-56.