A CASE STUDY ON STUDENTS' ACADEMIC PERFORMANCE EMPLOYING EDUCATIONAL DATA MINING

[1]SALEM M. LAYLO, [2]JASMINNIGUIDULA

Technological Institute of the Philippines
E-mail: [1]salemcoe@yahoo.com, [2]jasniguidula@yahoo.com

Abstract - Data Mining Technology helps an organization to easily recognize the trends, patterns and visualization of the data available in a database. In an educational institution, the process is called Educational Data Mining which was applied on this study. The academic performance, gender, subjects and type of originating schools are the attributes which were derived from the database source. In order to process the extracted data, the Knowledge Discovery Process model was used. The Classification technique was utilized in order to categorize the attributes of the data collected and the rule based classifier using a set IF-THEN rules was also used to represent the information that was derived. The primary goal of the case study is to determine the level of academic achievement of the Grade 10 Junior High School from a select secondary private high school when grouped according to their originating schools, identify who perform better between male and female and determine the courses in which the students excel most and on the other hand find difficulties. The results of the study implied that originating school has a factor on academic performance of students, it was also proved that female performs better than male. Furthermore, it was identified that the academic performance rating of students in most subjects fall under the Proficiency level and revealed that Mathematics is the most difficult subject.

Keywords - Rule-based Classification, Academic Performance, Data Mining, Knowledge Data Discovery

I. INTRODUCTION

Academic performance of every student is a great factor on the achievement of a particular academic institution since it covers the measurement of how an academe promotes the quality of education to its clientele. Researchers have been interested on identifying the variables associated with the academic performance and this is through the use of the available data in the educational database.

In every research conducted in an institution, the most reliable source of information is the student data. Every information associated with it is to be considered in order to come up with good ideas for its succession plan. Continuous development of the various decision-making procedures such as planning, counselling, evaluation and other related steps should be done to meet the highest standards in education system. One approach to obtain this was to identify the unknown significant information by the top management of the educational institution. This significant information is concealed among the available data sets in the academe institutionand it can be retrieved by the use of various data mining technology. [1]

Education sector has a lot of datain the form of student's information. The utilization of computer in the education can extract valuable data that can be transform into an information which can be useful in decision making procedures. Due to the merging of education and data mining techniques, there are growing number of researchers became interested in this field of study. Data mining is utilized to describe significant information in data storage that involves retrieving and knowing meaningful information and the extracted significant data utilizing statistical, mathematical, artificial intelligence and machine learning technique.[2] Most of data that is available in an educational institution can be used to have a historical and operational data.

Knowledge Data Discovery which is one of the processes being used in data mining, pertains to the general way of discovering meaningful information from the stored data. It encompasses the assessment and possibly representation of the illustrative outputs to make the valuable resultsfrom which it succeeds as information. It also includes the various options of encoding methods, pre-processing, sampling, and projections of the data thru data mining functions. [3]

This study demonstrates the importance of data mining in an educational institution, which Knowledge Data Discovery model will be utilized. Its main objective is to identify the level of academic achievement of the Grade 10 Junior High School students when grouped according to their originating schools, to identify who perform better between male and female and to determine the courses in which the students excel most and on the other hand find difficulties. The study also aims to identify the level of understanding of the students using classification model. The proponent would like to come up with some interventions needed in order to solve the particular problems that will arise.

II. RELATED WORKS

Siraj&Abdoulha (2011), assessed enrolment data through Data Mining and utilized the CRISP – DM Method. From the school year of 1998 to 2006, a sum of 8510 students' enrolment inSebha University was gathered. The study was conducted to determine the

correlation of the following variables:teacher, sex and student status. The possible values from student status variable includes Enroll, Move, Expel, Quit and Completed the Study. Having the results of the study, it was found out that more female students have completed the study compared to male students who were taking Science (Obari and Sebha), Arts (Sebha) and Medicine (Sebha). However, Sport (Ghat) and Law (Sebha) courseswere dominantly completed by the male students. It was also revealed that there are63% students of the total population enrolled at the university were female compared to 37% male students. Furthermore, it was identified that the male enrollment rate was almost half of the female rate. The data obtained from the study can be beneficial to the office of the registrar and administrators to formulate plan necessary for the improvement of university's admission process. Additional variables such as student's year/semester of study and the academic achievement could be included to improve the previous model and to deliver other prediction models. Moreover, automation process was recommended for the information and the delivered knowledge. The results of the study were employed to improve the proportion of students' intake based on gender inSebha University and all public universities in Libya. [4]

Patil, investigated the academic performance of students by the use of classification, clustering and association rules. WEKA data mining tool was also employed to find out poor academic records of students and propose solutions to improve it. [5] In addition, Dhekankar&Datar, conducted an analysis of student performance by utilizing data mining technique, they grouped the respondents into grade level and employed data mining concepts to enhance students' performance and identify students' reasons of obtaining low academic performance. [6] Moreover, Kabakchieva focused on using data mining methods for classification to predict student performance and obtain new knowledge from data gathered by one university in Bulgaria. The main goal of the study revealed the advantages of data mining application for university administrators although the study revealed that the prediction rates are not remarkable. The classifiers workinversely for the five groups. Based on the study conducted, the factors that mostly influenced the classification process were the data attributes of the University admission results, the frequency of how failures occur and the first year university exams. [7]

Furthermore, the study of Siraj&Abdoulha focused on the demographic profile of students in Sebha University. It is similar with the study on terms that the gender of the students will be analyse to determine who performs well in academic. [4]

The researches of Patil, Dhekankar&Datar andKabakchieva was also parallel to the study being conducted. It was focused on the student performance and the solutions on how to resolve problems related to it. [5][6][7]

III. METHODOLOGY

Data Collection

The student data of the Grade 10 Junior High School students of an academic institution was used in the study. It covers the entry year which started from School Year 2014- 2015 up to School Year 2016-2017. The records utilized in the study consists of 10, 360 and preceding table shows the different attributes, description and possible values of the records that came from the source database.

Attribute	Description	Possible Values
Type of Originating School	The originating school of the students	Home-grown Transfer-In from Private Transfer-In from Public
Courses (Subjects)	The subjects taken by the students	AralingPanlipunan (APP)/ Christian Living Education (CLE)/ Computer/ English/ Filipino/ Foreign Language / Music, Arts, PE & Health (MAPEH)/ Math / Science/ Technology & Livelihood Education (TLE)
Gender	The Gender of students	Male/Female
Grade	Rating of the students on different courses	100 % - 70%
Descriptors	The prescribed equivalent of rating from Department of Education	(A)Advanced (AP)Proficiency Approaching (P)Proficient (D)Developing (B)Beginning

Table 1. Students' Record Its Attributes, Description and Possible Values

Table 1presented the variables used in data analysis and includes the type of originating school, courses, gender, grade, and descriptors, from the databasesource. There were ten (10) courses included on the conduct of the study.

Derived Model

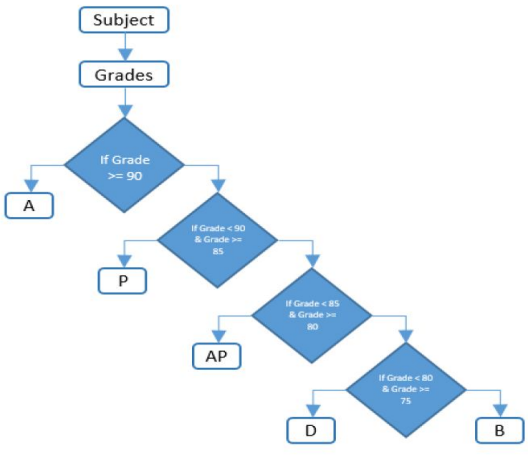


Figure 1. Derived Model for the Study

The derived model of the study was shown in figure 1, the rule based classifier using a set of IF – THEN rules for classification was used in order to represent the information or bits of knowledge that was derived in the database source. [8] To have a better understanding of the level of proficiency of the students in different subjects, the numeric values were converted and counted according to the level of proficiency that was set by the Department of Education. The grades of students were classified as A (Advanced), P (Proficient), AP (Approaching Proficiency), D (Developing) and B (Beginning) . [9]

Descriptors	SUBJECTS										MEAN	%
	APP	CLE	COM	ENG	FIL	FLANG	MAPEH	MATH	SCIG	TLE		
ADVANCED	27	58	78	34	36	56	45	46	40	54	47	20.21
PROFICIENT	73	99	98	79	62	85	79	75	81	84	82	35
APPROACHING PROFICIENCY	76	65	35	76	90	68	84	81	64	72	71	30
DEVELOPING	58	12	25	44	47	26	27	33	50	26	35	15
BEGINNING	0	0	0	1	0	0	0	0	0	0	0	0.043
TOTAL	235	235	235	235	235	235	235	235	235	235	235	100.00

Table 2. Number of Students According to its Level of Proficiency

Employing the derived IF-THEN rules, table 2 represents the frequency of the converted value of grades into its level of proficiency.

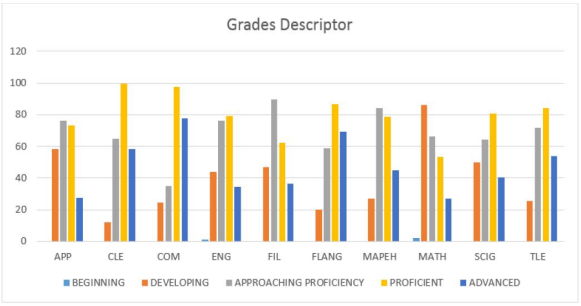


Figure 2. Grades Descriptors According to the Level of Proficiency

Moreover, to further validate the academic achievement of the Grade 10 students, Figure 2 presents the mean averages of students for three consecutive years on the different subjects. Noticeably, most of the students fall under Proficient Level particularly in CLE, Computer Subject, Foreign Language, TLE, Science and Technology and English. Based on the Level of Proficiency set by the Department of education the results indicate that the students have established the important knowledge and skills essential understandings, and can transfer them independently through authentic performance tasks. There were also students who fall under the Advanced level, the Computer subject has the most number of counts, followed by Foreign Language then Christian Living Education. It shows that the learner has exceeded the core requirements on the said subjects in terms of knowledge, skills, and understanding, the learners can show manifestations on bringing out the acquired knowledge and dynamically complete the performance tasks. [9]

In additionally, Filipino, MAPEH and APA subjects have the greatest number of frequency under the Approaching Proficiency. The students at this level have developed fundamental knowledge and skills and core understanding, can accomplish the tasks with a little guidance from the subject teacher and / or with some assistance from peers [9]

On the other hand, Mathematic subjects falls under Developing Proficiency. It indicates that the learner possesses the minimum knowledge and skills and core understandings, but needs help or supportin carrying out particular task. [9] There are also minimal number of learner that falls under Beginning on this particular subject, it implies that the learner is academically challenge and having a problem with his / her understanding; prerequisite and fundamental knowledge or skills which have not been acquired or developed adequately to support understanding. [9]

Data Processing

The Knowledge Data Discovery (KDD) employed on the study. The data were extracted from the database source. The inconsistent data were removed, then data needed was selected and perform the analysis. Figure 3 illustrated the KDD model.

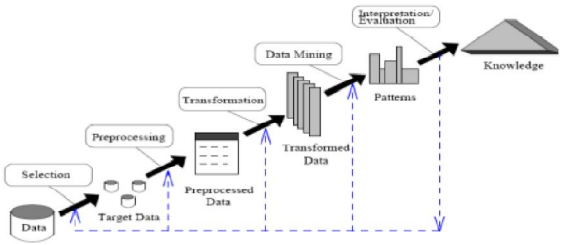


Figure 3. Knowledge Discovery Model (Priyadharsini, et.al, 2014 [3]

The data were analysed using the Tableau Desktop tool. Rule-based classification and visualization were used in data treatment. Tableau Desktop software was one of the data mining tools that offers data connection from different sources, data analysis, discovery, an interactive data visualization, build powerful calculations, drag and drop reference lines & point with trend analysis and generatereal-time reports. [10]

Classification technique was being used to arrange the data into different classes or groups based on its similarities or resemblances. [11] Rule-based is one of the classifications employed in the study.

On the other hand, data visualization was employed to easily recognized the trends, patterns and correlations of the data. Line graph and pie chart were used to represent the treated and analyzed data and support in the discussion of the results visually.

Table 2. Student Population According to Originating Intermediate School

Type of School	Population	Percentage
HOMEGROWN	55	23.40%
TRANSFER-IN FROM PRIVATE	97	41.28%
TRANSFER-IN FROM PUBLIC	83	35.32%
TOTAL	235	100.00%

Table 2 was defined in order to identify the number of home grown students, transfer in from other private schools and from the nearby public schools. The percentage was used to determine its allocation.

Type of School	Male	Female	Total
HOMEGROWN	24	31	55
TRANSFER-IN FORM PRIVATE	39	58	97
TRANSFER-IN FORM PUBLIC	38	45	83
TOTAL	101	134	235

Table 3. Number of Male and Female Students

Table 3 was identified to determine the number of male and female that came from the different originating schools. This was used to identify who performs better between male and female students. Based on the data collected, it was shown that there are more female students than male.

IV. RESULTS AND DISCUSSION

Based from the data that have been processed and analysed,the following results were derived:

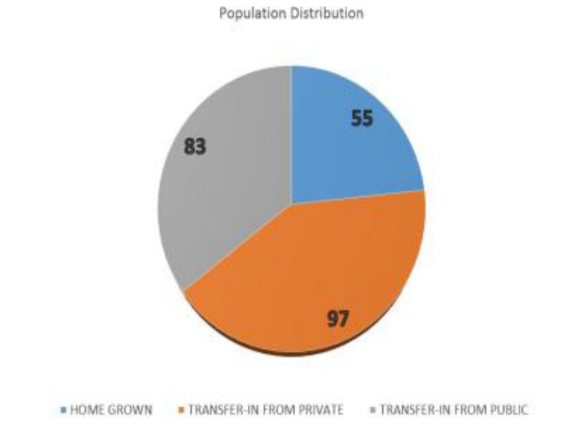


Figure 4. Student Population According to Originating Intermediate School

Figure 4 shows that majority of the Grade 10 students are transferees from the other private schools, it comprises the 41% of the population. Thirty five percent (35%) came from public schools and the remaining 24% came from the home grown students from the Grade School Department.

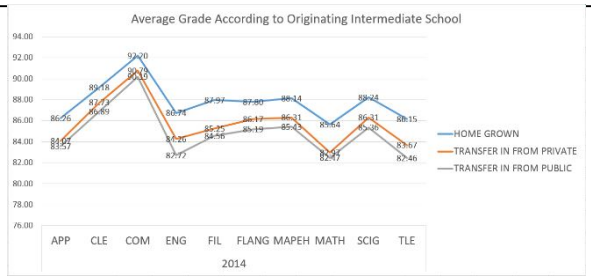


Figure 5. Average Grades of Students in Different Subjects According to Originating Intermediate Schools for School Year 2014 - 2015

Figure 5 indicates that for School Year 2014 -2015, home grown students got the highest average in all academic subjects, followed by the students from other private schools while students from public schools got the lowest average. Based from the graph, most of the students from different schools got a high grades in Computer subjects and they find difficulties on English, Mathematics and TLE subjects.

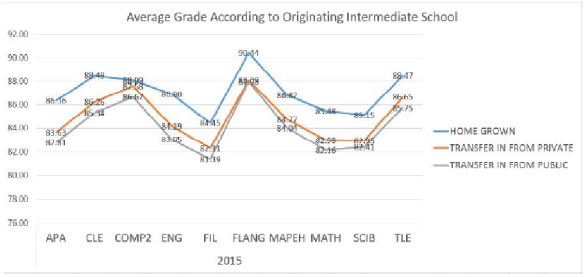


Figure 6. Average Grades of Students in Different Subjects According to Originating Intermediate Schools for School Year 2015 - 2016

For School Year 2015 – 2016, the home grownstudents also got the highest grades in all academic subjects. It was also followed by the other private schools and the least is from the public schools. It was shown on the graph that the students excel on Foreign Language, followed by the TLE, CLE and Computer subjects. The students from other private schools and public schools obtained a lower grade in Filipino, followed by the Math, Science, APA and English subjects.

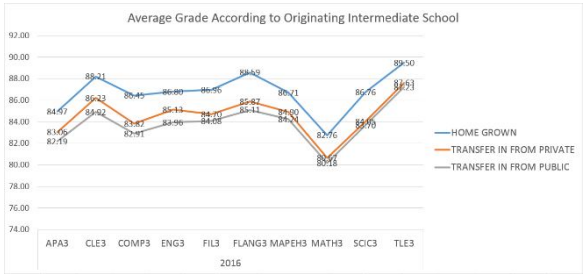


Figure 7. Average Grades of Students in Different Subjects According to Originating Intermediate Schools for School Year 2016 – 2017

From the data presented in Figure 7, it indicates that home grown students obtained the highest averages on the different academic subjects. While there is a

slight decrease on the averages when the data of the other private schools are compared with the public schools. During their Grade 9 level their academic performance is quite low when compared to the previous school year. Math is the subject that most of the students find difficulty. They got a highest mean average of grades in TLE.

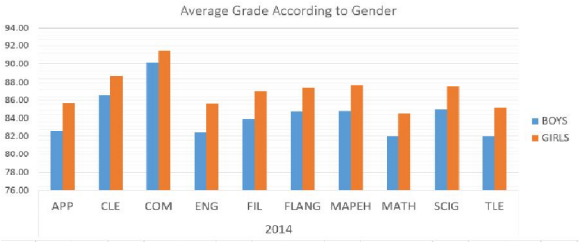


Figure 8. Average Grade of StudentsAccording to Gender for School Year 2014-2015

The result illustrated on figure 8shows that, female performs better than male students. For School Year 2014-2015, they acquired high grades in Computer, followed by the Christian Living Education, then by Science and Technology, MAPEH and Foreign Language that falls on the same level. It was observed that grades of students in Math was seen to be on the lower part of the graph followed by English though it was seen that female students got higher grades compared to male students on both courses, still these two courses are in the lower grade range.

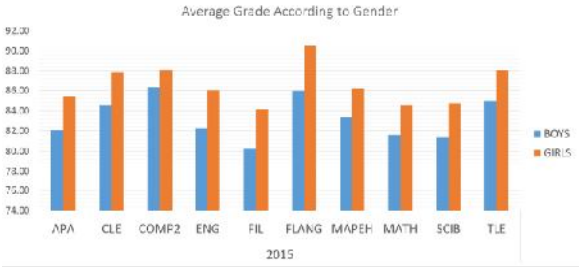


Figure 9. Average Grade of Students According to Gender for School Year 2015 – 2016

Furthermore, home grown Grade 10 female students are still the prevailing students as far as academic achievement is concerned as reflected on Figure 9. They got the highest grade in the Foreign Language subject while the male obtained the highest average in Computer subject. At this year level both male and female had a difficulty in Filipino subject.

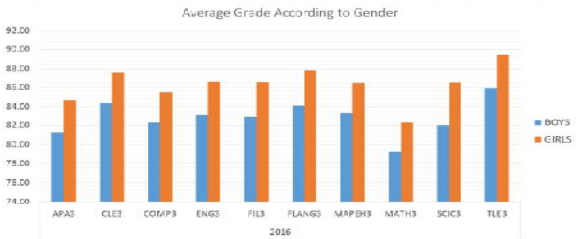


Figure 10. Average Grade of Students According to Gender for School Year 2016-2017

However, it is noticeable on Figure 10 that Grade 9 level is difficult as reflected on their academic performance compared to the data on the previous school year (2014-2015and 2015-2016). It was reflected that Mathematics is the most difficult subject.

CONCLUSION

This case study introduced the importance of data mining in an academic institution. The organization became aware that there is valuable knowledge that can be found on the existing database that is useful for the enhancement of the academic programs. The study revealed that based from the originating schools, the home grown students mostly excel on the academic subjects, followed by other private schools and from the public schools. The institution should provide intervention programs for the transferee student for them to cope up with the educational system of the institution understudy. This maybe in the form of bridging program, tutorials, peer tutoring, remedial classes and alike.

The academic team should review the learning modules of the Mathematics subject since most of the students fall under Developing level. The institution may provide learning activities that would urge student to study and learn Mathematics.

REFERENCES

[1] R. B..Mohammad , "Application of enhanced analysis model for data mining processes in higher educational system". ResearchGate Publication, 2005,pp.1-2

[2] E. Turban,et al, "Decision Support and Business Intelligent Systems. Prentice Hall of India, 2007

[3] C. Priyadharsini&Dr. A. SS.Thanamani, "An Overview of Knowledge Discovery Database and Data mining Techniques",International Journal of Innovative Research in Computer and Communication Engineering,vol.2, special issue 1, pp. 1-2, March, 2014

[4] F. Siraj&Abdoulha, "Mining enrolment data using descriptive and predictive approaches." inKnolwledge – Oriented Application in Data Mining, Prof.KimitoFunatsu (Ed.), January 2011, pp. 53-55

[5] D. P. Patil, "A Study Of Student's Academic Performance Using Data Mining Techniques", International Journal Of Research In Computer Applications And Robotics, vol.3,issue 9, p. 61,2015

[6] M. P. M. Dhekankar& D. S.Datar, "Analysis of Student Performance by using Data Mining Concept", International Journal on Recent and Innovation Trends in Computing and Communication, vol. 3,issue 5, May 2015 pp. 2942.

[7] D. Kabakchieva, "Predicting Student Performance by Using Data Mining Methods for Classification",The Journal of Institute of Information and Communication Technologies of Bulgarian Academy of Sciences, vol. 13, issue 1, pp. 61 -71., 2013

[8] J. Han et al, "Rule-Based Classification," in Data Mining Concepts and Techniques, 3rded, Elsevier Inc, 2012, p.355

[9] Luistro, Armin (2012, April). "Policy guidelines on the implementation of grades 1 to 10 of the k to 12 basic education curriculum (bec) effective school year 2012-2013)." [Online].Available: http:// www.deped.gov.ph/sites/defaults/files/order/2012/DO_s2012_21.pdf, p.13, [Sept. 5, 2017]

[10]Tableau Software. (2017). Tableau Desktop. Internet: https://www.tableau.com/products/desktop [Sep. 5, 2017].
[11]Ranjay (2013, January). "Define 'Classification'. Explain the purpose and methods of classification of data giving suitable examples."Internet: https://ignoubcom.wordpress.com/2016/01/13/define-classification-explain-the-purpose-and-methods-of-classification-of-data-giving-suitable-examples/ [Sept. 5, 2017]

★★★