International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

An Application of J48 Classification Algorithm in Predicting Students' Academic Performance

Hershey R. Alburo

***Abstract:** This paper sets out to use J48 classification algorithm to predict students' academic performance towards the end of the semester in the Data Structure course under the Computer Science Program. This algorithm aimed to help faculty in forecasting who among the students would likely to fail and who would make it until the end of the semester. In this way, the faculty could make remedial measures to help those struggling students pass the subject and advance to the next level, thus, increasing students' success rate and retention in a Higher Education Institutions (HEI). This research employed a descriptive correlational design using Exploratory Data Analysis (EDA) for Data Mining in testing and verifying data to generate new information. Data mining is part of the Knowledge Discovery in Databases (KDD) process where it follows six steps: data selection, data pre-processing, data transformation, data mining, interpretation, and knowledge discovery. Step 1 includes gathering and selecting data for the study and for this purpose, a total of 103 students' records were collected from the instructors for a period of two semesters, S.Y. 2014 -2015 & 2015 – 2016. Different evaluative criteria contained in the class records were utilized as attributes in predicting students' academic performance. Steps 2 and 3 is pre-processing and transforming the data where it involves discarding those students who dropped/withdrawn from the semester, and converting the excel file into a comma separated values or .csv file, respectively. After these steps, step 4 or the application of J48 classification algorithm was utilized to discover classification rules. Step 5 refers to the tree visualization results where it identified the strongest predictor that most likely influence the students' final average grade. Finally, step 7 shows the extracted information from the tree or the extracted rules that can be used by the administration, faculty and other stakeholders to improve the academic performance of the students. In particular, they might consider redesigning and restructuring teaching pedagogies to assist and focus more on struggling students.*
***Keywords** : Academic Performance, Classification Algorithm, Educational Data Mining, J48 Decision Tree*

I. INTRODUCTION

Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.

Revised Manuscript Received on February 28, 2020.
* Correspondence Author
Hershey R. Alburo*, Faculty, College of Computer Studies, State-Owned University in Eastern Samar, Philippines,

Data mining is extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data [1]. The availability of these large amount of data prompted the researcher to transform raw data into useful information that could be used by educators to take informed decisions and appropriate actions using EDM.

Al-Barrak and Al-Razgan [2] conveyed that many universities have been using data mining techniques to analyze educational report stored in the educational institute such as enrolment data, students' performance, teachers' evaluations, gender differences, and many others. Classification, clustering and association rules are just some of the numerous techniques in data mining that can be applied to these data.

Scholars [3], believed that once these data have been correctly analysed, this will help educational institutions assess, evaluate, plan, and decide their educational programs. This new knowledge is expected to reveal hidden patterns that will assist academic programs utilize resources more effectively.

Academic institutions in general, typically evaluate students based on examinations, activities, assignments, projects and quizzes carried out during the semester. Other programs have the so-called inner assessment which include quizzes, midterm, lab work, projects, etc. and outer assessment which is based on their final scores [4].

This setting is almost similar in the case of ESSU Salcedo particularly the Computer Science program. Assessment of students is based on major examinations (prelim, midterm, prefinal and final), laboratory exercises, quizzes, projects and assignments in a particular subject. A student may have a passing or a failing grade depending on his/her average grade in the enrolled subject. Thus as educators, it is important to have prior knowledge on those students who are likely to fail the subject so there could be remedial efforts taken by teachers and students as well.

In this paper, the researcher applied one of the data mining techniques named classification to predict students' performance in one of the most difficult subjects like Data Structures in the Computer Science program. Classification is one of the supervised learning techniques that build a model to classify a data item into a predefined class label. The aim is to predict future output based on the available data. Data Structures was considered by the researcher as a subject under study because of its high rate of failure. Data were gathered via obtaining the class records from a faculty handling the subject during the second semesters of the school year 2014-2015 & 2015-2016 with 103 students enrolled in the course.

Since most of the studies focused more on differentiating and evaluating different classification algorithms in the different areas of concern like health and education, other authors used socio-demographic profile of students to predict students' performance [5], and still others used data for female students, the contribution of this paper is to utilize J48 classification algorithm (based on the performance comparison done by several authors) in generating a predictive model to forecast students' success or failure in the course using both male and female records in a higher education [6]. Moreover, the researcher found out that there exists only a handful of literature in this area in the Philippines.

Guided by the concepts and findings of various studies, this research aimed to use classification algorithm to build a model that will help predict students' performance in the Data Structure subject and would eventually provide useful insights to educators in re-designing pedagogies for struggling students.

B. Objective of the Study

This research aimed to generate a predictive decision tree model that will forecast the highly influencing factors affecting students' performance in the Data Structure course using J48 Classification Algorithm.

C. Research Framework

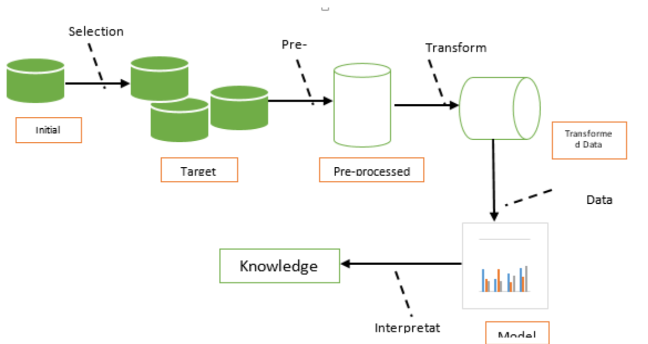

Figure 1. Knowledge Discovery from Data (KDD) Process

Figure 1 is the methodological framework adopted by the researcher to carry out the procedure in the study. [7] described the KDD process as it includes selecting the data needed for data mining process. Pre-processing consists finding incorrect or missing data. It also includes noise or outliers, collecting necessary information to model or account for noise, accounting for time sequence information and known changes. The next step is the transformation wherein it is concerned on converting data into a common format for processing. In this case, the researcher after categorizing some data, it was encoded or transformed into more usable format like .csv file format. Data mining is the task performed to generate the desired result or outcome. Interpretation/evaluation involves presenting the results to the users. This stage is important because the usefulness of the result will depend on it. This is where various kinds of representation (e.g. classification, clustering, association rule, etc.) are used in this step.

C. Scope and Limitation of the Study

The algorithm used in this study is only capable of predicting categorical values. The accuracy of prediction is only evaluated as to how high is the percentage of the attributes/values classified. Moreover, the model is just suggestive which means other criteria or attributes must be included and carefully measured to achieve more accurate results.

II. METHODOLOGY

A. Research Design

This research utilized a descriptive correlational design using Exploratory Data Analysis (EDA) for Data Mining to test and verify data to generate new information and build the classification model.

B. Research Methods

a. Building the Model

Figure 2 is the model of the study depicting how data was prepared and represented.

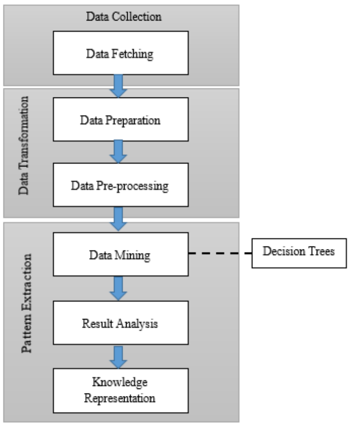

Figure 2. Building the Model of the Study

a.1.Data Collection

Students' records for the course Data Structure were collected from the faculty handling the said subject at the College of Information & Communication Technology of ESSU Salcedo Campus from School Year 2014-2015 to 2015-2016. A total of 188 records were collected. Records showed that each student had student number, student name, grades in their quizzes divided into four terms (quiz 1, quiz 2 to quiz 11), major examinations (prelim, midterm, prefinal and final exam), laboratory grade and requirement and the total average grade. Students should get at least 3.0 to pass the course. That means, 3.1 to 3.5 is considered "Conditional Grade" while 3.6 to 5.0 is "Failed".

a.2. Data Preparation

Data pre-processing is necessary prior to applying any data mining technique. Three steps were adopted by the researcher to prepare the data:

1. Eliminate records of students who withdrew from the class because some relevant values might be missing.
2. Discretize the final grade attribute to seven categories: Outstanding, Excellent, Very Good, Good, Fair, Conditional, Failed

Data discretization means that one is going to use a predefined set of intervals and grouping future values according to that interval. After pre-processing the data, it was loaded to WEKA software to apply classification algorithm. Table 1 shows the sample of the dataset used in this study.

Table I. Description of the dataset used in this study

Attribute	Description	Categorized Values
Quiz 1, 2, …13	Grades for Quiz (1, 2, …12)	None
Laboratory Grade	Laboratory Grade	None
Prelim Exam	Prelim Grade	None
Midterm Exam	Midterm Grade	None
Prefinal Exam	Prefinal Exam	None
Final Exam	Final Exam	None
Requirement	Requirement Grade	none
Final Grade	Final Average Grade	1.0 Outstanding 1.1 - 1.5 Excellent 1.6 - 2.0 Very Good 2.1 - 2.5 Good 2.6 - 3.0 Fair 3.1 - 3.5 Conditional 3.6 - 5.0 Failed

a.3. Data Classification

The goal of this research is to predict students' academic performance for the Data Structure subject. Classification is used here as it is the process of placing an object into a class or category. Decision tree is used to classify unknown samples. One of the objectives in classification is building decision trees which are flowchart-like tree structures, with internal nodes and leaf nodes. All internal nodes have two or more child nodes which denote a test on an attribute. Leaf nodes represent class labels or class distribution [8]. Decision tree generation consists of tree construction and tree pruning. All of the training examples in the beginning are at the root of the tree. Partitioning is done recursively based on the selected attributes. Tree pruning identifies and removes branches that reflect noise and outliers partitioning [9].

a.4. Decision Tree Induction

ID3 developed by J. R. Quinlan is the central algorithm in building decision trees. It employs a top-down, greedy search through the space of possible branches with no backtracking. A decision tree depicts rules for dividing data into groups. J48 (in WEKA tool) builds decision trees from a set of training data in the same way as ID3 using entropy and information gain. ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is equally divided it has entropy of one [10].

At each node of the tree, J48 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is chosen to make the decision. The J48 algorithm then recurs on the smaller sublists .[11]

The following are the conditions for stopping the partitioning [9]:

- All samples for a given node belong to the same class.
- There are no remaining attributes for further partitioning – majority voting is employed for classifying the leaf.
- There are no samples left (Apte & Weiss, 1997) (Fayyad, 1994 as cited by [9]).

a.5. Extracting Rules from the Tree

From the study of [9], below are some extraction rules from the decision trees:

1. Knowledge is represented in the form of IF-THEN rules.
2. A rule is created for each path from the root to a tree.
3. Each attribute-value pair along a path forms a conjunction.
4. The leaf nodes holds the class prediction.
5. Rules are easier for humans to understand (Kamber, et.al., 1997)

III. RESULTS AND DISCUSSION

For this experiment, the researcher used all the attributes (which are in numerical forms): quiz 1 to quiz 13, final lab grade, grades in the major exams (prelim, midterm, prefinal, final) and the final average grade which were categorized as (outstanding, excellent, very good, good, fair, or conditional).The data were encoded into the MS Excel application with a CSV (comma separated values) file format. The dataset were splitted into training set (80%) and into test set (20%). Training set was used to generate the predictive model while test set was used to validate the accuracy of the model. It is remarkable from the tree that a student should perform better on major examinations to have an improved final average grade.Figure 3 shows the pseudo code of the developed model where different attributes like Prefinal Exam, Quiz 13, Quiz 2 and Final Exam were depicted. This means that if a student obtained a grade of 1.85 and higher, likewise in quiz 13 and quiz 2 then a student would likely to receive a "very good" final average grade in the subject. However, if a student receives a grade lower than 1.85 in the Prefinal Exam and quiz 13 is lesser than 2.32, most probably the student would get just a "Fair" rating in the subject.

```
IF Prefinal Exam <= 1.85
   IF quiz 13 <= 1.71
            IF Quiz 2 <= 2.91 then Final Average Grade is "Very
Good"
      ELSE Final Average Grade is "Good"
   ELSE
             IF Final Exam <= 2.42
                   IF Quiz 4 <= 3.1 then Final Average Grade is "Very Good"
         ELSE Final Average Grade is "Good"
                     ELSE IF Final Exam > 2.42
         then Final Average Grade is "Good"
ELSE IF Prefinal Exam > 1.85
   IF Quiz 13 <= 2.32 then Final Average Grade is "Good"
   ELSE Final Average Grade is "Fair"
```

Figure 3. Decision Rules Generated from the Tree

To measure the accuracy of the classification, Table II gives the predictive performance done by the WEKA's evaluation tool. The set of measurement is derived from the training data. It shows statistics on the accurate prediction of the classifier to classify correct classes. It is evident that 92.68% \was the accuracy of the prediction which means there were only about 7% incorrectly classified instances. The mean absolute error of the probability is 0.08 and the root mean square error is .20, meaning that is the square root of the quadratic loss. Total number of the training data set after split is 82. Errors are neither 1 or 0 because it did not classify all training samples correctly.

Table II. Detailed Summary of the Classifier Output

Correctly Classified Instances	76	92.6829%
Incorrectly Classified Instances	6	7.3171%
Kappa statistic	0.8232	
Mean absolute error	0.0856	
Root mean squared error	0.2068	

Relative absolute error	28%	
Root relative squared error	53%	
Total Number of Instances	82	

Confusion matrix, reflected in Table III is a more detailed class breakdown of the classifier output. It demonstrates the actual and predicted classes done by the algorithm and can be read diagonally (Goyal & Mehta, 2012).

Table III. Confusion Matrix

a	b	c	Classification
57	0	0	a = Good
3	17	0	b = Very Good
3	0	2	c = Fair

In this study, a is the number of correct predictions that an instance is classified as "Good", b as "very good' and c as "fair". True Positive (TP) for class a is 57, class b is 17, and class is 2, while the False Positives (FP) for the classes a to c is 0, 3, and 3, respectively. The total number of true positive is 76, and the total number of False Positive is 6.

TP means that if the outcome of a prediction is n and the actual value is also n then it belongs to True Positive.

FP, on the other hand, if the actual value is not n then it is False Positive.

Table IV. Detailed Accuracy by Class

TPR	FPR	Precision	Recall	ROC Area	Class
1.000	0.240	0.905	1.000	0.919	Good
0.850	0.000	1.000	0.850	0.925	Very Good
0.400	0.000	1.000	0.400	0.461	Fair
0.927	0.167	0.934	0.927	0.904	Avg.

Reflected in Table IV are True Positive Rate (TPR), False Positive Rate (FPR), and Precision/Recall which are computed using the formula:

$$\text{TPR} = \frac{\text{diagonal element}}{\text{sum of relevant row}}$$

$$\text{FPR} = \frac{\text{non} - \text{diagonal element}}{\text{sum of relevant row}}$$

$$\text{Precision} = \frac{\text{diagonal element}}{\text{sum of relevant column}}$$

Thus, solving for class good, very good, and fair (TPR), respectively, would result to the answers below which are similar to Table IV:

$$\text{TPR} = \frac{57}{57+0} = 1$$

$$\text{TPR} = \frac{17}{17+3} = 0.85$$

$$\text{TPR} = \frac{2}{2+3} = 0.4$$

False Positive Rate gave the following results which is also comparable to Table IV:

$$\text{FPR} = \frac{6}{(3+17)+(3+2)} = 0.24$$

$$\text{FPR} = \frac{0}{57+5} = 0$$

$$\text{FPR} = \frac{0}{57+20} = 0$$

True Positive Rate recorded 100% classified correctly as "Good", 85% as "Very Good" and 40% as "Fair".

False Positive Rate indicates that a small number was incorrectly classified as either "Good", "Very Good" or "Fair".

Receiver Operating Characteristic (ROC) curve represents a performance graphing method between True Positive Rate and False Positive Rate. When ROC skewed more to 1 (as in the case of the experiment done in this research, ROC Area is 91%), which indicates that the algorithm nearly classified all attributes correctly.

Figure 4 is the predicted model of the classification algorithm. The model depicted the strongest predictor or the highly influencing variable affecting students' academic performance in the Data Structure subject. It shows that students should perform best in major examinations to get at least a very good final average grade.

Two major examinations –Prefinal exam and Final exam were portrayed in the decision tree with the former as the strongest predicted attribute. When a student got less than or equal to 1.85 as the final average grade another attribute Quiz 13 (that is in the final term of the semester) will be evaluated. If the students got 1.71 or better and Quiz 2 (in the beginning of the semester) is also better than 2.91, most likely the students' final average grade will be "Very good" otherwise "Good".

The decision tree, also showed that a particular student will get only a"good" or "fair" rating if he/she does not do well in the final exam, quiz 4 and quiz 13 otherwise he/she will obtain a "Very Good" rating.

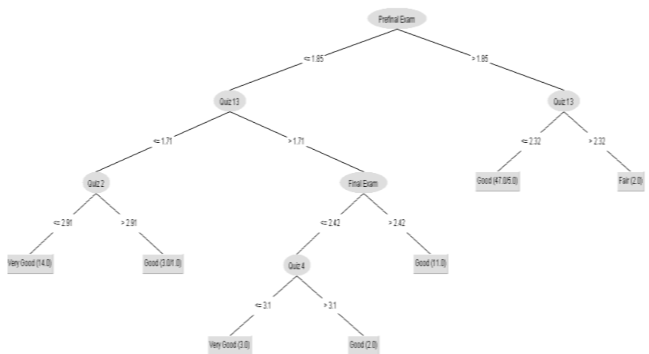

Figure 4. Predicted Model of the Study

The oval shape represents the attributes of the students' academic data in the Data

Structure subject – that is Prefinal Exam, Final Exam, Quiz 13, Quiz 2, and Quiz 4, where Prefinal Exam is called the root node. The numbers under each leaf nodes are the students final rating in the course. The squares represent the classification as either Vergy Good, Good, or Fair. The tree shows that students should have a mastery of the lessons in the prelim and midterm since exams in the prefinal and final terms are mostly taken from the previous lessons (prelim and midterm), that is why the strongest predictor from this experiment is prefinal exam which further disclosed that a stroung foundational concepts is vital in passing the subject.

IV. CONCLUSIONS

A data mining technique named classification was used in predicting students' academic performance. Specifically, this study utilized J48 classification algorithm to efficiently produce a predictive model based on historical data of the students that classified attributes into each class/category. From the predicted model, a student should have a strong foundational concepts in the learning areas of the subject so he/she could perform best in major examination to get a passing grade in the course. Major examinations (Pre-final & Final) and quizzes in the beginning of the semester were seen as contributory factors in the academic performance of the students. Otherwise, a student will likely to fail or have poor average grade if 2 of the major examinations will have a rating lower than 3.0. The result achieved 92.68% accuracy rate of the prediction after applying the classification algorithm.

V. RECOMMENDATION

The result from the study could be utilized by the administration, faculty, and other stakeholders to improve and deliver the academic requisites efficiently and effectively. Further, faculty concerned on the subject should conduct remedial classes and other measures to help students attain higher scores in major examinations. The College of ICT, in particular, might redesign and restructure teaching pedagogies to assist and focus more on struggling students. Due to time constraints, the data collected by the researcher was only limited to two semesters which might be insufficient to produce more accurate results. It is therefore suggested, that data for a period of 6 semesters or more might be considered to yield better accuracy results. Since the model is just suggestive, it is also recommended to have other attributes to help forecast students' success or failure in the course.

ACKNOWLEDGEMENT

The researcher is deeply indebted to Dr. Las Johansen B. Caluza, faculty of Leyte Normal University for sharing his time, intellect & expertise in this research study; to the Administrator and Instructors of ESSU Salcedo Campus for the assistance in all forms, and to Cebu Institute of Technology-University through Dr. Larmie Feliscuzo for the equal opportunity given to the researcher in presenting this research.

REFERENCES

1. International Data Mining Society. Available at educationaldatamining.org.
2. M. Al-Barrak & M. Al-Razgan "Predicting Students' Performance through Classification: A Case Study", Journal of Theoretical and Applied Information Technology, 20th May 2015, Vo. 75, No.2 pp. 167-175, 2015
3. M. Al-Razgan, A. Al-Khalifa, & H. Al-Khalif." Educational Data Mining: A Systematic Review of the Published Literature 2006 – 2013," in Proceedings of the First International Conference on Advanced Data and Information Engineering (Da-eng, 2013), Vol. 285, T. Herwan, M. M. Deris, and J. Abawajy, Eds. Springer Singapore, 2014, pp. 711-719.
4. Abu-Oda, & A.M. El-Halees, "Data Mining in Higher Education: University Dropout CaseStudy", International Journal of Data Mining and Knowledge Discovery process, Vol. 5, No. 1, January 2015.
5. Behrouz, et.al., "Predicting Student Performance: An Application of Data Mining Methods With The Educational Web-Based System" Lon-CAPA © 2003 IEEE, Boulder, CO., 2013
6. J. Ruby, & K. David. "Predicting the Performance of students in Higher Education Using Data Mining Classification Algorithms – A Case Study", International Journal for Research in Applied Science & Engineering Technology (IJRASET), www.ijraset.com Volume 2 Issue XI, November 2014, ISSN: 2321-9653
7. S. Aher & L.,L.M.R.J., "Data Mining in Educational System using WEKA", International Conference on Emerging Technology Trends, International Journal of Computer Applications, pp. 20-25., 2011
8. S.K. Yadav, S. Pal, & B.K. Bhardwaj, "Mining Data to Predict Students' Retention: A Comparative Study", www.researchgate.net/2012
9. A. Goyal & M. Rajni "Performance Comparison of Naïve Bayes and J48 Classification Algorithms "International Journal of Applied Engineering Research, ISSN 0973-4562, Vol. 7, No. 11, 2012
10. S. Sayad. "Introduction to Data Mining. Retrieved from: www.saedsayad.com, 2017
11. S.A. Kumar & M.N Vijayalakshmi, "Efficiency of Decision Trees in Predicting Student's Academic PerformanceD.C. Wyld, et.al. (Eds): CCSEA 2011, CS & IT 02, pp. 335-343, 2011. DOI:10.5121/csit.2011.1230, 2011
12. M. Pandey, & U.K Sharme, "A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction", International Journal of Computer Application (0975-8887), Vol. 61, No. 13, January 2013.
13. L. Ramanthan, D. Sasksham. & K. Suresh "Predicting Students' Performance Using Modified ID3 Algorithm, International Journal of Engineering and Technology, ISSN:0975-4024, Volume 5, No. 3, July 2013,pp. 2491-2497.
14. F. Siraj, & M.A. Abdoulla "Mining Enrolment Data Using Predictive and Descriptive Approaches", Knowledge Oriented Applications in Data Mining, pp. 53-72, 2007

AUTHORS PROFILE

HERSHEY R. ALBURO, is a faculty of the College of Computer Studies in a state-owned university in Eastern Samar, Philippines, and is currently pursuing Doctor of Information Technology at Cebu City. Her interests are both in qualitative and quantitative researches in computer science, information technology, and general education. Her teaching areas are in object-oriented programming, data mining, and data structures and algorithm. As a research enthusiast, she has presented various researches in local and international fora.