Development of Random Tree Based Student Competency Model in Java Programming

Devine Grace D. Funcion

Leyte Normal University, Tacloban City, 6500 Leyte, Philippines; dgd_35@lnu.edu.ph

Abstract

Objective: Java programming is perceived to be a difficult subject. The educated programmer is in demand in the local and international market in creating computer application. However, no model that would assist the teacher in identifying the needed competency in Java programming that will make it less stressful for the student to understand the topic.

Methods: data mining technique was explicitly utilized the random tree to predict the academic performance of the students. The purposeof gathering the data was through survey questionnaire for those fifty (50) are enrolled IT students in the IT_103 Java Programming subject answered and poll. Java Programming class schedules are: 10.30-12.00 MTh (Monday and Thursday), and 4.00-5.30 TFri (Tuesday and Friday). The researcher observed the schedule in conducting the survey.

Findings: The algorithm generated by the decision tree model shows that conditional statement, operators, arrays and loops are important Java programming competency. It means that the student should learn mastery in the topics for them to become expert in java programming.

Application/Improvements: In the formulation of the course syllabus for the Java Programming more contact hours should be given to the following topics conditional statement, Operators Arrays, and Loops for the student to understand more the problem. For more improvement of the subject area, the use of another type of that the data mining technique is recommended and with additional parameters to test the accuracy in predicting Java competency.

Keywords: Array, Data Mining, Information Technology, Java Programming, Loops, Random Tree Algorithm

1. Introduction

In recent years, the demand for computer programmers has conveyed so much attention as it is one of the in-demand jobs in the local and international market. The Bureau of Labor and Statistics identified the fastest growing occupation projected until 2026 among the top is a software developer with one hundred eighty-six (186.6) thousand employment needed. Information Technology (IT) companies are hiring graduates who have excellent programming skills. Becoming a computer programmer requires great analytical thinking skills to design code and debug a program.

Moreover, analytical skills are necessary for problem-solving according to which involves identifying relevant information, assessing needs, determining available input, specifying desired objectives and results in the organization. In programming, there are two components (data, and instruction). To work with data, students need to understand variables and types; to work with instructions the student has to understand control structures and subroutines. Programming languages always have commands for getting data into and out of variables and for doing computations with data. In the declaration, the object has a class which is composed of three members the field, method, and classes. Moreover, skills such as computational thinking (problem-solving, complex systems design, evaluation, and understanding human behavior) are foundations of an IT professional.

Java programming is an Object-Oriented Programming (OOP) tools to solve problems using class and object. Object-oriented programming as defined in webopedia a type of computer programming in which programmers applies not only the data structure but also the types of operations in data structure. Also, OOP is a method of programming that involves the creation of an intellectual model that data is associated. Java programming is the most popular programming language.
used in the world, and Java works in different platforms like in Windows, Mac, and Linux. Furthermore, Java programming is perceived to be a difficult subject of some IT student. The study of pointed out some reasons why Java programming is difficult for a student to learn. He identified three areas of programming difficulties; first multiple skills (syntax, structure, semantics, and style) other educational novelty (problem-solving and precision), and third various processes (translation, algorithm, and coding).

However, with the high demand in computer programmer, there is scarcity in this field of profession. A report issued by the Commission on Higher Education (CHED) stated that in School Year 2016-2017 there are about 398,765 thousand students enrolled in the IT profession all over the Philippines. However, there are only 73,646 thousand total students who graduated in the S.Y 2016-2017 with the degree in Bachelor of Science in Information Technology. Several kinds of literature show the reasons for the high turnout of IT dropouts. In the University of Tartu, Estonia in the school year 2014/2015, 25% of the students had dropped out in IT. Some of the reason identified was students indicated a math-related subject called discrete elements of math and programming to be too difficult.

Additionally, when coding specific programming language, it entails allotting of rules and processes to follow which makes programming challenging to learn. Another factor that influences the high dropout rate is a conflict of work while studying simultaneously.

Additionally, reasons for the low turnout of IT graduates are student retention. According to, the study of low retention has been prominent in programming subjects in tertiary level. Also, despite the popularity of high demand for computer programmer according to reason for low retention rate, student has a problem with the subject: learning how to program.

Another situation that makes programming challenging to learn are those students enrolled in none technological courses. In the K-12 program implemented by the Department of Education in the Philippines senior high school students can choose from the tracks such as Accounting, Business Management (ABM) track, Science, Technology, Engineering and Mathematics (STEM) track, Humanities and Social Science (HUMSS) track, and General Academic Strand (GAS) track. Moreover, those enrolled students in non-STEM track find it challenging to comprehend the topics in Java programming since computer programming is not part of the offered subject in the non-STEM course. The CHED issued a CHED Memorandum Order (CMO 10) series of 2017 policy on student affected by the implementation of the K to 12 programs. The Higher Education Institution (HEIs) has the option to implement a bridging program to address the problem of some non-STEM courses who wish to take Information Technology course in the tertiary level to assist the student to understand the basic concepts of computer programming.

Several studies have been conducted to identify the competency of students in programming. In the study of he mentioned that student should develop more on the cognitive and metacognitive process, knowledge structures and, content representation. Another study of the Evaluation of Programming Competency using Student Error Patterns he identified five (5) category of competency (Basic Programming, Sequencing, Tracing Non-iterative, Tracing Iterative, writing programs and Program Exception Handling). Students show that some problem and difficulty in error handling. It implies that competency in programming is essential so that IT students can create and solve real-world problems.

Hence, the ability to predict competency in Java programming is essential to lessen the difficulty in programming. Moreover, data mining as a technique is employed in this study to discover new knowledge from a large amount of data to identify relationships such as patterns, the association among variable in databases. Discovery of knowledge can be examined using machine learning technique. The study of student dropout rate in online education was conducted by and utilized data mining to predict reasons for dropout. A total of 189 students registered to the online Information Technologies Certificate Program in 2007-2009. The result of the research shows that online technologies self-efficacy, online learning readiness, and previous online experience as the most critical factors in predicting dropouts. The researcher deploys the used of random tree algorithm because of robustness in handling small datasets and numbers of fields. Also, the strength of the random tree is on classification and regression tree method. The random tree used two way of randomization first bagging, where it creates cloned training sets through sampling with replacement from the original datasets. Second, each split of the tree, is considered in a sampling of the input field for the impurity measure.

Furthermore, data mining was conducted by employing ability of IT students in their on-job training as the result of the study Business Operation is the most critical attri-
bute in the On-the-Job Training. Wherein, students are trained to be productive, responsible and cooperative and have the initiative to act. Since computer programming is a difficult subject as stated in the literature. There are no studies conducted that will predict student competency in Java programming that will lessen the difficulty of programming among IT student. However, findings are limited to theoretical approaches. Furthermore, for the programming teachers to identify what specific lessons in Java programming that need more time in the discussion to help the student understand and become an expert in programming. Further, the research study is used to develop a model for predicting competency in Java programming.

2. Methodology

The study utilized Knowledge Discovery in Database (KDD) process which refers to the overall operation of discovering useful knowledge from data. Data mining is a particular step in this process—application of specific algorithms for extracting patterns (models) from data. The objective is to unravel the hidden data behind the difficulty in learning computer programming and unleash the important competency necessary to improve the Java programming of the IT students. There is no other technique used to uncover hidden data except the KDD process.

2.2 Data Mining Process

2.1.1 Data Collection

The method of gathering the data was through a survey questionnaire. A total of fifty (50) out of two hundred eight (208) first year IT students were identified to answer the survey questionnaire. The researcher observed the following class schedule in conducting the survey 10.30-12.00 MTh (Monday and Thursday) and 4.00-5.30 TFri (Tuesday and Friday). Table 1&2 show the rating and limit scale used to identify the level of competency in Java programming.

2.1.2 Data Sampling

The researcher is handling two (2) Java Programming (IT_103) subject. In each subject a total of thirty-five (35) students. However, some students dropout and stop attending the classes during the conducting classes. Some students were absent during the conduct of the survey questionnaire. However according to the study of having a relatively small microarray dataset (n= 53 to 280) the predicted and actual classification error is between 1%-7%. Additionally, in small sample size, re-sampling strategies like the bootstrap or repeated/iterated k-fold cross validation are a most appropriate tool to sample test.

2.1.3 Software Used

The Weka software was utilized to analyze the data. Weka is a machine learning algorithm and data processing. It provides extensive support for the whole process of experimental data mining, including preparing the input data, evaluating learning schemes statistically, and visualizing both the input data and the result of learning.

2.1.4 Data Preparation and Processing

During this phase, a pre-processing of collected and prepared the data for the mining techniques. At first, we eliminated some irrelevant attributes, e.g., student name, Student Track, teacher’s name, and schedule. Then, each survey report will have the following characteristics as shown in Table 3.
Table 3. Sample datasets

<table>
<thead>
<tr>
<th>Basic in Java Programming</th>
<th>Operators</th>
<th>Conditional Statement</th>
<th>Loops</th>
<th>Array</th>
<th>Mean</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2.0</td>
<td>Newbie</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2.6</td>
<td>Newbie</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2.6</td>
<td>Newbie</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.8</td>
<td>Intermediate</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.8</td>
<td>Intermediate</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2.8</td>
<td>Intermediate</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3.0</td>
<td>Intermediate</td>
</tr>
</tbody>
</table>

Second step data preparation, the result of the survey was in MS Excel and later converted to Microsoft Excel Comma Separated Values File (.csv). The .csv file was then loaded to Notepad++. In this stage, data cleaning is done to eliminate unwanted symbols (i.e., comma, colon, spaces).

Additionally, in the notepad application declaration of syntaxes like @Relation, @Attribute, and @Data is included as a requirement in the Weka application. The converted text file to Attribute-Relation File Format (ARFF) in notepad describes the list of instances sharing a set of attributes and the accepted file format for WEKA application. Next, information is uploaded to the Weka Application and conducted the pre-processing of raw data to a more understandable file format. The third step, data modeling, in WEKA it utilized the Random Tree algorithm to predict competency in Java programming.

2.1.5 Data Visualization

After loading the data to WEKA, we set out some primary useful knowledge about the attributes before applying any data mining method by using the visualizing technique in the software. For example, we found that in the Remarks attribute there are eight (8) students show basic programming competency in java programming, eleven (11) students under operator’s competency, and twenty-five (25) students under conditional statement competency, four (4) students under Loops competency, one (1) students under Array competency. The problem is how what specific competency in Java programming does the teacher will give more emphasis to improve the level of competency among BSIT First year students.

![Figure 1. Primary visualization of the different attributes.](image-url)
2.1.6 Random Tree Algorithm

A random tree is a rooted tree following the Markov process. The nodes in a random tree with the arrays of the n-dimensional vectors are associated. The identified random tree can be a collection of vector arrays.

\[
\{S_0, S_1[b], ..., S_N[b] \ldots [b] \}
\]

Costs of n properties at time \(k\Delta t\) are enclosed inside n-dimensional vector \([S_k[i_1]...[i_k]]\). All nodes have the same number of branches, \(b\). All branches have direction, showing time transition. The node at the branch head from the tail node is originating. Initial vector \(S_0\), connected to the root, has an initial cost of \(n\) properties.

For given \(k\)th vector, \(S[i]...[i] = (S[i]...[i]) \ldots (1, ..., n)\). \(k\) is the index of the vector. \(b\) is created.

\[
R(t + \Delta t) = \exp \left[ \frac{\Delta t}{\sqrt{\Delta t}} \left( \sum_{j=1}^{k} R(t_{j}) \right) \right]
\]

Condition is \(S(k\Delta t) = S_k[i_1]...[i_k]\) at time \(k\Delta t\).

\[
S_{k+1}[i_1]...[i_k][j] = S_k[i_1]...[i_k][j] \exp \left\{ r - \delta_j - \frac{\sigma_j^2}{2} \Delta t + \sigma_j \sqrt{\Delta t} W_j(t) \right\}
\]

Random Decision Trees algorithm was shown to have nearly comparable accuracy to Random Forest but without using bootstrap samples based on the training set. The critical advantage of completely random trees for our experiment is that growing the trees does not depend on the training data, which means that a single forest can be generated once and used for case bases with different sets of case instances.

3. Result and Discussion

Figure 2 illustrates the graphical presentation of the pruned decision tree of Student Competency in Java Programming. Hence, the major topic in Java programming covers (Basic Programming, Operators, Conditional Statement, Loops, and Arrays). Wherein, Conditional Statement as the highest instances and become the first split between the (Operators <=2.5) and (Array >2.5 60) in predicting student competency in java programming.

Table 4 shows the correctly classified instances wherein there are 77.55% precisely classified instances and 22.44% incorrectly individual instances this is supported by Table 5 that shows the complete accuracy by class wherein the Precision Weight Average of the competency level of student in Java programming is 78.10%.

It implies that the conditional statement having a weighted average of 78.10% is an important attribute or topic in Java programming that the teacher needs to focus more attention during class discussion. Giving more computer programming problems and activities should be implemented in the use of conditional statements,
operator and arrays. The more activities are provided to the students to analyze the computer programming problem the higher the probability the student will be able to understand and find the subject not challenging to understand. In the study, of ten quick tips for teaching programming, he mentioned the best way to guide the student in programming use worked examples: a step-by-step guide in solving an existing programming problem. The instructor provides many similar programming examples for learning purposes. Worked samples should be broken down into sub-goals and labels. Predicting student competency in Java programming is essential to reduce the difficulty in understanding Java programming. Hence, to enhance the learning in a conditional statement, arrays, operators’ teachers may conduct scaffolding, coaching, and modeling to help students master intellectually challenging operations. The model should address ‘orientation’ by incorporating a right introductory section on programming, and its different types.

Table 4. Correctly classified instances

<table>
<thead>
<tr>
<th>Correctly Classified Instances</th>
<th>38</th>
<th>77.551%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incorrectly Classified Instances</td>
<td>11</td>
<td>22.449%</td>
</tr>
<tr>
<td>Kappa statistic</td>
<td>0.6482</td>
<td></td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>0.1061</td>
<td></td>
</tr>
<tr>
<td>Root mean squared error</td>
<td>0.3079</td>
<td></td>
</tr>
<tr>
<td>Relative absolute error</td>
<td>39.3793%</td>
<td></td>
</tr>
<tr>
<td>Root relative squared error</td>
<td>84.4445%</td>
<td></td>
</tr>
<tr>
<td>Total Number of Instances</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Detailed accuracy by class

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>MCC</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>NO Knowledge</td>
</tr>
<tr>
<td>0.636</td>
<td>0.026</td>
<td>0.875</td>
<td>0.636</td>
<td>0.689</td>
<td>Newbie</td>
</tr>
<tr>
<td>0.88</td>
<td>0.25</td>
<td>0.786</td>
<td>0.88</td>
<td>0.636</td>
<td>Intermediate</td>
</tr>
<tr>
<td>0.25</td>
<td>0.067</td>
<td>0.25</td>
<td>0.25</td>
<td>0.183</td>
<td>Advance</td>
</tr>
<tr>
<td>0</td>
<td>0.021</td>
<td>0</td>
<td>0</td>
<td>-0.021</td>
<td>Expert</td>
</tr>
<tr>
<td>Weighted Avg.</td>
<td>0.776</td>
<td>0.139</td>
<td>0.781</td>
<td>0.776</td>
<td>0.657</td>
</tr>
</tbody>
</table>

A Conditional Statement is statements that execute specific conditions whether the statement is ‘True’ or ‘False’. In Java, there are two forms of conditional statements: the if-else statement, can choose between two alternatives; the switch statement, to choose between multiple options. The basic structure of an If-Else Statement follows this structure;

```java
if (expression) {
    // codes
}
else (expression) {
    // codes
}
```

While operators are a symbol used to perform operations of addition, subtraction, division multiplications and even comparing values using the greater-than-equal symbol, less-than-equal symbol.

```java
if (age <= 18) {
    System.out.println("You are underage");
} else {
    System.out.println("You are legal age");
}
```

The above Java code will determine the age of the person. If the statement is true, then person age is Under Age, hence if the statement is false, then the personage is his at legal age. Observe in the above code conditional statement and operators are connected or intertwined if the student does not know how to construct correct conditional statement and use correct operations then the struggle is present. According to the study student ability to provide semantically correct code hinders their ability to recall and apply correct syntax.

4. Conclusion

Predicting student competency in java programming using random tree algorithm is helpful for the student who is facing difficulty in Java programming. Hence, for the student to become an expert in programming topics like a conditional statement operators, arrays, and loops are programming skills or competencies that need to have mastery from the student. Also, for the teachers handling programming subject, these will serve as guide and awareness of the topics in Java programming that needs more time and attention more complete activities and exercises that will enhance more there cognitive thinking and rea-
soning during class discussion to give more emphasis on the importance Java programming.

5. Recommendation

The following are some recommendations:

1. In the formulation of the course syllabus for the Java Programming, more contact hours should be given to the following topics conditional statement, operators arrays, and loops for the student to understand more the problem and become an expert in Java programming.
2. Other State, Colleges, and Universities might consider the topics in a conditional statement, operators loop and array as essential skills in Java programming that the IT student must understand,
3. Bridging program be conducted to equipped students with basic knowledge in C programming as low-level and structured programming to the non-STEM course,
4. During class discussion, teachers should not focus more on the theoretical components of the subject but more on the practical application of the topics in java programming. It is recommended to give more programming activities to the student's to help the student recall the syntax is used in Java programming, and
5. More comprehensiveconduct of research study in the use of Random Tree in predicting student competency in Java programming to have a more accurate result.

6. Reference


