Analysis of Student-at-Risk of Dropping out (SARDO) Using Decision Tree: An Intelligent Predictive Model for Reduction

Maricel A. Timbal

Abstract—The advancement of the country is exceptionally reliant on the education of its citizens, which is an incontestable truth of the worldwide perspective. Education assumes a focal part of developing an individual to become a productive citizen and is the most important factor contributing to the progress of the nation. However, children in developing countries have denied their right to education by dropping out for different reasons. Student-At-Risk of Dropping Out commonly known as SARDO is a term coined by the Philippines’ Department of Education, defined as, a student who is likely to become a candidate to drop out. Philippine Statistics Authority had conferred that there is a total of 36,238 people ages from six to twenty-four who are out-of-school. This poses a predicament if this will continue without even attempting to intervene or at least reduce it gradually, to the governing body (DepEd) and ripple its effect to the entire country. In this paper, a classification data mining technique such as decision tree was used to generate a rule-based classifier. Based on results, there were three extracted rules from the decision tree generated using partykit and rpart libraries as a predictive model, a basis for forecasting who among the enrolled students will be a dropout. The first rule is, if a student has an experience being retained from previous grade level/s, he/she is more likely to drop. The second is, if he/she has not experienced retention but his/her number of siblings is greater than six, most probably he/she will drop from school. The last rule is if a student has not experienced retention nor has greater than six siblings, but his/her general average of previous grade level is fairly satisfactory and satisfactory, then he/she will probably stop from schooling. Having this intelligent predictive model, the educational institution with its stakeholders can now have a proactive measure on addressing to reduce the inflation of dropout rate of which is the sole purpose of why this study was conducted.

Index Terms—Dropout, data mining, decision tree, rule-based.

I. INTRODUCTION

The advancement of the country is exceptionally reliant on the education of its citizens, which is an incontestable truth of the worldwide perspective [1]. Education assumes a focal part of developing an individual to become a productive citizen; and is the most important factor contributing to the progress of the nation. However, children in developing countries have denied their right to education through dropping out for different reasons. Student-At-Risk of Dropping Out commonly known as SARDO is a term coined by the Philippines’ Department of Education, defined as, a student who is likely to become a candidate to drop out. National Center for Education Statistics defined dropping out as discontinuing the program and leaving school before graduation without achieving a diploma. Dropping out from school is not just a teacher adviser’s issue because of the reason it will reflect his or her performance of how many students were promoted to the next grade level nor an issue of a dropout student, but it also affects the entirety of the nation. Hoff, et al. [2] supported this argument, stated that this issue impacts more than just the student who makes this decision, but it also affects his or her family, the community, and society as a whole. Campbell [3] studied the socioeconomic consequences of high school dropouts. He evidently displayed in his paper that there is a strong effect of dropout in the increase of socioeconomic disadvantages.

Chua [4], Chua [5] showed a Public Secondary School Exploratory Analysis that quantify the dropout rate from years 2013 to 2015. In it, Region XI landed on the thirteenth place among the regions with highest dropout rates in all three years. Whilst, Region II listed as the first in rank. In addition, Division of Davao del Norte had 23,956 enrollees and a dropout rate of 3.25% in 2013; 26,042 enrollees and a dropout rate of 3.16% in 2014; and 27,128 enrollees and a dropout rate of 4.48% in 2015. In 2014, there is a decrease in dropout rate but on the following year – 2015, the line is started to climb because of the full implementation of the Republic Act No. 10533 entitled “An Act Enhancing the Philippine Basic Education System by Strengthening Its Curriculum and Increasing the Number of Years for Basic Education”, otherwise known as the “Enhanced Basic Education Act of 2013” - adding two years to a 10-year acquisition of basic education; it is also commonly known as “K-12 program”. Lisay [6], Mateo [7] posted a news that there were an estimated 400,000 junior high school completers have dropped out with a dropout rate of 27%. Philippine Statistics Authority had conferred on [8] that there is a total of 36,238 people ages from six to twenty-four who are out-of-school. This poses a predicament if this will continue without even attempting to prevent or at least reduce it gradually, to the governing body (DepEd) and ripple its effect to the entire country. Martin [9] discussed that UNESCO warns Southeast Asian countries including the Philippines to take urgent measures to lessen the number of out-of-school youths and enroll them to avoid economic consequences in the future such as losing billions of dollars and a decrease of the percent of gross domestic product (GDP).

There are few empirical pieces of research had set its focus on the issue about dropouts but only a little research in

Manuscript received November 20, 2019; revised April 24, 2019.
Maricel A. Timbal is with the Department of Education, Division of Davao del Norte, Philippines (e-mail: maricel.timbal@deped.gov.ph).

doi: 10.18178/ijmlc.2019.9.3.798 273
number applied educational data mining set in the Philippines. This paper targets to address this gap. It aims to analyze the data set of SARDO’s from the Division of Davao del Norte specifically the senior high school students of Kapalong National High School who had dropped out from school using the decision tree and then convert to rule-based supervised learning to provide a predictive model to forecast who, among the enrolled students, might drop and intend to leave the school in relation to the given combination of attributes which had caused a student to drop. Moreover, with the findings it provides, this study will redound to the following beneficiaries: (1) Policy maker could amend if there is already an existing policy governing on funding Department of Education solely for the dropout and Department of Social Welfare and Development to encourage out-of-school youth to go back to school and continue their studies; (2) Department of Education Officials could make this as a guide in making a strategic plan towards intensifying Drop Out Reduction Program (DORP); (3) School Administrators will be encouraged to create an action plan for reducing the dropout rate and give much attention to strengthening “Oplan Baling” program; (4) Teachers will be notified of the predictors causing SARDO to drop and exercise early remedial session; and (5) Students will be given emphasis to manifest their right for education.

II. THEORETICAL FRAMEWORK

A. Student Dropout Factors

Student dropout is prevalent in all educational institution that is the reason why numerous of empirical studies attempt to develop efficient methods for prediction enabling to reduce risk and manage to adopt proactive measure [10].

In Doll et al. [11] study, they had made a comparative analysis of seven nationally representative studies to understand why students drop out of high school. They categorized it into three factors that affect the student’s decision; these are: (1) Push factors include school-consequence on attendance or discipline; (2) Pull factors include out-of-school enticements like jobs and family; and (3) Falling out factors refer to disengagement in students not caused by school or outside pulling factors. The first two factors are extrinsic, and the latter is intrinsic.

In the United States of America, a research conducted by Burrus et al. [12] about dropping out of high school came up with four categories of factors affecting them to drop out; these are (1) Demographics include a student comes from low-income family, he/she is a member of racial or ethnic minority group, male, and older than the average student in their grade; (2) Performance include lack of credits earned, poor attendance, and poor grades especially in core courses; (3) Self-identified factors about include class not interesting, lack of engagement with school, test too difficult, and poor attendance; and lastly, (4) Self-identified factors about others include adults did not expect them to perform in school and parents not involved in education. Another study made by Fall er al. [13] was conducted in the same country about dropouts suggested that social contextual includes family support, teacher support, and peer support; self-system includes perceived identification with school and perceived control by the extent to which it meets or ignores, as coined by Fan et al. [14] as student’s educational expectations; and school engagement mediates the relation between self-system process and student outcomes, are variables contribute to decisions to drop out. Moreover, Eck et al. [15] cited that chronic absence is one major problem in schools. School climate impacted student chronic absence rate and eventually leads to dropping out. In 2017, Hughes et al. [16] investigated the effect of grade retention in elementary school found out the implication of the students who were retained in the elementary grades would be likely to leave school than their promoted peers.

In Zimbabwe, a research conducted by Chinyoka [17] figured out that poverty in household, child labor or household chores, broken families, poor supervision by parent, involvement in bad company or peer pressure, drug abuse, and malnutrition and health-related issues, low self-motivation [18] and lack of interest in education, are variables of dropping out.

In Turkey, Alkan [19] had found out that the predictors are non-academic factors. Students’ dropout intention relates to intrinsic indicators such as school membership, loneliness, and coping strategies.

In Lebanon, Al-Hroub [20] explored the perspective of school dropout’s dilemma, on his findings, socioeconomic status, school curriculum and services, family involvement, and domestic laws governing the participation of Palestinian refugees.

In Brazil, Fernades et al. [21] identified the main factors influencing secondary dropout. These are all intrinsic factors significant to influence the decision to drop out, such as difficulties faced with courses (subject), desire for different school, the perception of better job opportunities, and importance assigned to school choice.

In India, Hegde [22] reduced the fifty-one (51) attributes affecting students to drop out. A multidimensionality of dataset was reduced to seven (7) through Principal Component Analysis. These were financial instability, an ability to cope with advanced courses (subject), university examination policy, lack of attendance, lack of motivational factor, and policy matter of the institution. On the other hand, Sharma et al. [23] cited that the reason why the student will likely to drop from school is the aggression they had experienced whether in face-to-face violence or electronic bullying.

In Canada, a study conducted by Ricard et al. [24] revealed that there is a role of parent and teacher to support basic psychological needs of the student. Other significant factors are reciprocal friendships and academic motivation considerably evident contributing to the prediction of dropping out.

B. Data Mining Concept

Data mining is a process of extracting actionable information from a large scale of data through identifying valid, interesting, novel, useful, and ultimately understandable patterns hidden in the data [25], [26]. Through data mining techniques it can yield the benefits of automated prediction of trends and behaviors, or automated
discovery of previously unknown patterns [27]. Data mining is also popularly known as Knowledge Discovery from Data (KDD) [28]. It consists of an iterative sequence of the following steps: (1) Data cleaning, (2) Data integration, (3) Data selection, (4) Data Transformation, (5) Data mining, (6) Pattern evaluation, and (7) Knowledge presentation.

C. Related Work

As shown in Table I, there were five similar studies that had set its focus on dropout and are using data mining technique. Nonetheless, there was no study conducted in Southeast Asia of which was warned by the UNESCO of the negative implication to its economic status.

III. OPERATIONAL FRAMEWORK

Fig. 1 shows the architectural design of the study. The information being used in the study were gathered from the Guidance Counselor of Kapalong National High School. It is the home visitation form of all senior high school SARDO’s. Guidance Office was the repository of all home visitation form of the Department of Education and were preselected in relation to the student dropout factors evidently presented by related studies.

A. Data Used

As shown in Table II, the attributes of SARDO dataset were renamed for the reason of convenience in the scripting in RStudio. These eighteen attributes were taken from the actual home visitation form of the Department of Education and were preselected in relation to the student dropout factors evidently presented by related studies.

B. Decision Tree Classifier

A decision tree is a flowchart, with structure of tree, where the internal nodes represent validations on the attributes, the branches represent the outputs of the validations, and leaf nodes represent classes. The node in the upper part of the tree is known as the root node. To classify an "unknown" instance, the flow of the tree from the root and down, according to the values that have the attributes for each node, and when you reach a "Leaf node", the instance is classified according to class assigned by said node [33].

C. SARDO Analysis

As a first step, a sample was drawn of student data. The information was generated in an Excel file separated by commas (CSV) to have greater flexibility when exporting the data. It should be noted that this first sample contains sensitive information from Home Visitation Forms that the teacher adviser fills in at the time of visiting the SARDO of which the guidance counselor takes care of after the visit.

Once the data were loaded in RStudio, two partitions were made for the training data and testing data. Out of fifty-six (56) observations, thirty-eight (38) tuples were assigned to training data and the remaining eighteen (18) were assigned in testing data.

As shown in Fig. 2, XRET was the root of the tree. This means that if SARDO had already experience being retained in his/her previous grade level/s, it is most likely that he/she will drop from school. On the left subtree, GAPL is what it means to previous grade level (e.g. if SARDO is grade 11, then his/her grade 10 GAPL is 11). XRET is the child of XRET. This means that if SARDO, if he/she did not experience retention from previous grade level/s but has a fairly satisfactory scholastic record as general average of previous grade level (e.g. if SARDO is grade 11, then his/her grade 10 GAPL is still grade 11) attained, then it bears a significant implication to dropping out. Generated rules were extracted from the decision tree on Table III.
Based on Fig. 3, NOS is the prime predictor. It says that if SARDO has more than six siblings, then he/she will likely to drop. However, if he/she has lesser than six siblings, GAPL is considered as the next predictor. It interprets that if his/her GAPL is either fairly satisfactory or satisfactory, and NOS is greater than 4 but less than 6, then he/she will likely to drop. Otherwise, he/she will continue to study. On Table IV is the generated rules based from decision tree using rpart.

The decision tree of training data as shown in Fig. 4, XRET was the root of the tree implies that it is the most significant factor after thirty-eight tuples were being processed. Five from it was classified as dropout. Furthermore, another condition is sprouted from the root which is the GAPL. It simply means that if SARDO is not a returnee but his/her GAPL is either from satisfactory or fairly satisfactory, then he/she will be a dropout.

The decision tree of testing data as shown in Fig. 5, NOS was the root of the tree. In here, testing data which has eighteen tuples happened to have a NO value in DO in all records. Upon cross-validation with training data, prediction says if SARDO is having more than 4.5 siblings, then he/she will likely to drop. Although there is no fractioned number of siblings, the value there is a result of analysis having a continuous data.
teacher adviser to expect from their students who have the same attributes seen in the rule-based classifier compared to their student information sheet upon enrolment so that teacher adviser will not necessary practice home visitation and will not wait for the student to commit five (5) consecutive absences.

Furthermore, it implies that if student has an experience being retained from previous grade level/s, he/she is more likely to drop. If he/she has not experienced retention but his/her number of siblings is greater than or equal to six (6), most probably he/she will drop from school. If the student has not experienced retention nor has greater than or equal six (6) siblings but his/her general average of previous grade level is fairly satisfactory and satisfactory, then he/she will probably stop from schooling.

IV. CONCLUSION

This study had presented an intelligent predictive model generated from DepEd's Home Visitation Form of senior high school students of Kapalong National High School.

Having this intelligent predictive model, educational institution with its stakeholders can now have a proactive measure on addressing to reduce the inflation of dropout rate of which is the sole purpose of why this study was conducted.

This study has a small size of observations due to the fact that senior high curriculum has just started its full implementation, and Kapalong National High School has just a population of nine hundred twenty-eight (928) as of school year 2017-2018, this pose for further study having greater observations to generate more reliable rule-based classifier that would represent the entire DepEd's SARDO. Moreover, the author recommends that this intelligent predictive model will be evaluated for its precision.

REFERENCES


Maricel A. Timbal is a senior high school teacher of Kapalong National High School in the Division of Davao del Norte.

She’s been teaching since 2008 in STI Tagum. She teaches office productivity, computer programming of the following languages/framework: java, jsp, php (with MySQL), .net, C#, android; data structure (with java programming); and systems analysis and design. In 2014, she transferred to Kapalong College of Agriculture, Sciences and Technology as a program coordinator of bachelor of science in information technology. In the advent of the full implementation of K-12 curriculum in the Philippines, she then transferred to the Department of Education teaching ICT related courses. In her ten years’ teaching period, she finds her profession a noble one influencing young individuals to be responsible citizens. In the place where she lives, ICT is not fully realized that is why she remains to teach in a rural place to uplift their being as to be globally competitive.