Prediction Models for Licensure Examination Performance using Data Mining Classifiers for Online Test and Decision Support System

Asia Pacific Journal of Multidisciplinary Research
Vol. 5 No.3, 10-21
August 2017
P-ISSN 2350-7756
E-ISSN 2350-8442
www.apjmr.com

Ivy M. Tarun
Isabela State University, Cabagan, Isabela, Philippines
ivy.m.tarun@isu.edu.ph

Date Received: April 3, 2017; Date Revised: June 21, 2017

Abstract – *This study focused on two main points: the generation of licensure examination performance prediction models; and the development of a Decision Support System. In this study, data mining classifiers were used to generate the models using WEKA (Waikato Environment for Knowledge Analysis). These models were integrated into the Decision Support System as default models to support decision making as far as appropriate interventions during review sessions are concerned. The system developed mainly involves the repeated generation of MR models for performance prediction and also provides a Mock Board Exam for the reviewees to take.*

From the models generated, it is established that the General Weighted Average of the reviewees in their General Education subjects, the result of the Mock Board Exam and the instance when the reviewee is conducting a self-review are good predictors of the licensure examination performance. Further, it is concluded that the General Weighted Average of the reviewees in their Major or Content courses is the best predictor of licensure examination performance.

Based from the evaluation results of the system, the system satisfied its implied functions and is efficient, usable, reliable and portable. Hence, it can already be used not as a substitute to the face-to-face review sessions but to enhance the reviewees' licensure examination review and allow initial identification of those who are likely to have difficulty in passing the licensure examination, therefore providing sufficient time and opportunities for appropriate interventions.

Keywords – *Performance Prediction Models, Data Mining, Decision Support System*

INTRODUCTION

The advancement of computers and Internet has been growing in leaps and bounds. This can be seen particularly in the education community as they are the forefront of most of the changes in information technology. Many educational institutions make use of databases that are far more efficient when it comes to storage and retrieval of large volumes of data. Such data includes employee records, student financial records, and student academic records. Undoubtedly, the student academic records occupy a greater space in the storage. More and more data are being accumulated over time. Unfortunately, these data were practically left as mere data. Hence, useful information from the database are hidden. Supposedly, these data can be valuable to capture information and knowledge.

One of the major breakthroughs in discovering potential useful information is the process of data mining. Generally, data mining is the process of analyzing large volumes of raw data to discover knowledge. It has several techniques that are used to discover meaningful information that may be used not only in business but in formative evaluation to assist educators in improving the quality of managerial decisions. Knowledge may refer to patterns, relationships, rules or predictions. Fan and Bifet (2012) concluded in their study that big data mining will continue to grow as data continue to be more diverse and large [1].

Sembiring, Zerlis, Hartama, Ramliana and Wani (2011) indicated that "there are increasing research interests in education field using data mining" (p. 110) especially so that data mining in education is relatively new and has a great potential to education institutes [2]. This implies that database administrators and developers and even researchers are given the chance to discover new knowledge and approaches in educational context which will benefit the general public as data mining can lead to better services.

For several years great effort has been devoted to the study of Data Mining in the different fields. It was only in the recent years that this was applied in the educational domain and is gaining importance. One of the key areas of data mining that is particularly interesting is the prediction of licensure examination performance of reviewees. Through data mining, an educational institution could predict who, among the students, will pass or fail an examination. They can use this information for additional support and review assistance on students who are at risk of failing the exam.

However, there are still very few researches that have been carried out concerning prediction of licensure examination performance. Most of the researches dealt with performance prediction relating to student transferability, retention, and success in classes like test scores in a particular subject and differences in learning. Baker and Yacef (2009) said that most of the studies about educational data mining were even undertaken in North America, Western Europe, and Australia/New Zealand with relatively low participation in other countries particularly in the Philippines [3]. On the other hand, there has been an absence of feedback and support after the pre-board exam in most of the licensure examination reviews. No interaction and less communication take place.

Thus, the researcher finds it necessary to explore educational data mining in the context of licensure examination performance prediction especially that educational institutions are continuously striving to improve their passing rate. This is because they are more focused at present on the number of students passing the board exam as output rather than the number of students enrolled as input [4]. This output manifests the institution's high standard of instruction and quality of students that they produce.

Results of this study will help enhance the teacher education graduates' success on the Licensure Examination for Teachers. The use of the predictors will allow initial identification of reviewees who are likely to have difficulty in passing the licensure examination, therefore providing sufficient time and opportunities for appropriate interventions during review sessions. The licensure examination performance prediction model that is generated in this study through data mining will help predict the eventual performance of the reviewees in the licensure examination. This appears to be a more important part of the study as this determines how much assistance will still be given to them.

While this is significant, it should then be applied to make it more valuable. Hence, the model was integrated to the Online Test and Decision Support System which will make way to the contemporary approach to pre-board exam management. This opens the opportunity for immediate feedback and support on the pre-board exam performance of the reviewer and eventually make remedial prior to taking the real licensure examination. This would help them cop success as well as for the institution to help improve its licensure examination performance rating. This can be seen as another essential way of strengthening the institution's accomplishments and may eventually lead to the path toward transformation in order to cope with the developing information and communications technology culture.

This study will likewise provide information for researchers, data mining enthusiasts and information technologists for the processes and application of data mining and decision support system integration.

OBJECTIVES OF THE STUDY

This study introduces the fusion of data mining and decision support system where licensure examination performance models are generated for prediction purposes and integrated into a decision support system to aid in the institution's and reviewees' decision making activities. Generally, it focused on the prediction of licensure examination performance using Multiple Regression and PART classification techniques of data mining for Online Test and Decision Support System.

Specifically, the study sought to generate a licensure examination performance prediction model based on Multiple Regression and PART Classification technique using the following factors: (1) Students' general weighted average during College in General Education, Professional Education and Major subjects, (2) Learning Style categorized into Visual, Auditory and Kinesthetic, (3) Review and Participation categorized into Conduct of self-review, Participation in peer study groups, Participation by asking questions, Taking down of notes, and Giving supplemental ideas, (4) Mock-Board Exam Result; and (5) Licensure Examination Performance. It also has the objective of developing a Decision Support System integrating the licensure examination performance prediction models. Prediction models generated are tested and implemented using real institutional data. The study also determined the performance evaluation of the system developed using the following criteria: (1) functionality, (2) efficiency, (3) usability, (4) reliability and (5) portability.

METHODS

As to the generation of the Licensure Examination Performance Prediction Models**,** Multiple Regression and PART classification techniques of data mining were used. Multiple Regression is one of the most popular mathematical models in making predictions [5]. In MR, the value of the response variable, Y, also called dependent variable or response attribute, is predicted using a linear function of predictor variables, X_1, X_2 … X_n, also called independent variables or predictor attributes, whose quantities are known. The MR equation takes the form of:

$$Y - \bar{Y} = \beta_1(X_1 - \bar{X}_1) + \beta_2(X_2 - \bar{X}_2) + \ldots + \beta_n(X_n - \bar{X}_n) \qquad \textbf{(1.0)}$$

where β_1, β_2 and β_n are the unknown coefficients, $\bar{X}_1$, $\bar{X}_2$ and $\bar{X}_n$ are the computed Means of the predictors, while $\bar{Y}$ is the computed Mean of the response variable. By simplifying equation 1.0, the MR model becomes:

$$Y = \beta_1 X_1 - \beta_1\bar{X}_1 + \beta_2 X_2 - \beta_2\bar{X}_2 + \ldots + \beta_n X_n - \beta_n\bar{X}_n + \bar{Y} \qquad \textbf{(1.1)}$$

To completely generate the MR model, the unknown coefficients should be supplied. These are computed using the covariance equation:

$$\text{cov}(X, Y) = \sum_{i=1}^{n}(X_I - \bar{X})(Y_i - \bar{Y}) / (n - 1) \qquad \textbf{(3.2)}$$

On the other hand, PART is a rule-based classifier that outputs a rule from partial decision tree in each iteration [6]. The PART model is in the form of

(Condition) → y

where Condition is the conjunction of attribute tests, and y is the class label.

The Data Mining Tool

The data mining tool used to generate the models was WEKA which offers a wide range of classification algorithms that can be easily applied to any dataset. WEKA is one of the most widely used data mining systems because of the many powerful features that are sometimes not found in commercial data mining software [7]. It contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. Since WEKA can be downloaded for free and has gained extensive adoption by researchers, then this is believed to be a successful medium for this study. Specifically, WEKA 3.7.4 version was used.

The Dataset

Table 1. The Attributes and their Values

Attribute	Description	Values MR	Values PART
GenEd (Predictor)	This is the general weighted average of the student in his general education subjects taken from his academic records.	1 to 3	E, VS, S, FS, G, FG, F, BF, P
ProfEd (Predictor)	This is the general weighted average of the student in his professional education subjects taken from his academic records.	1 to 3	E, VS, S, FS, G, FG, F, BF, P
MajorCore (Predictor)	This is the general weighted average of the student in his Major or Content subjects taken from his academic records.	1 to 3	E, VS, S, FS, G, FG, F, BF, P
Audio (Predictor)	This indicates that the learning style of the reviewee is audio.	1, 2, 3	F, S, T
Visual (Predictor)	This indicates that the learning style of the reviewee is visual.	1, 2, 3	F, S, T
Kinesthetic (Predictor)	This indicates that the learning style of the reviewee is kinesthetic.	1, 2, 3	F, S, T
SelfReview (Predictor)	This denotes that the reviewee conducted self review.	1, 0	Y, N
PeerStudy (Predictor)	This denotes that the reviewee participated in a peer study.	1, 0	Y, N
AskQuest (Predictor)	This denotes that the reviewee asked questions during the review.	1, 0	Y, N
TakeDown (Predictor)	This denotes that the reviewee took down notes during the review.	1, 0	Y, N
GiveIdeas (Predictor)	This denotes that the reviewee gave or shared her own ideas during the review.	1, 0	Y, N
MBResult (Predictor)	This is the score of the reviewee in the Mock Board Exam.	0 to 150	F, G, VG
LETPerf (Response)	This is the LET performance of the student which makes use of 2 classes.	0 to 100	P, F

The original dataset consisted of 73 instances, the values of which were obtained from the Students' Academic Records and Information and Accounting System (SARIAS) of Isabela State University Cabagan campus, the answers to survey questionnaire from the respondents, the respondents' performance in the mock-board exam, and the respondents' LET performance. These data were transformed to match that of the requirements of the algorithms used. The attributes were discretized from numeric to categorical ones for PART and from categorical to numeric for MR producing two (2) separate datasets but using the same data. The predictor and response attributes derived were shown in Table 1.

The categories used for the grade predictors were taken from the adjectival rating used by ISU while the other categories were set by the author. Their corresponding grade range and their meaning are given in Table 2.

Table 2. Meaning of Categories Used

Value	Meaning
E	Excellent, range is 1.0 – 1.12
VS	Very Satisfactory, range is 1.13 – 1.37
S	Satisfactory, range is 1.38 – 1.62
FS	Fairly Satisfactory, range is 1.63 – 1.87
G	Good, range is 1.88 – 2.12
FG	Fairly Good, range is 2.13 – 2.37
F	Fair, range is 2.38 – 2.62
BF	Below Fair, range is 2.63 – 2.59
P	Passed, range is 3.0
F	First
S	Second
T	Third
Y	Yes
N	No
F	Fair, range is 1 - 50
G	Good, range is 51 - 100
VG	Very Good, range is 101 - 150
P	Passed
F	Failed

The transformation process was continued by removing the duplicate instances which reduced the original dataset to 70 instances for PART and 57 instances for MR. The actual distribution of data used in predicting the LET performance is shown in Figures 1 and 2 where classes P and F are represented by colors blue and red respectively.

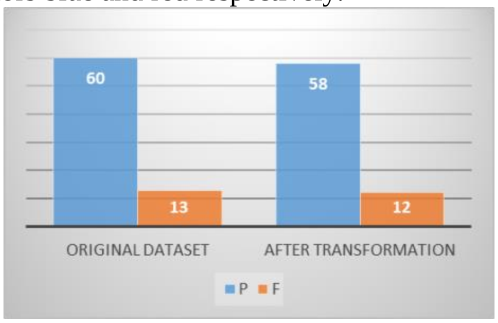

Fig. 1. Actual Distribution of LET Performance Data for PART Classification

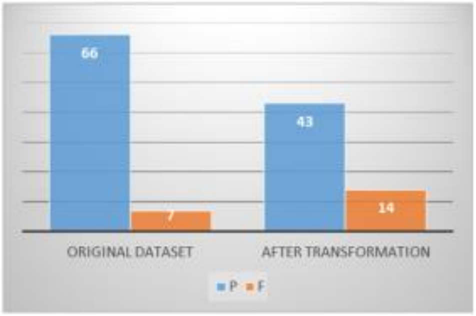

Fig. 2. Actual Distribution of LET Performance Data for Multiple Regression

Validation Procedures

In this study, the MR dataset and PART dataset were each divided into two, the *training and validation dataset*, and the *test dataset*. This is for the purpose of estimating the model parameters and assessing the predictive ability of the models. By doing so ensures the accuracy of the model. As far as validation in data mining is concerned, there is no set rule as to the proportion of data in each dataset but in this study 80/20 is used. This means that 80% of the data was utilized for the training and validation dataset while the remaining 20% was used for the test dataset.

The training and validation dataset is used to build the models and at the same time to confirm whether or not the model built does not over-fit the data. In order to perform the validation procedure for the training and validation dataset systematically, k-fold cross validation was used. This was chosen because of the scarcity of data.

The k-fold cross validation is done by first splitting the data into k equal sets or folds, where k in this study is equal to 10. Subsequently, k iterations of training and validation are performed such that within each iteration, a model is trained using k-1 folds of training data while the resulting model is validated using a different fold of the remaining data. The performance measure of the k-fold cross-validation is equal to the average of the values computed within the loop [8].

The model is then put to test using the test dataset to verify if the models have a similar fit.

Attribute Selection Procedures

A model is generated from the training and validation datasets of PART and MR with and without preprocessing procedures. The model generated without preprocessing covers all the attributes while the model generated with preprocessing determines the best predictors of the licensure examination performance. Comparison between the two models was done in order to determine their difference in terms of their performance and the models produced.

The training and validation datasets that involved preprocessing were classified with supervised attribute selection using CfsSubsetEval of WEKA. CFS (Correlation-based Feature Selection) evaluates subsets of attributes rather than individual attributes. It considers the worth of individual features for predicting the class together with the level of inter-correlation among them [9]. Use of attribute selection ensures that the model generated does not involve irrelevant variables [10].

OTDSS Development

As to the development of the decision support system, *Rapid Application Development* (RAD) process was followed. The RAD was chosen because of its suitability to the project's scope, data, decisions, team, technical architecture and requirements. Figure 3 depicts the stages that were undertaken following the RAD process.

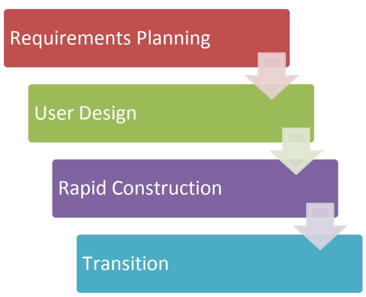

Fig. 3. Rapid Application Development Stages

Requirements planning is the first stage that was carried out in the development of the OTDSS. During this stage, the current situation was studied, the data were gathered, the system scope was defined, and the cost and schedule were estimated.

After careful identification of requirements, user design stage was carried out. Throughout the stage, the detailed activities and data associated with the OTOTDSS were analyzed, the screen layouts were developed, and the construction approach was selected.

In this stage, the system design that was described in the User Design stage was implemented. Hence, OTDSS was developed and tested accordingly. During this stage, the fastest possible cycle of designing, coding, testing, modifying, and re-coding the system was done to produce the acceptable OTDSS.

Traditionally, it is in this stage that the newly developed system gradually replaces the existing system. Since this project intends to strengthen the review and assistance extended to the reviewees of the licensure examination, this stage concentrated on how the OTDSS will be used to foster learning, to supplement review assistance and to make performance prediction for decision making rather than replacing the traditional face-to-face review sessions. Hence, it is within this stage where the system was prepared for utilization by the respondents and eventually for evaluation.

Data Flow Diagram of the Developed OTDSS

The Data Flow Diagram (DFD) is a graphical representation of the movement of data through the different processes in the OTDSS. Figure 4 shows the context-level DFD which represents the overall interaction between the OTDSS and its outside entities while Figure 9 shows the expanded DFD which represents the main functions of the OTDSS.

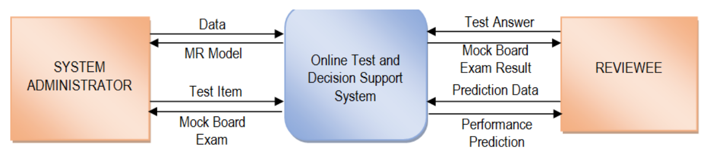

Fig. 4. Context-Level DFD of the developed OTDSS

The DFD contains two entities, the system administrator and the reviewee. The Raw Data and the Test Item are inputs coming from the System Administrator which will be used to generate the MR Model and Mock Board Exam respectively. On the other hand, the Test Answer and the Prediction Data are inputs from the Reviewee to produce the Mock Board Exam Result and Performance Prediction respectively.

Evaluation Procedures

System evaluation was done using the following criteria as defined in ISO/IEC 9126, the international standard for the evaluation of software quality. ISO/IEC 9126 does not provide requirements for software, but it defines a quality model which is applicable to every kind of software [11].

1. **Functionality -** A set of attributes that bear on the existence of a set of functions and their specified properties. The functions are those that satisfy stated or implied needs.

2. **Efficiency -** A set of attributes that bear on the relationship between the level of performance of the software and the amount of resources used, under stated conditions.
3. **Usability** – A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users.
4. **Reliability** – A set of attributes that bear on the capability of software to maintain its level of performance under stated conditions for a stated period of time.
5. **Portability** – A set of attributes that bear on the ability of software to be transferred from one environment to another.

The system was evaluated using a qualitative approach. A questionnaire was prepared which consisted of statements that were adapted from the ISO/IEC 9126. Graduating students for 2015 who are expected to participate in the review sessions were asked to respond to the statements in the questionnaire in terms of the extent to which they agree with them. Specifically, a five-point Likert Scale was used to measure the opinions of the respondents. Table 3 shows the five responses used and their corresponding numerical value which were used to measure the quality of the system under evaluation.

Table 3. Likert Scale

Response	Numerical Value
Strongly Agree	1
Agree	2
Uncertain	3
Disagree	4
Strongly Disagree	5

Data were summarized and analyzed using Mode, Frequency and Percentages to determine the Central Tendency of the data while Inter-Quartile Range was used to measure the dispersion of the data.

RESULTS AND DISCUSSION

The data were gathered and then treated in response to the objectives presented in this study. Four objectives prompted the collection of the data and data analysis. Those objectives were to generate a licensure examination performance prediction model based on Multiple Regression and PART Classification technique, to develop the OTDSS, to test the OTDSS and to evaluate the OTDSS. These objectives were attained. The results presented establish the potential for educational data mining as well as the integration of data mining and decision support system.

The PART Models

The original dataset for PART was divided into two using the 80/20 rule: the training and validation dataset which consisted of 56 instances and the test dataset which consisted of 14 instances. Figure 4 shows the PART decision list generated using WEKA in a 10-fold cross validation and a confidence factor of 0.25 without preprocessing of attributes. This is named as PART Model A in this study. It consists of five rules namely:

If MBResult is G and SelfReview is Y and ProfEd is G, then P.
If MBResult is G and ProfEd is FS, then P.
If ProfEd is S, then P.
If Auditory is F, then P.
Otherwise, F.

```
PART decision list
------------------

MBResult = G AND
SelfREview = Y AND
ProfED = G: P (18.0)

MBResult = G AND
ProfED = FS: P (16.0)

ProfED = S: P (6.0)

Auditory = F: P (9.0/3.0)

: F (7.0/1.0)

Number of Rules  :      5
```

Fig. 4. PART Model A

When the training and validation dataset of PART was subjected to preprocessing using the supervised attribute selection, only four attributes were selected out of 13. Figure 5 shows the selected attributes when CfsSubsetEval using Best First search method was applied.

```
Attribute selection output

=== Run information ===

Evaluator:    weka.attributeSelection.CfsSubsetEval
Search:       weka.attributeSelection.BestFirst -D 1 -N 5
Relation:     PART for training and validation-weka.filters.supervised.attribute.AttributeSelection-
Instances:    56
Attributes:   4
              GenEd
              SelfREview
              MBResult
              LETPerf
Evaluation mode:    10-fold cross-validation



=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

number of folds (%)  attribute
         10(100 %)    1 GenEd
          8( 80 %)    2 SelfREview
         10(100 %)    3 MBResult
```

Fig. 5. Attribute Selection Output for PART

Based from the attribute selection, it can be established that attributes GenEd, Selfreview and MBResult are relevant features as they are good predictors of the class LETPerf (LET performance) (Figure 11). This supports the study of Ong, Palompon and Bañico (2012) that the students' academic performance and their performance in the pre-board examination are significant determinants of the success and failure of their licensure examination performance [12].

From the preprocessing of attributes, only two rules were generated using the same validation and confidence factor as shown in Figure 6. This is named as PART Model B in this study. The rules are:
If MBResult is G, then P.
Otherwise, F.

```
PART decision list
------------------

MBResult = G: P (47.0/3.0)

: F (9.0/3.0)

Number of Rules  :      2
```

Fig. 6. PART Model B

Table 4 shows the confusion matrices of the two PART models generated with and without preprocessing of attributes. It compares the accuracy of the models as to the number of correctly and incorrectly classified instances. Correctly classified instances in PART Model A had higher percentage than that of the PART Model B. The correctly classified instances in PART Model A had an overall percentage of 89.29, while PART Model B had an overall percentage of 87.50.

Table 4. PART Confusion Matrices

	Predicted Class					
Actual Class	**PART Model A**			**PART Model B**		
	Passed	**Failed**	**Percent Correct**	**Passed**	**Failed**	**Percent Correct**
Passed	46	1	97.87	44	3	93.62
Failed	5	4	44.44	4	5	55.56
Overall %	90.20	80.00	89.29	91.67	62.50	87.50

Figure 7 examines the difference between the performance of the two PART models as far as classifier accuracy is concerned when subjected to testing using the supplied test dataset. It could be gleaned from the figure that PART Model B had a higher percentage on the correctly classified instances and lesser percentage on the incorrectly classified instances than that of the PART Model A. However, there is only a little difference of 7.14% on their accuracy.

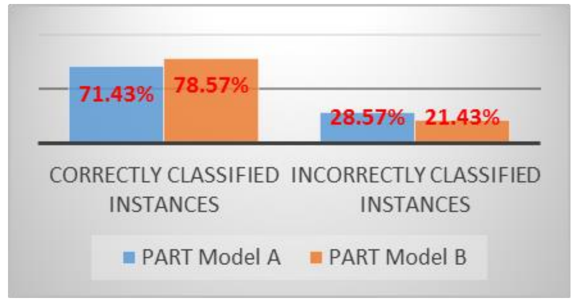

Fig. 7. Classifier Accuracy Performance of the PART Models

Results show that the PART models generated, whether or not there is attribute selection, has a high percentage of accuracy (Table 4). PART Model A has a higher percentage accuracy of 89.29% while PART Model B has 87.50%. PART Model B gave a very slim range of only two rules (Figure 6) but of higher accuracy when subjected to testing (Figure 7). These two rules were already incorporated in PART Model A (Figure 5). PART Model A which summarizes all the rules is better during the training and validation while PART Model B is better during testing. An important implication of this finding is that imposing a random limit on the number of attributes that can be considered when building a model is insignificant if there are only few attributes and instances.

The Multiple Regression Models

The original dataset for MR was also divided into two using the 80/20 rule: the training and validation dataset which consisted of 46 instances and the test dataset which consisted of 11 instances.

```
Linear Regression Model

LETPerf =

      5.3095 * GenEd +
     -5.4424 * ProfED +
     -8.9076 * MajorCore +
     -1.1974 * PeerStudy +
      1.4045 * AskQuest +
      0.4584 * GiveIdeas +
     -1.139  * Visual +
     -1.4949 * Auditory +
     -1.2066 * Kinesthetic +
      0.0855 * MBResult +
     97.0025
```

Fig. 8. MR Model A

Figure 8 shows the MR equation generated using WEKA in a 10-fold cross validation and a confidence factor of 0.25 without preprocessing of attributes. This is named as MR Model A in this study.

When the training and validation dataset of MR was subjected to preprocessing using the supervised attribute selection, only two attributes were selected out of 13. Figure 15 shows the selected attributes when CfsSubsetEval using Best First search method was applied.

```
Attribute selection output

=== Run information ===

Evaluator:    weka.attributeSelection.CfsSubsetEval
Search:       weka.attributeSelection.BestFirst -D 1 -N 5
Relation:     MR for training and validation-weka.filters.supervised.attribute.AttributeSelection
Instances:    46
Attributes:   2
              MajorCore
              LETPerf
Evaluation mode:    10-fold cross-validation



=== Attribute selection 10 fold cross-validation seed: 1 ===

number of folds (%)  attribute
          10(100 %)  1 MajorCore
```

Fig. 9. Attribute Selection Output for MR

From the preprocessing of attributes making use of the same validation and confidence factor, the MR model generated is shown in Figure 10. This is named as MR Model B in this study. MR Model B tells us that the LET performance is expected to decrease by 9.4925 when the general weighted average in Major or Content subjects increases by one (Figure 10). The finding was quite confusing because of the negative computed coefficient of the MajorCore attribute which usually means a decrease effect. But this is inappropriate on this study since it uses a grading system where the lower the grade the higher it is.

```
Linear Regression Model

LETPerf =

     -9.4925 * MajorCore +
     96.8363
```

Fig. 10. MR Model B

Table 5 shows the error measures of the two MR models generated with and without preprocessing of attributes. It compares the fitting of the models as to the differences between the observed values and the model's predicted values. The correlation coefficient of MR Model B is higher than that of the correlation coefficient of MR Model A. On the other hand, MAE and RMSE of MR Model B is lesser than that of MR Model A's. This implies that MR Model B is better than MR Model A since it has higher correlation coefficient and lower MAE and RMSE. The correlation coefficient of MR Model A is 0.3531 which implies a weak positive linear relationship between various predictors while MR Model B is 0.4899 which implies a moderate positive linear relationship between various predictors. Based from the attribute selection, it can be established that attribute MajorCore, is the only relevant predictor of the class LETPerf (LET performance) (Figure 9). Nevertheless, there remains a clear need for further research on this matter.

Table 5. Error Measures of the MR Models

	MR Model A	**MR Model B**
Correlation Coefficient	0.3531	0.4899
Mean Absolute Error (MAE)	3.1906	2.9525
Root Mean Squared Error (RMSE)	4.0494	3.4841

Figure 11 shows the evaluation of MR Model A when subjected to testing using the supplied test dataset.

```
Classifier output

=== Predictions on test set ===

 inst#     actual  predicted      error
     1     78.5       75.608     -2.892
     2     81.8       75.97      -5.83
     3     80.81      74.58      -6.23
     4     75         74.606     -0.394
     5     83.2       81.893     -1.307
     6     76.4       78.345      1.945
     7     72         78.254      6.254
     8     76         75.674     -0.326
     9     73         77.733      4.733
    10     82         78.91      -3.09
    11     77.4       81.032      3.632
```

Fig. 11. MR Model A Evaluation on Test Set

```
Classifier output

=== Predictions on test set ===

 inst#     actual  predicted      error
     1     78.5       76.921     -1.579
     2     81.8       77.092     -4.708
     3     80.81      77.718     -3.092
     4     75         76.427      1.427
     5     83.2       78.269     -4.931
     6     76.4       76.779      0.379
     7     72         78.8        6.8
     8     76         76.304      0.304
     9     73         78.089      5.089
    10     82         79.75      -2.25
    11     77.4       80.224      2.824
```

Fig. 12. MR Model B Evaluation on Test Set

Figure 12 also shows the evaluation of MR Model B when subjected to testing using the same supplied test dataset.

Figure 13 examines the difference between the performances of the two MR models, as far as evaluation on test set is concerned. It could be gleaned from the figure that MR Model A had a higher MAE and RMSE than that of the MR Model B. Lower values of MAE and RSME indicate better fit.

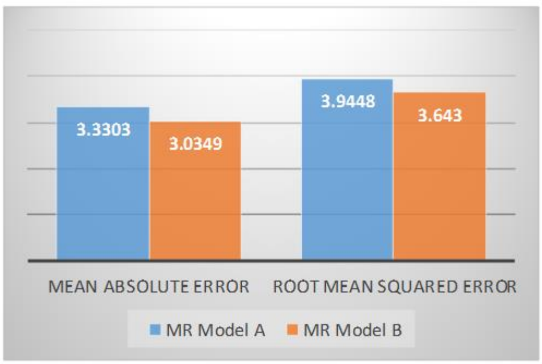

Fig. 13. Performance of the MR Models based from Evaluation on Test Set

Testing and Evaluation of the Developed OTDSS

Testing and evaluation were undertaken as soon as the OTDSS was finalized. For the testing, both sample and real institutional data were entered into the system.

When a sample dataset was entered in Microsoft Excel, Figure 14 shows the coefficients of the predictors from the regression output.

	A	B	C
1		*Coefficients*	
2	Intercept	91.58132564	
3	GenEd	11.54954642	
4	ProfED	-6.163809775	
5	MajorCore	-15.53089353	
6	MBResult	0.080670307	
7			

Fig. 13. Regression Output in Microsoft Excel

The same dataset was entered into WEKA for further accuracy testing of the MR Model. Figure 14 shows the MR Model generated by WEKA.

```
Classifier output

Linear Regression Model

LETPerf =

     11.5495 * GenEd +
     -6.1638 * ProfED +
    -15.5309 * MajorCore +
      0.0807 * MBResult +
     91.5813

Time taken to build model: 0.01 seconds
```

Fig. 14. WEKA Regression Model Output

The same dataset was uploaded into the developed OTDSS to generate the MR Model as shown in Figure 15. The simulation results in the OTDSS match the calculations of Microsoft Excel and WEKA.

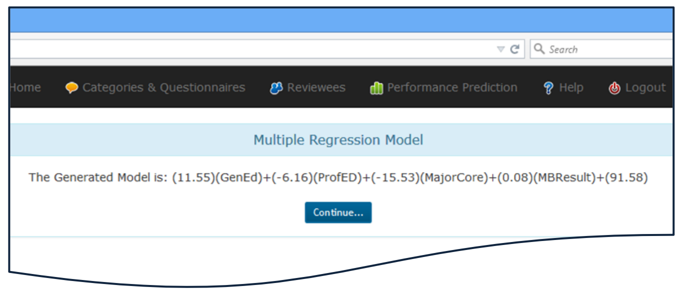

Fig. 15. ***Multiple Regression Output of OTDSS***

Based from the results of testing, the MR model generated by the OTDSS was accurate since the coefficients of the predictors calculated using Microsoft Excel, model generated by WEKA and that of the output of OTDSS were all the same.

On the other hand, the evaluation of the OTDSS went as expected with no unusual events. Figure 16 shows the home page of the OTDSS that was presented to the respondents.

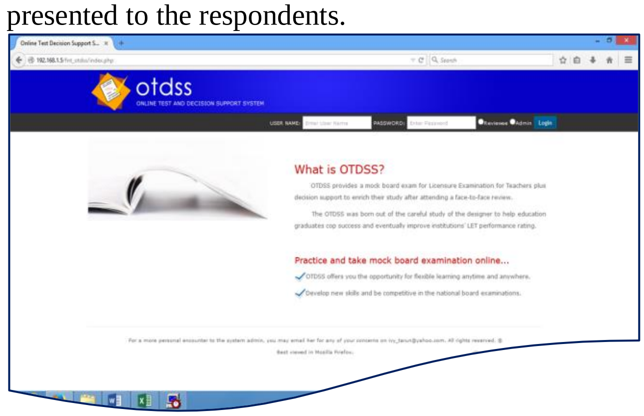

Fig. 16. OTDSS Home Page

Table 6. OTDSS Functionality

	Mode	%	Inter-Quartile Range
The set of functions on a specified task is appropriate.	1	88.89	0
The system provides the right and agreed results.	1	74.07	0.5
The system has the ability to prevent unauthorized access, whether accidental or deliberate, to programs or data.	1	81.48	0
The system works as per intended application.	1	96.30	0
The system provides output that conforms to the base requirements.	1	85.19	0
The system generates output the same with that of the expected output.	1	92.59	0
The system has no broken links and spelling mistakes.	1	92.59	0
There is no confusing application flow and crashes.	1	88.89	0
The response time is fast.	1	88.89	0
The system serves its purpose well.	1	96.30	0

The functionality of OTDSS as measured by the mode as well as frequency and percentage with respect to the mode, and Inter-Quartile Range appears in Table 6. Data revealed that the responses on the system's functionality were clustered together as evidenced by the small values of IQR. However, there is a very little variation on the responses when it comes to the system's provision of right and agreed results. Most respondents indicated strong agreement with the idea that the OTDSS satisfied its implied functions as far as mode is concerned (Mode=1).

Table 7. OTDSS Efficiency

	Mode	%	Inter-Quartile Range
User specifications are achieved by the system.	1	85.19	0
The system produces desired output with optimum time.	1	88.89	0
The system does the required processing on least amount of hardware.	1	88.89	0
The system uses an optimum amount of memory and disk space.	1	85.19	0
The system performs greater useful work transactions.	1	88.89	0

The efficiency of OTDSS as measured by the mode as well as frequency and percentage with respect to the mode, and Inter-Quartile Range appears in Table 7. Data revealed that the responses on the system's efficiency were also clustered together as evidenced by the zero value of IQR. Most respondents indicated strong agreement as to the efficiency of the OTDSS as far as mode is concerned (Mode=1).

Table 8. OTDSS Usability

	Mode	%	Inter-Quartile Range
The system is easy to use.	1	88.89	0
The system enables the user to learn how to use it.	1	96.30	0
The system has an attractive interface.	1	66.67	1
The system is easy to understand.	1	96.30	0
The system is fit to be used by both reviewers and reviewees.	1	96.30	0

The usability of OTDSS as measured by the mode as well as frequency and percentage with respect to the mode, and Inter-Quartile Range appears in Table 8. Data revealed that the responses on the system's usability were clustered together as evidenced by the small values of IQR. However, there is a very little variation on the responses when it comes to the system's attractive interface. Most respondents indicated strong agreement with the idea that the OTDSS satisfied its usability characteristics as far as mode is concerned (Mode=1).

Table 9. OTDSS Reliability

	Mode	%	Inter-Quartile Range
The system handles errors systematically.	1	85.19	0
Transactions are simple.	1	85.19	0
The system has the ability to continue operating properly in the event of the failure of some of its components.	1	85.19	0
The system provides consistent results.	1	92.59	0
The system performs consistently well.	1	88.89	0

The reliability of OTDSS as measured by the mode as well as frequency and percentage with respect to the mode, and Inter-Quartile Range appears in Table 9. Data revealed that the responses on the system's reliability were clustered together as evidenced by the small values of IQR. Most respondents indicated strong agreement regarding the reliability of the OTDSS as far as mode is concerned (Mode=1).

Table 10. OTDSS Portability

	Mode	%	Inter-Quartile Range
The system is re-useable.	1	96.30	0
The system is easy to install.	1	81.48	0
The system can be transferred from one computer to another.	1	85.19	0
The system can run on different operating systems without requiring major rework.	1	81.48	0
The system can be easily integrated into another environment with consistent functional correctness.	1	88.89	0

The portability of OTDSS as measured by the mode as well as frequency and percentage with respect to the mode, and Inter-Quartile Range appears in Table 10. Data revealed that the responses on the system's portability were likewise clustered together as evidenced by the small values of IQR. Most respondents indicated strong agreement regarding the reliability of the OTDSS as far as mode is concerned (Mode=1).

CONCLUSION AND RECOMMENDATION

From the study that has been carried out, it is concluded that the GWA of the reviewees in their General Education subjects, the result of the Mock Board Exam and the instance when the reviewee is conducting a self-review are good predictors of the licensure examination performance. It is suggested however that further research should still be undertaken considering a greater number of instances using the same attributes or other relevant variables.

It has been demonstrated in this paper using the developed Online Test and Decision Support System that the fusion of data mining in education and decision support system can be readily used in practice. The approach of integrating data mining in decision support system can be successfully used for other applications aside from performance prediction.

The developed Decision Support System produced accurate Multiple Regression models. This was exhibited during testing and simulation of real institutional data and displayed the same output with that of the two reliable application programs, the Microsoft Excel and WEKA. Furthermore, the developed Decision Support System satisfied its implied functions and is efficient, usable, reliable and portable. The respondents during the evaluation of the system have shown their enthusiasm of using the system. This suggests that the system could be used by the reviewees to enhance their licensure examination review. This would allow initial identification of reviewees who are likely to have difficulty in passing the licensure examination, therefore providing sufficient time and opportunities for appropriate interventions during review sessions. It should, however, be used with caution as the system was built as an encouragement for the reviewees to review very well and not as a substitute to the face-to-face review sessions.

The developed Decision Support System does not include other statistical computations and other data mining classifiers. As such, it is recommended that future researchers may upgrade the developed system by incorporating more complex features both on the data mining and decision support stages like statistical computations of standard error, t- statistic and p-value together with the generation of the MR Model. Other data mining classifiers may also be included such as decision tree algorithms in support with the MR model.

REFERENCES

[1] Fan, W. and Bifet, A. (2012). Mining Big Data: Current Status, and Forecast to the Future. SIGKDD Explorations, 14(2), 1-5. Retrieved from https://goo.gl/qyUf2O

[2] Sembiring, S., Zarlis, M., Hartama, D., Ramliana, S. and Wani, E. (2011). Prediction of Student Academic Performance by an Application of Data Mining Techniques. International Conference on Management and Artificial Intelligence, 6, 110-114. Retrieved from https://goo.gl/EHrdaa

[3] Baker, R. and Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1(1). Retrieved from https://goo.gl/R1oRJM .

[4] Tarun, I. M., Gerardo, B. D. and Tanguilig III, B. T. (2014). Generating Licensure Examination Performance Models Using PART and JRip Classifiers: A Data Mining Application in Education.

International Journal of Computer and Communication Engineering, 3(3), 202-207.
[5] Multiple Linear Regression in Data Mining (n.d.) Retrieved from https://goo.gl/jhpz5g
[6] Pentaho Data Mining (n.d.) PART. Retrieved from http://wiki.pentaho.com/display/DATAMINING/PART
[7] Markov, Z. and Russell, I. (n.d.). An Introduction to the WEKA Data Mining System. Retrieved from www.cs.ccsu.edu/~marko Pedregosa et al., v/weka-tutorial.pdf
[8] Pedregosa et al. (2011). Cross-validation: evaluating estimator performance. Retrieved from https://goo.gl/5DYPkJ
[9] Hall, M. A. and Holmes, G. (2003). Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. IEEE Transactions on Knowledge and Data Engineering, (15(3). Retrieved from https://goo.gl/BCWrbE
[10] Hamad, M. M. and Qader, B. A. (2014). Knowledge-Driven Decision Support System Based on Knowledge Warehouse and Data Mining for Market Management. International Journal of Application or Innovation of Engineering and Management, 3(1), 139-147.
[11] ISO 9126: The Standard of Reference. Retrieved from http://www.cse.dcu.ie/essiscope/sm2/9126ref.html
[12] Ong, M. B.., Palompon, D. R. and Bañico, L. (2012). Predictors of Nurses' Licensure Examination Performance of Graduates in Cebu Normal University, Philippines. LDCU-RPO| Asian Scientific Journals, 2(1). Retrieved from https://goo.gl/6yjVKj