Predicting Student Performance in a Computer Network Certification Program: using the J48 Algorithm

Lowell A. Quisumbing

Leyte Normal University, Tacloban City, Philippines, Orcid: 0000-0002-1914-2274; lowellquisumbing@lnu.edu.ph

Abstract

Objective: Academic institutions worldwide are continually trying to improve their curricula to address the needs of industry and government. Many technology curricula emerged to satisfy well recognized IT standards and critical components. Colleges and universities are now inclined to include IT certification in their curricula. Moreover, this study aims to develop a model that will predict outcomes based on the performance of IT students in their Cisco Certification Program.

Method: The data sets used in this field are the forms of the 3rd year BSIT students enrolled in the Cisco Routing and Switching Certification Program from S.Y 2016-2017. Findings: There were 165 datasets with five variables: Pre-Test, Quizzes, Hands-on Skills Exam, Final Exam, and Remarks. The Data Mining process predicts the performance of the students in the certification program, individually. The J48 Algorithm derived from a C4.5 decision tree model. The results revealed that Quizzes had the highest instances of achieving a passing mark and became the first split between (Finals<=79.67) and (Finals> 79.67) in predicting student performance in the CISCO Certificate Program. Moreover, it showed the decision tree model Quizzes have the highest factor that students will get Passed, Failed, or Conditional marks.

Application/Improvements: In order to obtain more precise results, we can employ other reputable data mining techniques and prediction algorithms or a combination of both to assess and guide student performance in future studies.

Keywords: Academic Performance, Certification, Data Mining, J48 Algorithm, CCNA

1. Introduction

The future needs an educational transformation. Education should align the “how” and “what” of finding out by themselves and the creation of work that awaits them after they leave school. In keeping with this trend, academic institutions worldwide are continually trying to improve their curricula to address the needs of industry and government, which resulted in the emergence of IT standards and critical components. Colleges and universities are now inclined to include IT certification in their curriculums. This new trend drives academic institutions worldwide to face the challenge of developing a course of study that produces graduates with strong foundations in Information Technology and strengthens the skills that make them readily employable.

In the Philippines, higher education institutions are redesigning their curricula to include IT certifications to provide the students with a solid foundation in IT. The emphasis is to have a paradigm shift from traditional teaching strategies to outcomes-based education were graduates are expected to acquire analytical, problem-solving, and lifelong learning skills. With the revision of curricula, tertiary institutions are compelled to use different schemes and teaching methods to intensify the academic performance of students. In the current job market environment, college graduates are expected to make an impact and become productive employees immediately after joining the workforce. Therefore, a need for professionals with IT certifications is commanding. In accordance with the law, public universities in the Philippines, with Information Technology courses have

*Author for correspondence
integrated IT Certificate programs to their syllabi to address this global need.

One such programs the Cisco “Introduction to Networks” CCNA Routing and Switching (R&S) provides an inexpensive and concentrated venue for learners and professionals to develop or upgrade their skills. For example, the Cisco courses offer a curriculum designed to augment traditional technical education with hands-on learning using the latest networking technologies. The certification program utilizes performance-based evaluation systems to measure student achievement in which they must attain a certain level of competency. The CCNA R&S program also maintains high standards through interactive course content and live simulation assessments. Online courseware is made available to students and instructors using Webinars, videos on demand, and podcasts, along with supporting online course materials. Interactive course guides, instructor reference guides, and case studies provided. The program takes a pragmatic, business oriented approach to measuring success by collecting data and applying metrics to outcomes.

However, despite these innovations, not all students that take the CCNA certification course are guaranteed to pass. Among the problems is the difficulty student’s encounter with the assessments, their lack of computational thinking strategies and poor study habits. Some students also struggle to acquire the practical skills needed in setting up Computer networks and configuring Networking devices.

This results in student retention problems and significantly contributes to student turnover. As surveys on the subject show, in addition to students not passing their courses, traditional problems contribute to failures, such as poor academic performance due to lack of involvements in school programs, insufficient motivation, emotional problems and the negative influence of peers. Surprisingly, academic shortcomings are not only frustrating to the students and the guardians but as well as to the teachers. The absence of concrete learning models that predict academic performance in this area is scarce.

Over the years, students’ have devised ways on how to resolve or bring down the impact of poor academic performance in students by using Data Mining techniques - using classification, clustering, and rule mining as a pragmatic approach to extract hidden knowledge from educational data. Classification techniques, like decision trees and Bayesian networks, are employed as useful tools to foretell the student's behavior in an educational environment or predict his result in an examination. demonstrated the use of genetic algorithms for classifying students to conceive of their final grade based on features extracted from logging data in an educational web-based system trained 354 student records from a module provided by an Open University then utilized regression algorithms to predict all of the student's pass or fail marks studied 533 student samples from three Belgian Universities anticipate and classify academic success into low, medium, and high-risk using neural network and mining techniques a similarity in the sense that it created a decision tree model to estimate levels of students who studied the C++ course at Yarmouk University.

Kumar and Vijayalakshmi used decision trees to predict the outcomes of the final test of their learners to help professors identify students who needed help, thus improving their performance in the exam. Hence, all of the studies previously mentioned have proven the effectiveness of using a decision tree algorithm, coupled with mining techniques, for predicting student performance in a classroom setting. However, studies on predicting student outcomes involving Certification programs in the Philippines are absent.

To help the students succeed in passing the CCNA certification course, the researcher employs the effectiveness of the WEKA as a tool in data mining with J48Decision Tree Algorithm application to calculate the efficiency and accuracy instances generated by confusion table to foresee the academic performance of LNU IT students in their Cisco CCNA R&S Certification Class. J48 is a simple form of predicting outcomes that could quickly generate decision trees with high accuracy, but is more cost effective compared to other algorithms. The study objective is to mine the student grades from the previous Ended Classes so the teacher can identify useful patterns and relationships in their class performance. A Cross-validation and Receiving Operating Characteristics Curve (ROC) was deployed to make a graph that can envision the model and test the solution based on identified variables such as Pre-Test, Quizzes, Hands-on Skills Exam, Final Exams, and Remarks. The predictive model will generate an accurate prediction list that will identify students who tend to need the support of the teacher in the different areas in the course. The outcome is a model
that can serve as a foundation for teacher intervention in future CCNA classes.

2. **Methodology**

The study focuses on the Knowledge Discovery in Database (KDD) procedure. The procedure starts with the accumulation of data and preprocessing, followed by model construction and conclusions with a model evaluation and interpretations\(^\text{11}\).

2.2 **Data Mining Process**

2.2.1 **Data Collection**

The collected datasets of this work are from the previous ratings of the Leyte Normal University 3rd year BSIT registered students in the CCNA Routing and Switching (Introduction to Networks) Certification Program from S.Y 2016-2017. There were 165 datasets with five variables (Pre-Test, Quizzes, Hands-on Skills Exam, Final Exam, Remarks). Table 1 shows the descriptions and values of the attributes.

- **Pre-Test** - is a variable that measures the actual knowledge of the students towards computer networks. It composes ten percent of the total grade.
- **Quizzes** - are the evaluation at the end of every chapter lessons. Quizzes are a total number of chapter exam multiplied by fifty plus fifty to get the equivalent rating. It comprises forty percent of the overall score.
- **Hands-on Skills Exam** - to measure the hands-on technical proficiency of the student in routing and switching technologies. It composes twenty-five percent of the total grades.
- **Final Examination** - is the summative exam of all topics. It composes twenty-five percent of the total grades.

Table 1. The variables are defined to understand the process

| Variables | Description                  | Possible Values |
|-----------|------------------------------|-----------------|
| P         | Pre-Test                     | 0-100           |
| Q         | Quizzes                      | 0-100           |
| S         | Hands-on Skills Exam         | 0-100           |
| F         | Final Examination            | 65-100 grade    |

2.2.2 **Software used**

To apply the classification algorithm in predicting data, we used the WEKA toolkit. This toolkit provides a wide range of different data mining algorithms implemented in JAVA, including J48. Furthermore, WEKA is commonly used in educational data mining researches and for teaching purposes\(^\text{12}\).

2.2.3 **Data Preparation and Processing**

During this phase, pre-processing and collection of the student data occurred in training for the mining techniques. During the first measure, the researcher eliminated some irrelevant attributes such as the student's name, student's registration number, teacher's name, and student's schedule. Table 2 displays the raw datasets of the pre-processed Student Records. In the next step, data preparation, the students' grades were stored in MS Excel and subsequently converted to Comma Separated Values File (.CSV) file format. The .csv file was then loaded to Notepad++ wherein data cleaning eliminated some unwanted symbols like space, colon, and comma. Likewise, syntaxes like @Attribute, @Data, and @Relation were declared as variables to converge in the application of WEKA. The converted notepad++ text file generated an Attribute-Relation File Format (ARFF) file. Then the information was uploaded to the WEKA Application and the preprocessing of raw data was conducted to a more understandable file format.

Table 2. Sample datasets of student grades from the previous semester

| Pre-Test Exams | Quizzes     | Hands-on Skills Exam | Final Exam | Remarks |
|----------------|-------------|----------------------|------------|---------|
| 42.6           | 81.1454545 | 90                   | 59.8       | PASSED  |
| 57.4           | 83.7545455 | 100                  | 80.4       | PASSED  |
| 66             | 92.5        | 100                  | 46.5       | PASSED  |

2.2.4 **Data Modelling**

This phase consists of five steps: testing, outcome evaluation, training, practice, and knowledge representation. This phase in the prediction of the Student Performance in the CCNA R&S class was analyzed and generated the decision tree, confusion matrix, and receiving characteristic operating curve.
3. Results and Discussion

3.1 The Model

Figure 1 illustrates the graphical presentation of the pruned decision tree on Student Performance in the CISCO Certificate Program. Wherein, Quizzes as the highest instances and become the first split between the (Finals < 79.67) and (Finals > 79.67) in predicting student Performance. Moreover, in Figures 2-3, show the skills decision rule that Quizzes has the highest factor that students will get PASSED, FAILED, CONDITIONAL remarks in the CISCO Certificate Program.

![Figure 1. Decision Tree Model on Student Performance on the CCNA R&S program.](image)

![Figure 2. Student Performance Decision Rule](image)

The Confusion Matrix table below shows the following:

Table 3 displays the decision tree has correctly classified One Hundred Five (105) instances as PASSED and incorrectly Classified Three (3) as FAILED and Nine(9) as CONDITIONAL leading to Misclassification.

- The decision tree correctly classified Twenty-four (24) instances as FAILED, and incorrectly Classified One (1) as PASSED and Four (4) as CONDITIONAL leading to Misclassification.
- The decision tree correctly classified Ten (10) instances as CONDITIONAL, and incorrectly Classified Three (3) as PASSED and Six (6) as FAILED, leading to Misclassification.

![Table 3. Confusion Matrix](image)

| A   | b   | c   | Classified as |
|-----|-----|-----|--------------|
| 105 | 1   | 3   | a = PASSED   |
| 3   | 24  | 6   | b = FAILED   |
| 9   | 4   | 10  | c = CONDITIONAL |

Table 4 presents the cross-validation summary wherein there are 84.24% correctly classified instances and 15.76% incorrectly classified instances. As supported in Table 5, it has been revealed by the algorithm that the Precision Weighted Average in 83.20%. With the classification of the accuracy rate of the following attributes PASSED
(90.5%), FAILED (86.4%), and CONDITIONAL (65.10%). Ultimately, the example has high acceptability and accuracy in predicting student Performance in the CISCO Certificate Program.

Table 4. Cross-Validation Summary

| Correctly Classified Instances | 139   | 84.2424 | %   |
|--------------------------------|-------|---------|-----|
| Incorrectly Classified Instances| 26    | 15.7576 | %   |
| Kappa statistic                | 0.672 |         |     |
| Mean absolute error            | 0.1202|         |     |
| Root mean squared error        | 0.3137|         |     |
| Relative absolute error        | 35.5184|       | %   |
| Root relative squared error    | 76.4811|       | %   |
| Total Number of Instances      | 165   |         |     |

3.2 Student Academic Performance

The anticipation of student academic performance plays a vital function in any student's life. The grade acquired by the student in the class determines his future. Thus, it is essential for every teacher to announce whether the student will reach or run out in the exams. If the prediction suggests that a student will most likely fall short in the test, a prior remedy or intervention program should start. Any extra efforts taken by the teacher to improve the learner's performance and help him to pass the examination is worth commending.

The cognition and skill in CCNA routing and switching (R&S) provided by Cisco presents an excellent platform for networking professionals and aspiring students who can apply for jobs like Network administration or technical support after they graduate. Enterprises only hire qualified network professionals. Hence, scholars should be well versed with the accomplishments and knowledge of Computer Networking because completing the CCNA R&S program is an indication of expertise, experience, and achievement in handling network issues promptly and efficiently. The CCNA R&S course adheres to the design philosophy of E-doing. It is a principle that believes that people learn best by doing. The curriculum includes embedded, highly interactive e-doing activities to help stimulate learning, increase knowledge retention, and make the whole learning experience much more vibrant, and that makes understanding the content much more straightforward.

As shown in the decision tree model, the J48 algorithm was able to predict 90.5% accuracy for PASSED, 86.4% accuracy for FAILED, and 65.1% accuracy for CONDITIONAL. As revealed by the result, Quizzes proved to be the highest indicator that can affect the certification performance of a student in the CCNA R&S class. Quizzes are lessons that are both theoretical and applied. It refers to the summary of the Chapters which are composed of the following topics: Network Operating systems, Protocols and Communications, OSI layers, IP Addressing, and IP Subnetting. The first three chapters require students to learn the fundamental process, tools, and equipment for communications. However, what make this lesson more challenging are the configurations that are needed to operate computers. Furthermore, the remaining chapter IP addresses and Subnetting focuses on number systems and calculations. Hence, a lot of discussion and lab exercises using Packet Tracer, a network simulation tool are required of the students to help them have a better comprehension and retention of the topics.

For a student to pass the course, the model predicts the student’s Quizzes grade should be higher than 79.67, and their score for their Final exams should be higher than 52.52. The student’s Pre-Test rating should be higher 48.8, and their Hands-on should be less than or equal to 95. In helping the students in achieving their target marks, Teacher intervention is essential. Evaluation of retention

Table 5. Class Accuracy Rate

| TP Rate | FP Rate | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | Class     |
|---------|---------|-----------|--------|-----------|-----|----------|----------|-----------|
| 0.963   | 0.214   | 0.897     | 0.963  | 0.929     | 0.781| 0.905    | 0.916    | PASSED    |
| 0.727   | 0.038   | 0.828     | 0.727  | 0.774     | 0.725| 0.864    | 0.682    | FAILED    |
| 0.435   | 0.063   | 0.526     | 0.435  | 0.476     | 0.403| 0.651    | 0.381    | CONDITIONAL |
| Weighted Avg. | 0.842 | 0.158 | 0.832 | 0.842 | 0.835 | 0.717 | 0.861 | 0.795 |
skills through discussion and hands-on demonstration can be applied to students to promote in them a more robust understanding of the topic at hand. Team collaborations among students through repeated Team Skills challenges can also help slower underachieving students and motivate them to strive harder in mastering the knowledge of the lessons. Ultimately, the CCNA R&S Certification course is a four-level program, and then it is not an easy chore for the students to attain. Merely with the continuous guidance and dedication of the instructor. Applying data mining techniques and prediction models will significantly ease the obstacles that the students have to look for them to complete the track.

4. Conclusion

Through the years, the certification requirements for all professionals in the IT field are turning and becoming more competitive. Therefore, it is a must that graduates should be capable of displaying the needed competencies, accomplishments, and knowledge relative to their aspired Job description. This study predicted possible outcomes based on the student performance of the IT students in their CCNA R&S Certification Program. The prediction model derived from the J48 algorithm with the application of various classifiers revealed Quizzes, followed by the final exams are the highest attributes that could predict the outcome of the student's grades. A highly acceptable result shown in the Receiving Operating Characteristics Curve (ROC), decision tree, Area under ROC Curve and confusion matrix, proves that the model is ideal and highly beneficial for both students and teachers. Practice tests, good study habit, and skills integration challenges can improve student performance on critical areas of their subject. Lastly, teacher intervention remains an important factor in determining student success or failure in the certification program.

5. Recommendation

The results can guide improvements in the teacher's strategies and techniques in the delivery of both the theoretical and practical lessons in the certification course. The model can guide policy creation in the IT program. It is essential for students to make preparations in their studies like practicing technical skills, gathering proper study materials, taking online practice exams ahead of the scheduled evaluations by their teachers. Moreover, a more advanced study using other reputable data mining techniques and prediction algorithms, or a combination of both to assess and guide the student and teacher performance maybe suitable future studies.

6. References

1. Christen A. Transforming the Classroom for Collaborative Learning in the 21st Century. Techniques: Connecting Education and Careers (J1). 2009; 84(1):28-31.
2. Randall MH, Zirkle CJ. Information technology student-based certification in formal education settings: Who benefits and what is needed. Journal of Information Technology Education: Research. 2005; 4(1):287-306. https://doi.org/10.28945/278
3. Ogallo GG. IoT-Enhancing Data-driven Decision-making in Higher Education. Case Study of Ohio University (Doctoral dissertation, Ohio University). 2018.
4. Bonde SN, Kirange DK. Survey on Evaluation of Student's Performance in Educational Data Mining. In 2018 Second International Conference on Inventive. 2018. https://doi.org/10.1109/ICICCT.2018.8473228 PMid:29708025
5. Minaei-Bidgoli B, Kashy DA, Kortemeyer G, Punch WF. Predicting student performance: an application of data mining methods with an educational web-based system. In 33rd Annual Frontiers in Education. 2003; 1:T2A-13.
6. Kotsiantis SB, Pintelas PE. Predicting students marks in the Hellenic Open University. Fifth IEEE International Conference on Advanced Learning Technologies. 2005; p. 664-8. https://doi.org/10.1109/ICALT.2005.223
7. Superby JF, Vandamme JP, Meskens N. Determination of factors influencing the achievement of first-year university students using data mining methods. Workshop on Educational Data Mining. 2006; 32:234.
8. Al-Radaideh QA, Al-Shawafka EM, Al-Najjar MI. Mining student data using decision trees. International Arab Conference on Information Technology (ACIT’2006), Yarmouk University, Jordan. 2006; p. 1-5.
9. Osmanbegovic E, Suljic M. Data mining approach for predicting student performance. Economic Review: Journal of Economics and Business. 2012; 10(1):3-12.
10. Patil TR, Sherekar SS. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. International Journal of Computer Science and Applications. 2013; 6(2): 256-61.
11. Yadav SK, Pal S. Data mining: A prediction for performance improvement of engineering students using classification. Journal of Information Technology. 2012; 2(2):51-6.
12. Al-Barrak MA, Al-Razgan M. Predicting final student GPA using decision trees: a case study. International Journal of Information and Education Technology. 2016; 6(7):528. https://doi.org/10.7763/IJIET.2016.V6.745

13. Kumar SA. The efficiency of decision trees in predicting student's academic performance. D.C. Wyld, et al. (Eds): CCSEA. 2011; p. 1-9.

14. Airi P, Anderson PK. Cisco Packet Tracer as a teaching and learning tool for computer networks in DWU. Contemporary PNG Studies. 2017; 26:88.