Predictive Decision Support System for Licensure Examination Performance through Integration of Multiple Regression and PART Classification Models of Data Mining

Ivy M. Tarun, Bobby D. Gerardo, and Bartolome T. Tanguilig III

***Abstract*—Data mining is the process of discovering knowledge which in turn can be used to predict future trends. On the other hand, decision support system is an information system that enables one to analyze data and compile information that will aid in decision making process. This paper presents the integration of data mining and decision support system in an educational context. thus, a predictive decision support system for licensure examination performance (PDSS-LEP) is designed that highlights the repeated generation of multiple regression model and the integration of another classification model which was derived using PART classification technique. The PDSS-LEP was found beneficial as it provides a good platform for generation of MR model that can be adapted by other institutions because of its model selection procedures and user-oriented interface. It is however suggested that data integration should be enhanced by considering multiple sources of data.**

***Index Terms*—Data mining and decision support system integration, LET performance, multiple regression, predictive decision support system.**

I. Introduction

Decision support system (DSS) is a computerized information system that is intended to help organizations identify and solve problems and eventually make better decisions [1]. DSS is widely used in business and management, medical diagnosis, agriculture, forest management [2] and the like. However, there has been an insignificant application of DSS in educational domain which involves data that are not business-oriented such as academic records, licensure examination performance and so on.

Licensure examination performance (LEP) is a growing concern of most of the educational organizations. LEP of the graduates is one of the major determinants of the quality of education an institution offers. This manifests the institution's high standard of instruction and quality of students they have. Hence, administrators are constantly monitoring the LEP in order to improve school achievements. Thus, one area of interest is the graduates of the institution who are about to take the licensure examination – the reviewees. A support must be extended to them prior to the taking of the licensure examination. Questions like "What subjects do I need to concentrate more? What possible score can I get? How likely will I pass the exam?" should be addressed. These questions, when addressed, would help determine who among the reviewees need further review assistance and what possible interventions are made by the reviewers. Eventually, this would help improve the institutions' LEP. A more innovative method of resolving these queries is through the fusion of data mining and decision support system. The integration of data mining and decision support may improve the quality of problem-solving methods, processes and achieve results [3].

Data mining is the process of detecting relevant patterns in a database using defined approaches and algorithms to look into current and historical data that can then be analyzed to predict future trends [4]. This is carried out simply by establishing a model from instances where the answers are known and then applied to other instances where the answers are not known [5]. If this process is adopted and integrated into a decision support system for data analysis, then information could be effectively revealed for easier decision making.

This paper introduces a decision support system prototype called predictive decision support system for licensure examination performance (PDSS-LEP) with the integration of multiple regression (MR) and PART classification models of data mining. multiple regression is a data mining technique that allow the use of more than one input variable and allow for the fitting of more complex models [6] while PART is a classifier that generates decision list [7].

The PDSS-LEP facilitates the repeated generation of MR model and the use of PART classification model. The PART model used in the system was previously derived from another study that has an accuracy of 82.54% [7].

II. Related Works

To demonstrate how this study stands in relation to the current state of data mining and decision support system, here are some of the studies that are relevant to this work.

Khademolqorani and Hamadani introduced an approach on the integration of DSS with data mining and multiple criteria decision making (MCDM). The proposed DSS includes the following steps: data understanding and select target, data integration, data selection, data preparation, data inspection, tools selection, presentation format, model implementation, evaluate finding, homogenize results, and

Manuscript received May 8, 2014; revised July 7, 2014.

I. M. Tarun is with the Isabela State University, Isabela, Philippines (e-mail: ivy_tarun@yahoo.com).

B. D. Gerardo is with West Visayas State University, Iloilo, Philippines. (e-mail: bgerardo@wvsu.edu.ph).

B. D. Tanguilig III is with the College of Information Technology Education, Technological Institute of the Philippines, Quezon City, Philippines (e-mail: bttanguilig_3@yahoo.com).

presenting results [8]. This approach is similar to the process model on this study only that MCDM was not considered since decision making are left to the users.

Zupan *et al*., also tried to bridge the data mining and decision support in their study. Specifically, they proposed a mechanism for communicating data mining models. The predictive model was developed separately within a data mining tool which is called Orange, then the model is encoded in XML [9] hence, models are hard coded and fixed.

Chau and Phung introduced the same concept by proposing a knowledge-driven educational decision support system for education with a semester credit system by taking advantage of educational data mining [10]. This is however different from this paper's objective though it was considered to be in the educational context.

Gasar, Bohanec and Rajkovic combined DM and decision support approach to the prediction of academic achievement. The models were developed by a combination of data mining using Weka and decision support techniques relying mainly on expert knowledge [11]. This combination of model development did not actually improve the accuracy of the model.

Hamad and Qader also merged the concepts of data warehouse (DW) and knowledge warehouse (KW) in their proposed Knowledge-driven DSS for market management. Their knowledge-driven DSS consisted of the following phases: collecting data, data pre-processing, loading the data into the DW, data selection for knowledge discovery, knowledge discovery using association rule mining, interpreting the rules to discover knowledge as output, representing the result using one of the visualization tools and decision making [12].

On the other hand, AL-Malaise proposed a DSS model with the integration of data mining abstract. The proposed model covers all the DSS components with some integration of DM abstracts and customer enquiry system [13].

All the above-cited works gave credit to the importance of integrating decision support system and data mining. However, the studies gave attention only to business and business-related processes that deliver customer-centered and marketplace support. No studies, so far, were undertaken that focus on student records in a DSS that will lead to the prediction of licensure examination performance of reviewees.

III. Work Done/ Contributions

The PDSS-LEP is a model-driven decision support system that is predictive in nature since it makes use of the multiple regression model and PART classification model of data mining for decision support. The design of PDSS-LEP was conceived primarily to make performance predictions through data mining and present these to users minimizing the complexity that arises when using data mining tools. The data mining tools use parameters and generate results that are hard to understand. Hence, the PDSS-LEP was designed to have a user-friendly interface for reviewees who are considered beginners to take advantage of the data mining capabilities presented in a decision support system. Likewise, the design was aimed at helping both the institutions and reviewees in achieving good results as far as licensure examination performance is concerned.

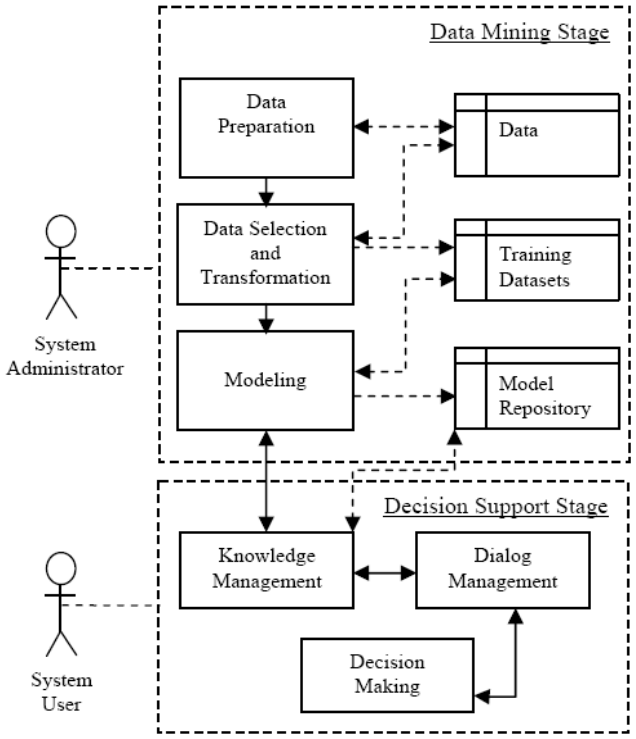

Fig. 1. Process model of PDSS-LEP.

A. Process Model of PDSS-LEP

The process model of PDSS-LEP was based on the two comprehensive models, the *CRoss Industry Standard Process for Data* Mining (CRISP-DM) and the basic model of DSS. The CRISP-DM is broken down into six phases: business understanding, data understanding, data preparation, modeling, evaluation and deployment [14]. On the other hand, the DSS model consists of five main phases: data management, model management, knowledge management, dialog management and decision making [15]. In this study, we have applied the same concept with some important changes and integration of approaches and tasks. Hence, the PDSS-LEP process model consisted of two stages – the Data Mining stage and the Decision Support stage as illustrated in Fig. 1.

The data mining stage presents the generation of the model in preparation for the decision support stage. The data mining stage begins with data preparation and ends with the modeling process. This stage is particularly handled by the system administrator knowledgeable in data mining.

First, the data are cleaned and integrated into a single data warehouse. Data cleaning is the process of removing the duplicate instances and those that have empty values. Before modeling, a training dataset must be prepared. Hence, data are selected from the data warehouse and transformed into meaningful groups within the attributes to match the requirements of multiple regression algorithm. Once the training dataset is ready, multiple regression algorithm can then be performed to generate the MR model. The broken lines that surround the data mining stage imply a cycle. Thus,

there can be repeated generation of MR model. The MR models are then stored in the model repository.

The decision support stage begins with knowledge management. This is the process where a model is selected and applied to the data. The knowledge is then presented to the user in a manner that is easy to understand. Dialog management is the process where a system user actually interacts with the system and inputs whatever data is needed to aid in the knowledge management. Results of the dialog will then facilitate in the decision making by the system user. The system user is responsible for the model selection to enable him to analyze the data as well as the knowledge which would allow him to make better decisions.

B. Functionalities of PDSS-LEP

The PDSS-LEP basically provides three major functionalities: the repeated generation of the MR model, the Mock Board Exam for system users to take and the Licensure Examination Performance Prediction for the system users to view and eventually make decisions as to what particular category of subjects needs special attention to. The PDSS-LEP prototype is shown in Fig. 2.

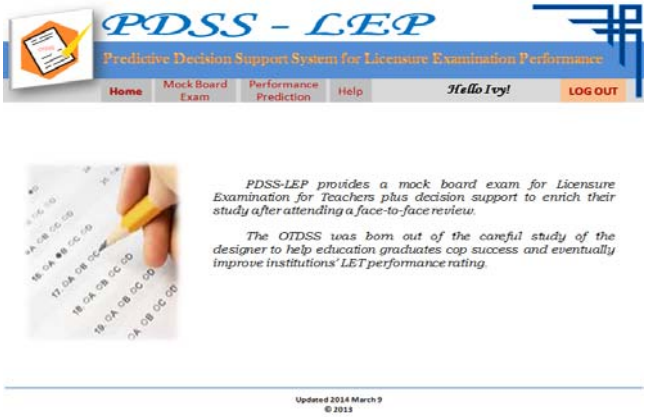

Fig. 2. PDSS-LEP home page.

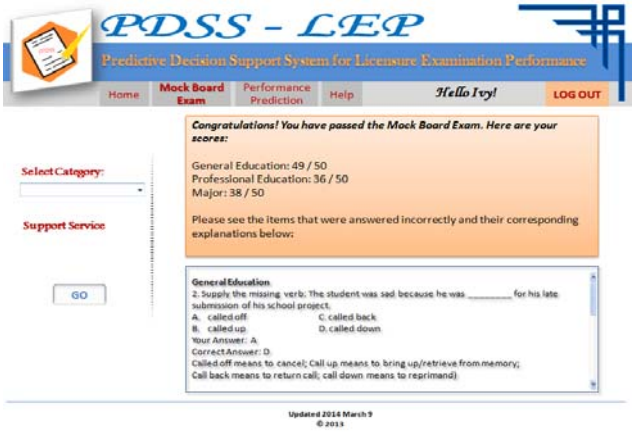

Fig. 3. Support service interface.

The PDSS-LEP goes beyond giving a test to the system users. Besides the mock board exam that will let the system users, specifically the reviewees, take the test within the time allotted, PDSS-LEP provides them the scores on the three different categories that they have taken: the General Education, Professional Education and Specialization for BSEd and Core for BEEd. Aside from the scores, the PDSS-LEP also provides additional support service. The support service of PDSS-LEP presents the items that were answered incorrectly by the reviewee. It gives support by explaining the concept therein. Fig. 3 illustrates this functionality.

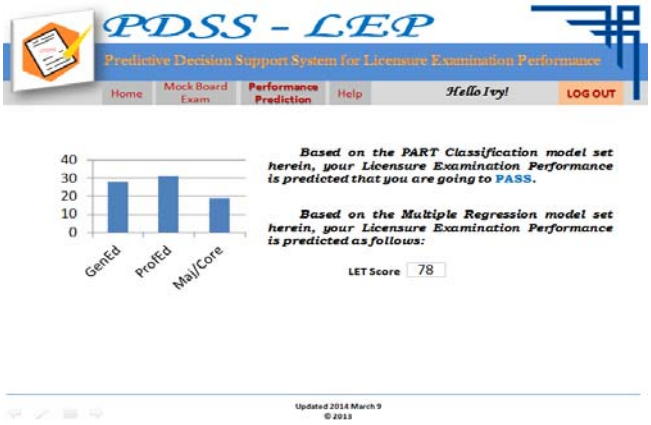

Fig. 4. Performance prediction interface.

Fig. 5. Dataset selection interface.

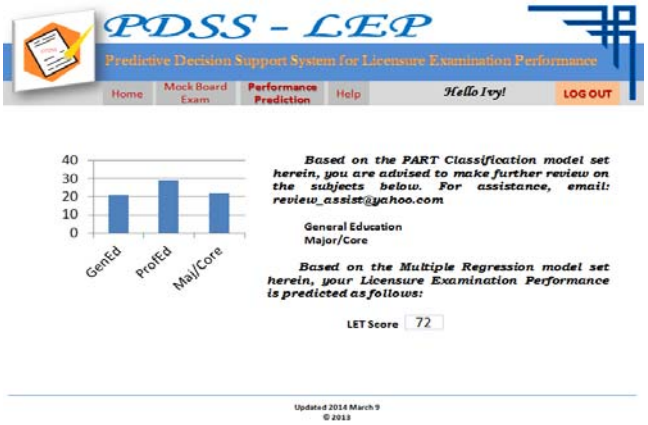

Fig. 6. Performance prediction interface.

PDSS-LEP also presents the licensure exam performance prediction by providing the predicted score of the reviewee after having entered the necessary input as shown in Fig. 4. The predicted score is based on the multiple regression model set by the user from the model repository. Only the system administrator has the authority to generate MR models while the system user is granted an opportunity to select a model for the prediction of his licensure exam performance.

The MR model can be repeatedly generated by the system

administrator using different training datasets which is selected from the data warehouse. This is illustrated in Fig. 5. The selected dataset will serve as input to the generation of the MR model.

To further aid in the prediction and support service of the system, PART prediction model is also used. This model which is uniquely integrated to the system consists of if-then rules previously generated from another study. An example of this functionality is shown in Fig. 6.

All models that have been selected and their corresponding predictions from the data entered by the users are stored in the database. This database can be accessed by the system administrator as shown in Fig. 7 for the purpose of continued data analysis, model evaluation and model improvement.

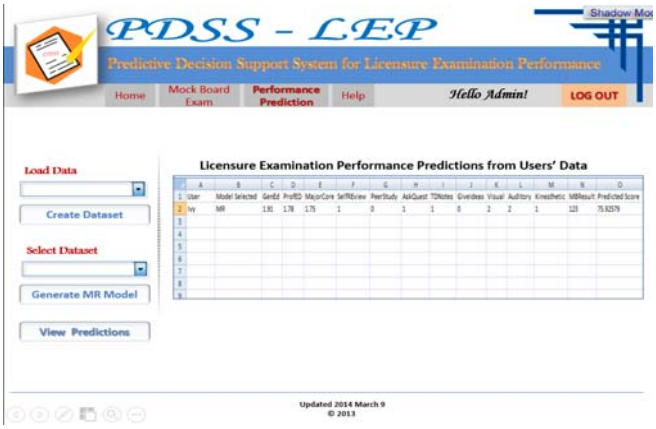

Fig. 7. Predictions from users' data.

IV. Discussion

Discussion on this work concentrated on the suggestions and the structured observations that were undertaken from the data mining experts and target users. However, the small number of respondents upon which the suggestions and structured observations are established signify a limitation of the analysis of data.

The simplicity of the design of PDSS-LEP led to the reviewees' approval for system use. Majority of them stated their appreciation on the system's user interface as it was designed with ease of use. As regards the dialog management, interactions between the system and the user are straightforward but not harsh especially when the performance prediction is negative to the part of the user.

The management of data and models are present as this is one of the basic capabilities that a DSS should possess. This capability was enriched since PDSS-LEP supports model integration through the repeated generation of MR model and the use of another model which is PART. Model integration is the process whereby individually developed sub-models are logically combined to create a large unified model [16]. However, it was suggested that data integration should also be improved along with model integration by allowing multiple sources of data.

V. Conclusions and Future Works

By the integration of data mining and decision support system, a system is designed for the prediction of licensure examination performance called PDSS-LEP. It is a model-driven decision support system that is predictive in nature since it makes use of the multiple regression model and PART classification model of data mining for decision support. PDSS-LEP is beneficial as it provides a good platform for generation of MR model that can be adapted by other institutions because of its model generation and selection procedures and user-oriented interface. Further, the design and concepts presented herein can serve as a benchmark for the optimal use of data mining and decision support in an educational context.

Data and model integration in the decision support system remains a challenge for future works. Multiple sources of data for data preparation and integration are our utmost concern in order to support reasonable decision making. Subsequently, implementation and evaluation of the PDSS-LEP is yet another undertaking that could be achieved in the future.

Acknowledgment

The authors would like to thank the selected Education Graduates of Batch 2014 of the Isabela State University Cabagan campus for their valuable comments and suggestions. The authors are also grateful to the selected IT experts for their inputs on data mining and decision support system.

References

[1] C. Boylan. Decision Support Systems – DSS (definition). [Online]. Available: http://www.informationbuilders.com/decision-support-systems-dss
[2] (February 2014). Decision Support System. [Online]. Available: http://en.wikipedia.org/wiki/Decision_support_system
[3] M. Bohanecl, B. Zupan. (January 2013). Integrating Decision Support and Data Mining by Hierarchical Multi-Attribute Decision Models. [Online]. Available: http://www.researchgate.net/publication/228753916_Integrating_decision_support_and_data_mining_by_hierarchical_multi-attribute_decision_models
[4] J. Silltow. (August 2006). Data Mining 101: Tools and Techniques. [Online]. Available: http://www.theiia.org/intAuditor/itaudit/archives/2006/august/data-mining-101-tools-and-techniques/
[5] D. Alexander. (July 1997). Data Mining. [Online]. Available: http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/
[6] M. Chapple. Regression. [Online]. Available: http://databases.about.com/od/datamining/g/regression.htm
[7] I. M. Tarun, B. D. Gerardo, and B. T. Tanguilig III, "Generating licensure examination performance models using PART and JRip classifiers: A data mining application in education," *International Journal of Computer and Communication Engineering*, vol. 3, no. 3, pp. 202-207, 2014.
[8] S. Khademolqorani and A. Z. Hamadani, "An Adjusted Decision Support System through Data Mining and Multiple Criteria Decision Making," in *Proc. Procedia - Social and Behavioral Science*s, 2013, pp. 388–395.
[9] B. Zupan *et al*. Orange and Decisions- at-Hand: Bridging Predictive Data Mining and Decision Support. [Online]. Available: http://www-ai.ijs.si/MarkoBohanec/pub/iddmorng.pdf
[10] V. T. N. Chau and N. H. Phung, "A Knowledge-driven educational decision support system," *Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)*, vol. 6, p. 1, 2012.
[11] S. Gasar, M. Bohanec, and V. Rajkovic, "Combined Data Mining and Decision Support Approach to the Prediction of Academic Achievement. [Online]. Available: http://silvana.telesat.si/dokumenti/IDDM11b2.pdf
[12] M. M. Hamad and B. A. Qader, "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*, vol. 3, no. 1, pp. 139-147, January 2014.
[13] A. S. A. Malaise. Integration of Automated Decision Support Systems with Data Mining Abstract: A Client Perspective. International Journal of Advanced Computer Science and Applications. [Online]. Available: http://thesai.org/Downloads/Volume4No2/Paper_26-Integration_of_Automated_Decision_Support_Systems_with_Data_Mining_Abstract_A_Client_Perspective.pdf
[14] R. Wirth and J. Hipp. CRISP-DM: Towards a Standard Process Model for Data Mining. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.5133&rep=rep1&type=pdf
[15] S. Srinivasan1, J. Singh, and V. Kumar. Multi-agent based decision Support System using Data Mining and Case Based Reasoning. [Online]. Available: http://www.ijcsi.org/papers/IJCSI-8-4-2-340-349.pdf
[16] S. Liu, A. H. B. Duffy, R. I. Whitfield, and I. M. Boyle, "Integration of decision support systems to improve decision support performance," *International Journal of Knowledge and Information System*, vol. 22, no. 3, pp. 261-286, 2010

Ivy M. Tarun was born in Cabagan, Isabela, Philippines, 1978. She is a graduate of bachelor of science in computer engineering last 2000 at the University of Saint Louis, Tuguegaro City. She earned her master's degree in information technology on 2004 at the University of La Salette, Santiago City. She is currently taking her doctorate studies in the field of information technology at the Technological Institute of the Philippines, Quezon City. She became a member of IAENG and IACSIT last February 2013.

She is a faculty member of the Isabela State University, Isabela Philippines. She has co-authored the book entitled "ICT @ First Glance: A Coursebook in Basic Software Applications" which was published last 2012. She has embarked last 2010 a research study entitled "Attitude of the Education Students toward Electronic Portfolio Assessment as a Substitute to Traditional Tests" which was awarded as Best Paper during the Agency In-house Review in the institution where she was working. The paper was eventually published in the institution's Research Journal last 2011. She is currently interested on researches involving the educational application of data mining, system development and web services.

Bobby D. Gerardo finished his BS in electrical engineering in 1994, with High Distinction from Western Institute of Technology at Iloilo, Philippines.

He took his master of arts in education major in mathematics from University of the Philippines in Diliman Quezon City in 2000 being the grantee of DOST-SEI scholarship for Math and Science Faculty. He pursued his Ph.D. in information and telecommunications with major in distributed systems at Kunsan National University, Korea in 2007 being the grantee of Korean Scholarship for Brain Korea (BK-21) project.

He is currently the vice president for administration and finance of West Visayas State University, Iloilo City, Philippines. His dissertation: Discovering Driving Patterns using Rule-based intelligent Data Mining Agent (RiDAMA) in Distributed Insurance Telematic Systems. He has published 54 research papers in national and international journals and conferences. He is a referee to international conferences and journal publications such as in IEEE Transactions on Pattern Analysis and Machine Intelligence and IEEE Transactions on Knowledge and Data Engineering. He is interested in the following research fields: distributed systems, telematics systems, CORBA, data mining, web services, ubiquitos computing and mobile communications.

Dr. Gerardo is a recipient of CHED Republica Award in Natural Science Category (ICT field) in 2010. His paper entitled SMS-based Automatic Billing System of Household Power Consumption based on Active Experts Messaging was awarded Best Paper on December 2011in Jeju, Korea. Another Best Paper award for his paper Intelligent Decision Support using Rule-based Agent for Distributed Telematics Systems" Asia Pacific International Conference on Information Science and Technology on December 18, 2008. An Excellent Paper award was given for his paper "Principal Component Analysis Mechanism for Association Rule Mining", Korean Society of Internet Information's (KSII) 2004 Autumn Conference on November 5, 2004. He was given a University Researcher Award by West Visayas State University in 2005.

Bartolome T. Tanguilig III was born in Baguio City, Philippines, in 1970. He took his bachelor of science in computer engineering in pamantasanng lungsodng maynila, Philippines in 1991. He finished his masters degree in computer science from De la Salle University, Manila, Philippines in 1999, and his doctor of Philosophy in technology management was earned at the Technological University of the Philippines, Manila in 2003.

He is currently the assistant vice president for Academic Affairs and Concurrent the dean of the College of Information Technology Education and Graduate Programs of the Technological Institute of the Philippines, Quezon City.

Dr. Tanguilig is a member of COMMISSION on Higher Education Technical Panel for IT Education, Board Chairman of Junior Philippine IT Researchers, member of Computing Society of the Philippines and Philippine Society of IT Educators.