USING REGRESSION ANALYSIS IN IDENTIFYING THE PERFORMANCE OF STUDENTS IN THE BOARD EXAMINATION

Sheila A. Abaya, Danzel Anerfee D. Orig, Richard S. Montalbo
University of the East, Caloocan – Philippines
sheila_abaya@yahoo.com.ph, danzelanerfee@gmail.com & rb_montalbo@yahoo.com

Abstract: This research focuses on developing a model prototype of predicting the possible student passers from a pool of board takers. The project focused on the performance of students who will take the board exam. Several attributes were included and identified as variables for prediction such as academic grades, age, gender and pre-board scores. The research project identified who amongst the pool of board takers will pass or fail the board exam and the passing percentage of the institution if it qualifies the national passing rate. The model prototype served as a preparatory tool for board takers to prepare for the examination and also an aid to the institution to plan and train even more their students before taking the board exam. Prediction is incorporated in the prototype using linear regression analysis of data mining. The predicted value was validated using a machine learning tool to identify its accuracy.

INTRODUCTION

Board examinations are given by professional regulatory agencies in various countries to its citizens to ensure that desired efficiency of the job is achieved, especially those which are critical to the society. The assessment also aims to measure the person's competencies and abilities in performing their jobs in the industry. Every year, new batch of examinees take their chance in passing the exam and becoming licensed professionals. However, there are times that the percentage exam result does not meet the national passing rate, resulting to an unfavorable reflection of an institution's performance.

Forecasting board examination results to set prognosis of the institution's efficiency in delivering instruction can be done using a data mining technique. Data mining (DM) is considered to be the latest technology in analyzing big data from various outlooks, summarizing voluminous pool of data and converting it into a useful and meaningful one, which can be of valuable resource to the business process. DM can identify behaviors of data and predict future trends through patterns, which can be of help in decision making.

Thus, this study aims to develop a predictive model that can identify figures of probable successful and unsuccessful examinees in the board examination and the possible passing rate percentage of a given set of board takers. Predictors were identified as bases of identifying the performance of the students before taking the exam. The technique of regression analysis of data mining was incorporated in the research project. It is hoped that using this technique would contribute further to the improvement of student performance in the pre-board examination and in identifying the probability rate of passing or failing the exam. This approach, if institutions allow, expects to better improve students' performance and permits anticipation of possible percentage result if it qualifies the national passing rate.

RELATED STUDIES

Part of the investigation process of a research is to have a survey of the previous studies which will acquaint you with needed knowledge about a certain area of research. It definitely helped in defining and understanding the problem due to its several suggestive techniques, tools and methods that have been used. Several studies have been considered to have a thorough glance, study and evaluation of the research.

Abubakak and Oguguo (2011) examined on the correlation of age and gender on academic achievement (CGPA) in mathematics and science students. Scatter-plot, mean and Standard deviation were used for the descriptive statistics while nivariate analysis of variance (ANOVA) and multiple regression were used for the inferential statistics. Results revealed a linear relationship between age and gender. A low positive correlation coefficients was obtained for ages and gender (r=0.030 and 0.111) which is significant. The predictor variables jointly accounted for 1.3% of the variance, gender was the better predictor.

Jabor and Kungu (2011) researched in determining if age and gender in high school influences the students' performance in mathematics. The study used the students' grade point (GPA) in mathematics during high school years to measure achievement. The study described the students by age, gender and academic achievement in mathematics. The comparison revealed that there were significantly differences in mathematics GPA scores between age group and gender, however, effect sizes were small.

Kanyongo, Certo and Launcelot (2006) focused on the use of linear regression analysis through structural equation modeling using AMOS 4.0 the relationship between home environment factors and reading achievement in Zimbabwe. The study revealed that social economic status (SES) is the strongest predictor in reading achievement in Zimbabwe.

McCarthy, Padgham and Bennett (2006) identified factors that influence student learning in college macroeconomics and microecono such as gender and personality type, college entrance exam scores, grade point average, class size, and whether the course was micro or macro and was measured by Test of Understanding College Economics III (TUCE). Results revealed that there is no statistically significant influence on student achievement from college entrance exam scores or class size. Gender and GPA were significant explanatory factors for performance in principles of both mircoeconomics and macroeconomics.

Ong, Palompon and Banico (2012) identified the predictors in identifying the nurses' licensure examination performance of graduates. The study used the descriptive design using inferential techniques. Factors such as College entrance examination performance on IQ test, aptitude test, composite scores of science, math and English tests, college grade point average and pre-board examination performance were studied. Results revealed that all variables are significantly correlated with the licensure examination performance and that there were only two significantly predicted licensure examination which are college grade point average and pre-board examinations. Thus, these variables are the relevant bases in determining the success and failure of students' licensure performance.

FRAMEWORK OF THE STUDY
The framework of this research was formulated to have a preconceived notion of the study and will serve as the research' guide to know what is going on and what is expected to learn from the study.
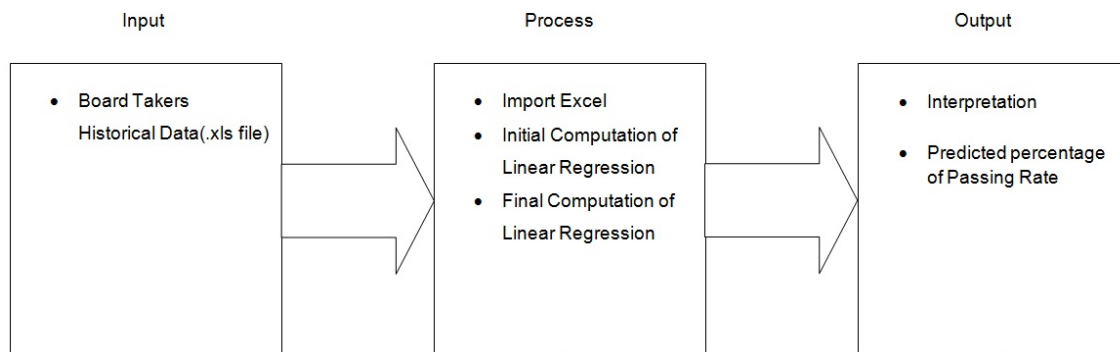

Figure 1.0 Conceptual structure of predictive model

Figure 1.0: Presents the prediction model wherein the historical file which was imported from MS Excel was loaded into the tool. The file is composed of the attributes defined in Table 1.0 (a), (b), (c) and (d). These attributes (age, GPA, gender and pre-board grade) were used as predictors in identifying the probable performance of students before taking the test. The file was loaded into the tool to apply the linear regression technique and as a result, students who pass or fail the exam have been identified and the possible percentage rate from the batch of students was determined which is an interpretation if the institution will or will not qualify for the national passing rate.

Table 1 (a): Defines the age data field used as a valid variable. It is used to identify the age of board takers which is considered as a factor in determining student performance.

Table 1(b): Defines the GPA or Grade Point Average of the student at the end of the program. It is used to determine the academic performance of the student before taking the board exam.

Table 1(c): Defines the gender of the board taker. It consists of nominal values only.

Table 1 (d): Defines the pre-board exam scores which could be the basis of student performance during the review process. Scores were taken before taking the actual board exam.

Table 1 (a) Age data field

Data Name	Age
Description	Age of the board taker
Data Types	Number [NUMERIC]
Length	2
Format	0-9
Use in	Determine the age of the board takers

Table 1 (b) GPA data field

Data Name	GPA
Description	GPA of board takers
Data Types	Number [NUMERIC]
Length	3
Format	0-9
Use in	Determine the GPA of the board takers

Table 1 (c) Gender data field

Data Name	Gender
Description	Gender of board takers
Data Types	VARCHAR
Length	1
Format	m/f
Use in	Determine the gender of the board takers

Table 1 (d) Pre-board data field

Data Name	Preboard
Description	Preboard grade of the examinee
Data Types	Number [NUMERIC]
Length	3
Format	0-9
Use in	Determine the grade of the board takers in Preboard

Figure 2: Illustrates the data flow of the proposed tool which started from the import historical data that has the records of the students who previously took the board exam. The tool only accepts the correct file that is saved in *.XLS* format and rejects the incorrect file. Then, data are plotted into a relation named as SPerformance where linear regression technique is applied. Finally, the predicted class result (Pass or Fail) is generated as well as the computed percentage rate.
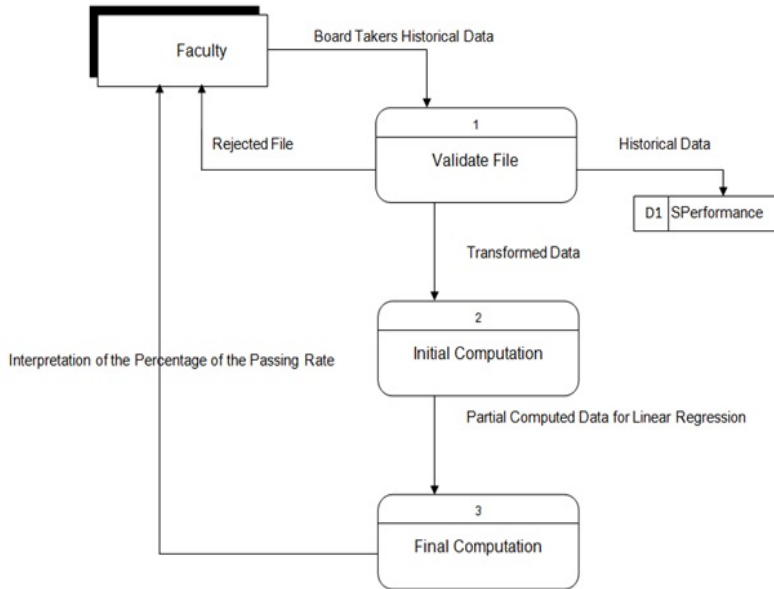

Figure 2.0 Data flow diagram of predictive tool

RESULTS AND DISCUSSION
This part consists of two parts: Predictive tool screenshots and the Model evaluation using a machine learning tool to determine the accuracy of the proposed prediction tool.

Predictive Tool
The following illustrations are presentation capability of the predictive prototype that was developed and presented as Figure 3, Figure 4, Figure 5 and Figure 6.

Figure 3: The loading window allows the user to open a file that is to be evaluated. The file contains the records of probable students who will take the board exam.
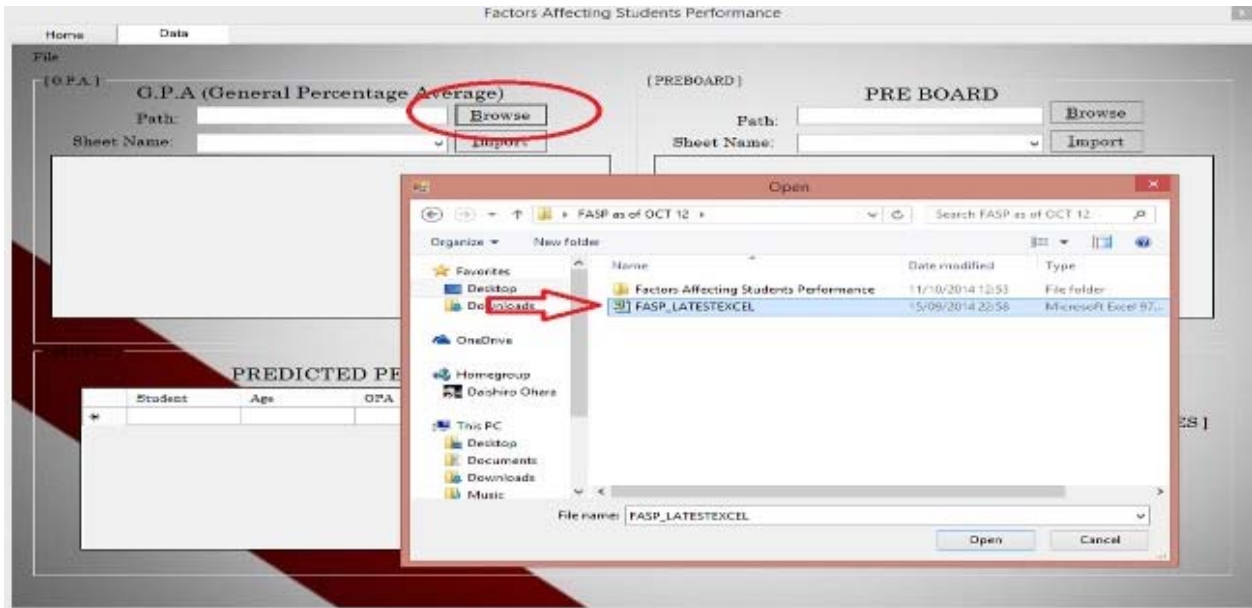
Figure 3.0 File loading window

Figure 4: A valid file will be loaded into a sheet table presented in columns and rows.
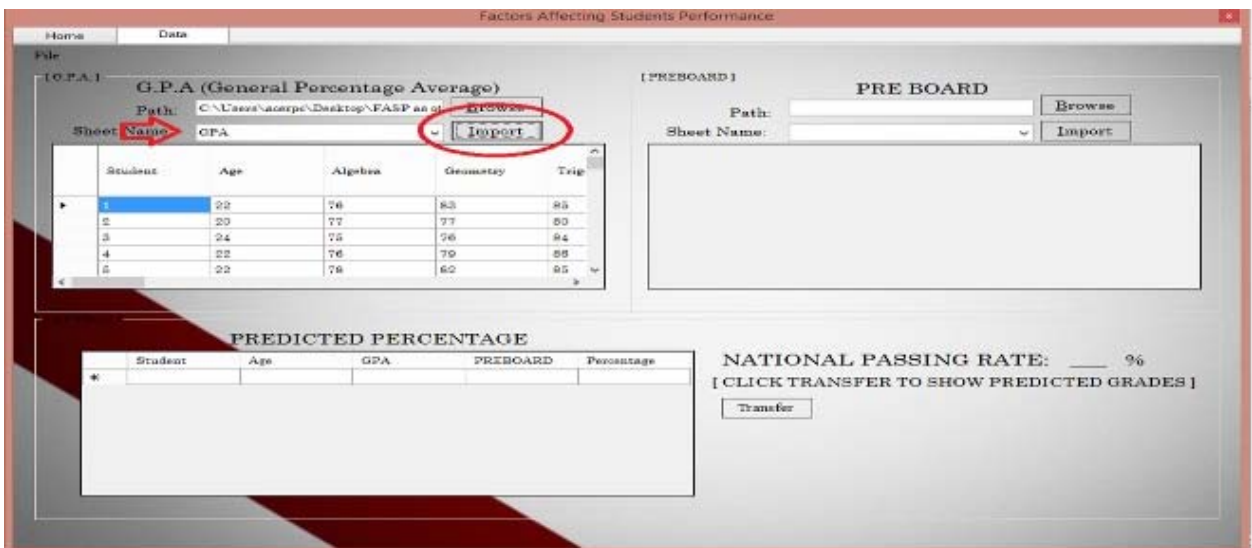

Figure 4.0 GPA loading window
Figure 5: Pre-board data was also loaded into the sheet table for evaluation vis-à-vis with the GPA records of the students
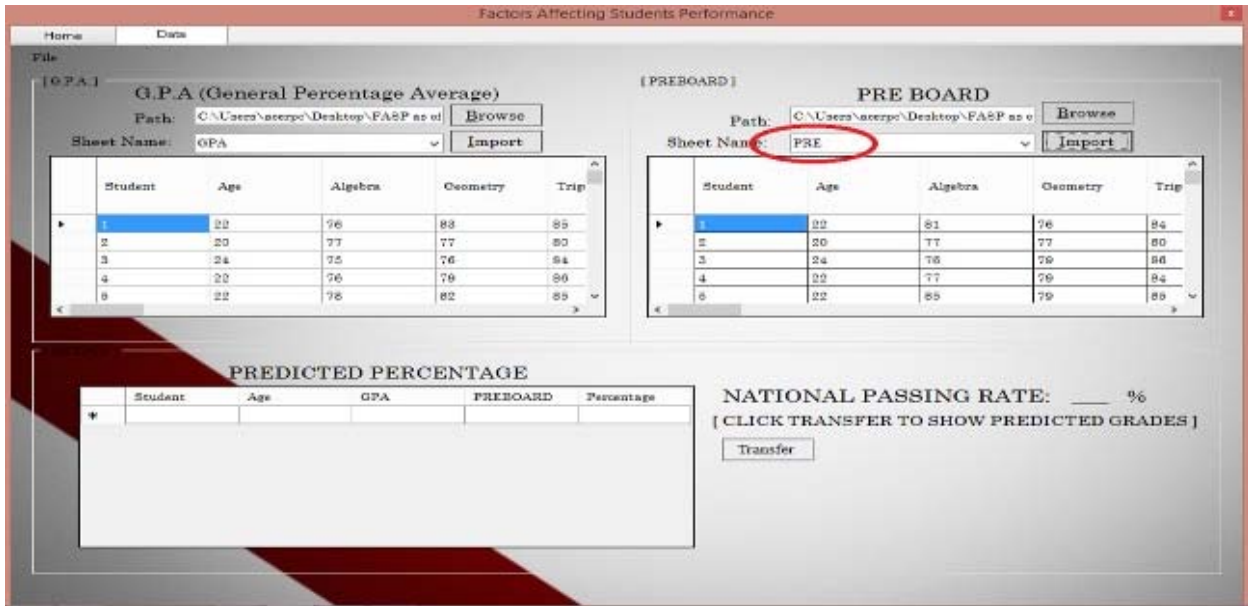
Figure 5.0 Pre-board file window

Figure 6.0: The "Transfer" button will evaluate GPA data with that of Pre-board data and the predicted percentage will be displayed which identifies the possible students who will pass or fail the exam. Based on the predicted percentage from a pool of data, the probable percentage rate will also be displayed to determine if the institution meets the required national passing rate.
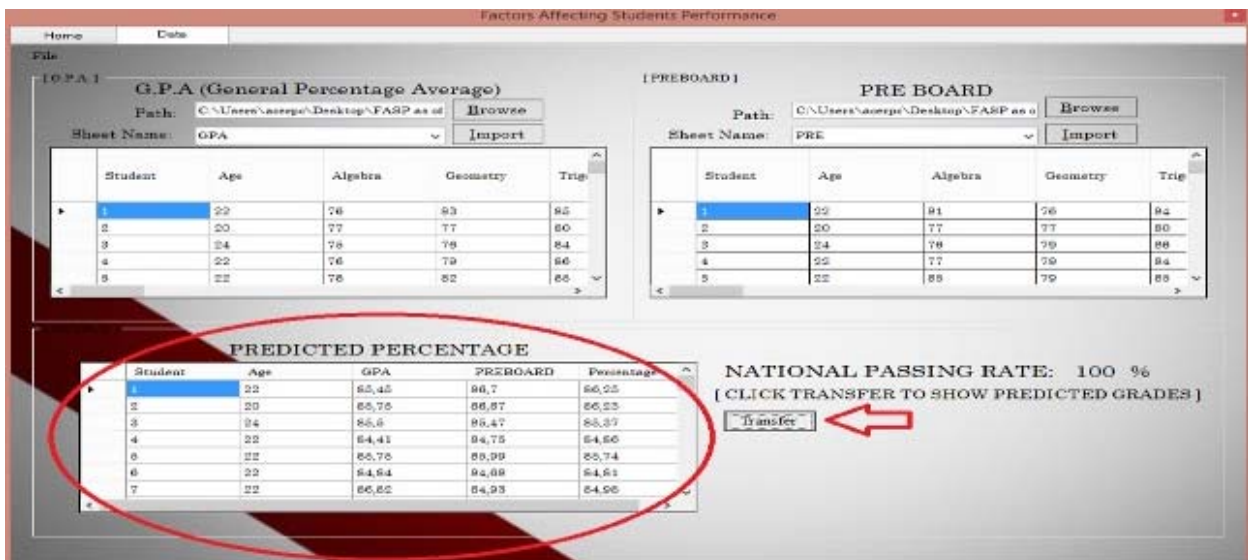

Figure 6.0 Final predicted result

MODEL EVALUATION
Data was evaluated using WEKA which is a machine learning tool from WAIKATO University, with the implementation of training set as a test option.

Table 2: Defines the Correlation Coefficient having a value ranging from +0.70 to +0.99 means there is a strong relationship between the actual value and the predicted value. Both the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) can range from 0 to ∞ and lower values show better results.

Table 2.0 Evaluation result using WEKA

TRAINING SET	
Correlation Coefficient	0.7003
Mean Absolute Error	8.5073
Root Mean Squared Error	10.3021

CONCLUSION AND RECOMMENDATION

With the new prediction tool, institutions can forecast the number of possible board passers from a batch of examinees and can also determine the percentage of passing rate. Results also revealed that factors such as age, gender, GPA and pre-board results can be used as determinants in predicting students' licensure examination performance. However, recommendations are advised such as:

- Implementation of other data mining techniques for prediction other than regression analysis
- Use of a larger dataset to be assessed using the proposed predictive tool to determine the behavior of the proposed tool
- Other factors to be considered as determinants for prediction

REFERENCES

Abubakar, R. & Oguguo, O. (2011). Age and gender as predictors of academic achievement of college mathematics and science students. *2011 Internatinal Conference on Teaching, Learning and Change.*736 – 742.

Kanyongo, G., Certo, J. & Launcelot, B. (2006). Using regression analysis to establish the relationship between home environment and reading achievement: A case Zimbabwe. *International Education Journal.* pp.632-641. 2006.

Jabor, M., Machtmes, K., Kungu, K., & Buntat, Y. (2011). The influence of age and gender on the students' achievement in mathematics. *2011 International Conference on Social Science and Humanity.*304-308.

McCarthy, C., Padgham, G., & Bennett, D. (2006). Determinants of student achievement in principles of economics. *Journal for Economics Educators.* 1-9.

Ong, M., Palompon, D. & Banico, L. (2012). Predictors of nurses' licensureexamination performance of graduates in cebu normal university, Philippines. *Asian Journal of Health.* 130 -141.