Predictive Analytics Assessment Tool For Information Communication Technology (ICT) Courses Of Technical And Vocational Education And Training (TVET) School In The Philippines

Joseph G. Gonzales, Ace C. Lagman, Roman M. De Angel

Abstract—The project is aims to develop a predictive analytics tool that will help TVETs Schools in the Philippines identify possible students who would fail in the NC certification so necessary interventions can be provided. The study aims to utilize the power of machine learning algorithms particular decision tree algorithm using J48 method to extract significant rules to embed in the proposed decision support system. This project will also adopt predictive analytics and will create a model to examine the students' performance inside the classroom, and predict students' performance, and help the institute to determine the possibility of students to pass the NC certification up to at least level II. The research methodologies used on this study were quantitative and qualitative methods. Qualitative research provides insights in the problems in developing hypotheses, while quantitative research has been used to measure the difficulty by having statistical data that provides figures. The study is mend to judge able to scrutinize summarize opinions from vast data gathered in the form of questionnaire and surveys.

Index Terms—CHED, Decision Tree, NC Certifications, Predictive Analytics, TESDA, TVET, Research Methods.

1 INTRODUCTION

Predictive analytics is now becoming popular among colleges and university, using advanced analytics to develop creative new techniques and tools to progress students' performance. In the Philippines, colleges and universities are now moving to the use of analytics to monitor the student performance to improve learning and to make the students more competitive. In the Philippines there are two agencies that is task to monitor the students in terms of higher education, it is the Commission on Higher Education (CHED) and for technical education, it is the Technical Education and Skills Development Authority (TESDA). With more than 2,299 higher learning institutions in the Philippines it's the responsibility of the schools to ensure that the graduates they produce will be competitive in the international market. "...All certificates relating to national trade skills testing and certification system shall be issued by the Authority through the TESDA Secretariat." From section 22 of TESDA law. This provides TESDA the Authority to provide the given certification. TESDA release a the official report on Philippine Technical and Vocational Education and Training (TVET) statistics (1995-2013) the report includes chronological data on TVET outputs such as past enrollment, past graduates, past assessment and past certifications. It also contains education and service-related figures which has been documented and analyzed to provide as inputs to policy making and program development. The report can be used as reference material for policy-makers, administrators and planning committee officers of TVET institutions, and other partner organization, trainings, researchers and students have significant in educational fields.

Table 1 shows that there are a total of 4,669,231 students were not able to finish the program due to several reasons.
Table 2 – Distribution of Certified Students by Year: 1995-2013

<table>
<thead>
<tr>
<th>Year</th>
<th>Assessed</th>
<th>Certified</th>
<th>Certification Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>49,686</td>
<td>24,414</td>
<td>49.1</td>
</tr>
<tr>
<td>1996</td>
<td>97,405</td>
<td>49,744</td>
<td>51.1</td>
</tr>
<tr>
<td>1997</td>
<td>68,797</td>
<td>28,304</td>
<td>41.1</td>
</tr>
<tr>
<td>1998</td>
<td>85,838</td>
<td>23,379</td>
<td>27.2</td>
</tr>
<tr>
<td>1999</td>
<td>72,952</td>
<td>38,822</td>
<td>53.2</td>
</tr>
<tr>
<td>2000</td>
<td>90,487</td>
<td>57,637</td>
<td>63.7</td>
</tr>
<tr>
<td>2001</td>
<td>185,399</td>
<td>144,766</td>
<td>78.1</td>
</tr>
<tr>
<td>2002</td>
<td>185,305</td>
<td>68,529</td>
<td>36.3</td>
</tr>
<tr>
<td>2003</td>
<td>111,251</td>
<td>39,208</td>
<td>35.2</td>
</tr>
<tr>
<td>2004</td>
<td>207,918</td>
<td>122,453</td>
<td>59.2</td>
</tr>
<tr>
<td>2005</td>
<td>223,984</td>
<td>108,361</td>
<td>48.4</td>
</tr>
<tr>
<td>2006</td>
<td>257,796</td>
<td>160,941</td>
<td>62.4</td>
</tr>
<tr>
<td>2007</td>
<td>398,711</td>
<td>293,219</td>
<td>73.5</td>
</tr>
<tr>
<td>2008</td>
<td>549,121</td>
<td>431,071</td>
<td>78.5</td>
</tr>
<tr>
<td>2009</td>
<td>836,131</td>
<td>690,836</td>
<td>82.6</td>
</tr>
<tr>
<td>2010</td>
<td>716,220</td>
<td>594,323</td>
<td>83.0</td>
</tr>
<tr>
<td>2011</td>
<td>835,572</td>
<td>703,360</td>
<td>84.2</td>
</tr>
<tr>
<td>2012</td>
<td>968,535</td>
<td>830,458</td>
<td>85.74</td>
</tr>
<tr>
<td>2013</td>
<td>534,613</td>
<td>636,284</td>
<td>118.27</td>
</tr>
<tr>
<td>Total</td>
<td>6,448,721</td>
<td>5,042,089</td>
<td>78.19</td>
</tr>
</tbody>
</table>

Table 2 revels that based on the data, out of more than 6 million students only 5,042,089 million were certified. Since TESDA aim to have quality workforce they should aim to have 100% passing rate of certified students. This certification gears toward graduates can perform expected task in the workplace. Students achieving these certifications are ensured that they pose quality to be globally competitive middle level workers.

With the resources that TVETs gives to its accredited schools, sometimes students does not attain the said certification, they failed the assessment exam, if they failed students will not be issued NC II certificates, in terms of looking a job, they will had less opportunity of landing a job than those who had passed the NC II certifications. By adopting predictive analytics and creating a model this measures will examine the students’ performance inside the classroom, and predict students’ performance, helping the institute to determine the possible of students passing the NC II certification.

2. Background of the Study

The lead agency of the government in-charge with National Competency for technical education courses is the Technical Education and Skills Development Authority (TESDA). Its vision is to produce highly competitive workforce with possible work attitude, to ensure each student acquires the needed competency TESDA strictly follow is Quality Policy through SERVICE.Strategic Decisions, Effectiveness, Responsiveness, Value Adding, Integrity, Citizen Focus, Efficiency.

TESDA Technology Institutions (TTIs) 2017 Annual Accomplishment Reports

TESDA Technology Institutions (TTIs) are the forefront in the conduct and delivery of TESDAs training programs and related services. There are 122 TTIs all over the country. They are composed of 56 TESDA Administration Schools (TAS), 18 Regional Training Centers (RTCs), 45 Provincial Training Center (PRTs), and 3 Specialized Training Centers (STCs).

In 2017, the TTIs have a combined enrollees and 319,580 graduates. These outputs were about 140 % above the 257,621 enrollees for the year. Likewise, a weeklong TTI Administrators’ Conference was held on May 2017 in Davao City. The conference objectives served as forum to discuss and clarify TTIs operational guidelines, the TTIs Institutional Development Plans (IDP) 2017-2022, Facilities and Equipment Maintenance systems (FEMS) and quality journey.

The conference provided an opportunity for the 112 participating Administrators to learn and benchmark the operations of Lupon School of Fisheries (LSF), the country’s first recipient of the Gold Level Award from Asia Pacific Accreditation and Certification Commission (APACC). On the same occasion, it was announced that Puerto Princesa School of Arts and Trades (PPSAT) became second Gold Level awardee among TTIs.

Certificates of Recognition were also issued to TTIs who have shown exemplary performance during the conference.

Theoretical & Conceptual Framework of the Study

Knowledge Discovery in Databases is the significant method of determining valid, novel, possibility useful, and eventually understandable design in data (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

Figure 1 shows the composition steps of KDD process. Any information given will go through the given process of KDD which will derive the knowledge as to help assessing the records of students who are taking ICT courses.
The model that will be created, and using some Data Mining paradigm which will have some iterative process for higher percentage of accuracy resulting to the student passing the assessment exam

The datasets consist of the following:
- Demographic Profile of the student – This includes the primary information of the student also additional information of educational background will be included.
- YP4SC (Youth Profiling for Starring Careers) – To guide the students to create a sound decision for their desire career, one organization was conceptualized which is YP4SC, to guide the students to select a right career that will matched on their respective degree by giving them more information on what is the current jobs is on trending at the industry and make some recommendations on what training school they can enroll for the preparations on their chosen career.
- Student Grades – Student grades are essential in prediction, this determines the interest of the students on specific courses.
- Graduate Status – Added as one of the predictors to determine the student availability to take the assessment. To create the model Data Mining paradigm will be used in predicting students who will pass the assessment exam.

The general objective of the study is to develop a tool using a derived model to help TVET schools to predict students who will pass the NC Certification.

Specific Objectives
1. Determine the academic profile of graduates.
2. Determine the problems encountered by students in taking the course.
3. Develop a predictive analytics tool to list the possible students who will pass NC certification.

Scope of the Study
The scope of the study is focused to NC ICT assessment of TVETs School in the Philippines. This includes all who offer NC ICT courses. The datasets that will be used are based on the gathered information of the students from TESDA Annual Reports such as demographic profile of enrolled and graduating students in their respective ICT courses, and
students who passed the assessment exam/NC certifications. The model creation will be based on classification decision tree and clustering k-means. Java Technology will be used to develop the application providing simple user interface. The application can be installed to any computer platform with JDK installed. Data that will be gathered is from year 2010 to latest result of year 2015. Only records that are listed on the master list will be included as part of the data sets information.

**Limitation of the study**
Input data will be limited to the datasets that were mentioned in this paper. This research focuses on the TESDA NC ICT assessment only for TVETs Schools in the Philippines were each branch are located on different parts of the Philippines. The software program that will be developed has one user only with no login credential. This will provide a forecast of students who will pass the assessment exam as well those students who will fail. Model will be created based on the data mining paradigm mentioned in this paper.

### 4. Methodology

#### A. Research Design

According to Gray definition of research design, researchers usually adopt a numerous methods of research techniques in order to gathered pertinent information that will support the study, such as series of interviews, participative or non-participative observations, case studies, and surveys (Gray, 2004) Qualitative Research provides understanding on the underlying reasons, provides motivations and opinions of different stakeholders in academe committee. Qualitative research provides insights in the problems in developing hypotheses. The common methods of Qualitative research are group discussion, gain opinions in the group discussion, set individual interviews, do participate and observe. Quantitative Research is gear to measure the difficulty by having statistical data that provides figures. The study is mend to judge able to scrutinize summarize opinions from vast data gathered in the form of questionnaire and surveys. The method includes paper, online survey, and interviews. In this study the researcher uses structured data in numerical forms or data that can be converted easily in to numerical values.

**Qualitative Research Method**

The qualitative research method is the process of real involvement in the study in order to prove the why, how, what, where, when or ‘who’ of decision making in the field of the study that the researcher wanted to prove. The three most common methods are the researcher observation, in-depth interviews and mostly done in small scale of groups of people in obtaining specific data.

**Quantitative Research Method**

Quantitative research method were embarked to learn normal practices by adapting many methods such as surveys, face-to-face interview, observations and techniques that uses numbers as data collection from the viewpoint of the objectives and the basis theory of the research. The main purpose is to determine the data and distinctiveness about what is being deliberate in terms of frequencies, averages, and other numerical computations.

#### B. System Methodology

According to Ghini, the ISDM (Information Systems Development Methodology) is a rapid growth modeling to the changing industry suitable for website development techniques. It is designed to provide the repeatable process of systems with the standard frameworks necessary to efficiently and effectively implement the system which includes: scope a project, conduct analysis, define and design the solution, create the system modules and evaluate the system after its implementation (Ghini, 2012).

**Figure 5. Information Systems Development Methodology (ISDM)**

Figure 5 shows the 7 phase of ISDM, and these are as follows:

1. Business Needs analysis
2. Scope Definition
3. Requirement Analysis
4. Design
5. Development
6. Integration, Test, and Acceptance
7. Implementation – Deployment (Ghini, 2012)

**Algorithm**

Decision Tree algorithm serves as one of the most popular data mining techniques because of its simplicity. The decision tree algorithm uses the entropy and information gain as its core foundations. Entropy refers to the uncertainty of data which present the negative and positive examples in the dataset. Information gain refers to the observations/signal of a variable combined to another variable. The formula for entropy is as follows:

**Data Model Extraction**

This section presents the model extracted using J48 algorithm. The J48 algorithm is using the entropy and information gain as key formula to extract and split trees. Entropy refers to the uncertainty of data which present the negative and positive examples in the dataset. Information gain refers to the observations/signal of a variable combined to another variable. The formula for entropy is as follows:

**Data Model Extraction**

This section presents the model extracted using J48 algorithm. The J48 algorithm is using the entropy and information gain as key formula to extract and split trees. Entropy refers to the...
uncertainty of data which present the negative and positive examples in the dataset. Information gain refers to the observations/signal of a variable combined to another variable.

The formula for entropy is as follows:

\[
\text{Entropy } H(S) = \sum_{x \in X} p(x) \log \frac{1}{p(x)}
\]

Where,
- \( S \) – The current (data) set for which entropy is being calculated (changes every iteration of the ID3 algorithm)
- \( X \) – Set of classes in \( S \)
- \( p(x) \) – The proportion of the number of elements in class \( x \) to the number of elements in set \( S \).

When \( H(S) = 0 \), the set \( S \) is perfectly classified (i.e. all elements in \( S \) are of the same class).

The algorithm can be represented in the tree structure below.

![Tree Structure](image)

Figure 6. Tree Structure

Where 2 presents as root node
- The Root node presents the highest calculated information gain.
Where 7,5,6,9 serve as decision nodes
- Where 2, 10, 5, 11, 4 serves as leaf nodes.
- Where rules can be formed by looking at the structure from the tree from top to bottom

If 2 and 7 and 6 then 11 where 11 refers to as outcome of the rules.

D. Data Gathering Instruments

The researchers will employ the following procedures in collecting data:

Interview.
The researchers will be using structured interviews or face to face interviews to gather more pertinent data and getting the camaraderie with the resource peoples. Fresh-hand information and sincerity of the people who will involve in the interview is an added accomplishment on the part of the researchers. Verifications of unclear statements can immediately clarify and understand.

Observation.
For this method, the researchers will employ non partippative observations so that people involves will not be intimidated that they are being watched and recording behaviors within a clearly defined area will not be tamished.

E. Statistical Treatment of Data

The following statistical treatment of data will be use to compute, analyzed, and interpret data and its results.

Percentage

Percentage used as statistical techniques in frequency distribution for the respondents of the study. The formula is given as:

\[
P = \frac{f}{N} \times 100
\]

Where:
- \( P \) = Percentage
- \( f \) = Frequency
- \( N \) = Total Population

Weighted Mean

This statistical tool will be used to compute and analyzed data from the tabulations of respondents who participated on the surveys. It will also use the likert scale table for the interpretations of the results

Standard Deviation

Standard deviation used as to measure the dispersion of the data. The formula is given as:

\[
SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}
\]

Where:
- \( SD \) = standard deviation
- \( x \) = scores
- \( \bar{x} \) = mean
- \( n \) = the number of respondents

<table>
<thead>
<tr>
<th>Scale</th>
<th>Range</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4.51 – 5.00</td>
<td>Strongly agree</td>
</tr>
<tr>
<td>4</td>
<td>3.51 – 4.50</td>
<td>Agree</td>
</tr>
<tr>
<td>3</td>
<td>2.51 – 3.50</td>
<td>Neutral</td>
</tr>
<tr>
<td>2</td>
<td>1.51 – 2.50</td>
<td>Disagree</td>
</tr>
<tr>
<td>1</td>
<td>1.00 – 1.50</td>
<td>Strongly Disagree</td>
</tr>
</tbody>
</table>
Table 3 will be used to interpret the results of survey using weighted means formula.

5. Results and Discussion
Closed monitoring on students performances and application of predictive analytics on the students who will take the National Certifications would provide better outcome as par as the National Certification rates on Technical and Vocational Education and Training (TVET) School in the Philippines were concerned.

Table 4 - Demographic Profile of the Students

<table>
<thead>
<tr>
<th>Group</th>
<th>f</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled</td>
<td>164,146</td>
<td>53.72</td>
</tr>
<tr>
<td>Graduates</td>
<td>141,389</td>
<td>46.28</td>
</tr>
<tr>
<td>Total</td>
<td>305,535</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 4 shows that there were a total of 305,535 students who participated on this study. There are 164,146 enrolled students who represent 53.72 % of the total populations, while 141,389 graduating students which represent 46 % of the total populations.

Table 5 – Enrolled & Graduate by NC Qualifications / Certifications in 2013

<table>
<thead>
<tr>
<th>NC CERTIFICATION</th>
<th>ENROLLED</th>
<th>GRADUATE</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D Animation NC II</td>
<td>2297</td>
<td>2466</td>
<td>109.44</td>
</tr>
<tr>
<td>Career Entry Course for Software Developers Using Oracle NC IV</td>
<td>3303</td>
<td>1890</td>
<td>51.15</td>
</tr>
<tr>
<td>Career Entry Course for Software Developers NC IV Using Java</td>
<td>1065</td>
<td>853</td>
<td>45.61</td>
</tr>
<tr>
<td>Computer Hardware Servicing NC II</td>
<td>1736</td>
<td>1159</td>
<td>66.95</td>
</tr>
<tr>
<td>Computer Hardware Servicing NC III</td>
<td>7063</td>
<td>6377</td>
<td>84.4</td>
</tr>
<tr>
<td>Computer Programming NC IV</td>
<td>56138</td>
<td>48538</td>
<td>86.42</td>
</tr>
<tr>
<td>PC Operation NC II</td>
<td>2097</td>
<td>2781</td>
<td>132.89</td>
</tr>
<tr>
<td>PC Operation NC III</td>
<td>1491</td>
<td>2312</td>
<td>153.06</td>
</tr>
<tr>
<td>Visual Graphics Design NC III</td>
<td>4908</td>
<td>1872</td>
<td>106.10</td>
</tr>
<tr>
<td>TOTAL ENROLLED</td>
<td>164146</td>
<td>141389</td>
<td>86.14</td>
</tr>
</tbody>
</table>

Table 5 shows that among the NC Certifications, Career entry course for Software Developers Using Oracle NC IV and Career entry course for Software Developers Using JAVA NC IV got the low ratings of 45.61% and 51.15% respectively. On the other hand, PC Operations NC II, 2D Animation NC II, and Visual Graphics Design NC III got more than 100% passing rates (155.06, 109.44, and 106.10) respectively, this is due to the other graduating students who takes the NC Certifications but not enrolled due to failure of passing the NC certifications which one of the requirements before the student can graduate. This will be avoided if predictive analytics has been integrated to monitor the student performance so the proper assistance may give to the students who need more attention to pass their NC certification before graduation.

Encountered Problems of the students on taking the course
These are the problems the students encountered on their journey to graduate and be NC certified which has been manifested on the fishbone diagram that has been addressed by the school administrator before sending their graduating students to TESDA for Assessment and NC certification. Some of the problems are as follows: financial problem due to economic reason; instructor credentials which refers to the instructors which are not yet NC certified; learning materials needs updates in order to cope up with new trends in ICT; lack of facility improvements specifically on the computer laboratory; low grades due to instructor credentials and learning materials; and finally, testing exam or assessment exam needs to be updated as well.

Conclusion
Closed monitoring on students performances and application of predictive analytics on the students who will take the National Certifications would provide better outcome as par as the National Certification rates on Technical and Vocational Education and Training (TVET) School in the Philippines were concerned.

Recommendations
Upon completion of datasets the researcher aims to use decision tree algorithm to predict the output of the certification. This is necessary as certification is an evidence of being technically equipped in a computing area. Decision tree rules sets are powerful if then else rules that will provide intelligence in the proposed system.

REFERENCES