Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

ISSN NO. 2320-5407

Article DOI:10.21474/IJAR01/8753
DOI URL: http://dx.doi.org/10.21474/IJAR01/8753

RESEARCH ARTICLE

THE USE OF DATA MINING TO MODEL PERSONALIZED LEARNING MANAGEMENT SYSTEM.

Mia Villar Villarica.
Laguna State Polytechnic University.

Manuscript Info

Manuscript History
Received: 20 January 2019
Final Accepted: 22 February 2019
Published: March 2019

Key words:-
Educational Data Mining, PLMS, Data Mining Algorithm

Abstract

Licensure Examination performance is a growing concern of most of the educational institution because it is one of the determinants of quality education and validates high quality instruction. Educational institution is focused on monitoring and improving the Licensure Examination performance particularly in Teacher Education Institution (TEI). The study intends to offer a possible solution to most TEIs apprehensions regarding LET performance by providing the students of Teacher Education a student support service in the form of a personalized Learning Management System with performance prediction and recommendation capability. This can be developed through drawing data model using several data mining techniques and tools. Previous literature suggested using data mining to classify students, predict student performance, improve student retention, enhanced student achievement and assess complex students' behavior to name a few.

This research project will provide the groundwork for the generation of a prediction model an innovative Student Support System with an Integration of Information Technology that would help the students by providing a Learning Management System (LMS) with personalized learning environment, performance prediction and recommendation engine that can be beneficial to board programs. This particular study covers the identification of the appropriate data mining algorithms to be used in modeling the PLMS.

After an intensive investigation and literature review conducted, the research the result shows that ID3 and J48 is the best data mining algorithms is student performance prediction. The researcher will be using these algorithms in the development of the PLMS to integrate the prediction function of the prototype.

Introduction:-

Teacher education is the most popular program offered in the country and yet most schools in the Philippines that offer the program are not performing well in the Licensure Examination for teachers (LET). According to the latest statistics from the Commission on Higher Education (CHED), for every 100 enrollees in teacher education programs in the Philippines, only 16 will eventually graduate, on average. A recent study by non-government organization Philippine Business for Education (PBEd) revealed that, for the last 5 years, most teacher education institutions

Corresponding Author:-Mia Villar Villarica.
Address:-Laguna State Polytechnic University Santa Cruz Campus

(TEIs) have not reached the national test-takers passing rates. According to PBEd the national passing rate for the elementary and secondary exams are 52% and 56% respectively. However, out of 1,025 TEIs for elementary, 601 or 59% did not reach the national passing rate. The same is factual data for secondary 63% (795 out of 1,258). The study identified 129 worse and worst performing TEIs in elementary. These are TEIs which had less than 20% of their students pass the LET. Also based on their study 17 TEIs had no LET passers for both exams from 2009-2013. The data for the study came from the Professional Regulation Commission [10] [11].

These alarming facts has placed Higher Educational Institution (HEI) specifically TEIs in a very complex state. TEIs challenged by different apprehensions such as closure of programs, revision curriculum, screening procedures to name a few.

The study aims to offer a feasible resolution to these apprehensions through providing the students of Teacher Education a student support service in the form of a personalized Learning Management System with performance prediction and recommendation capability. This can be developed through drawing data model using several data mining techniques and tools. Educational Data Mining (EDM) is defined as an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, knowledge transfer, learning situations and the organization which they learn in [14]. EDM will be applied in the analyzation, design and development of an automated system. The main research problem of this study is to generate a prediction model and discover factors that may impact the outcome of the licensure examination performance for the Bachelor of Secondary Education graduates who are planning to take the Licensure Examination for Teachers. More specifically to address the following research questions: (1) What are the predictors significant to the prediction of the LET performance of the students? (2) How will the data mining techniques give best results to predict the LET Examinee performance and produce data models for prediction of LET results? (3) How can the data models be used in the design and development of the recommendation engine in the Personalized Learning Environment?

The Personalized Learning Management System (PLMS) will be coined as ExAid short for Examinee Aid for the Teacher Education students specifically the Bachelor of Secondary Education graduates who are planning to take the Licensure Examination for Teachers.

The main objective of this study is to create a framework to be used in the design and development of ExAid: A Personalized Learning Management System with Performance Prediction and Recommendation Capabilities. Specifically it aims: (1) to identify the predictors significant to the prediction of the LET performance of the students (2) to utilize variety of data mining techniques to discover knowledge from the pool of data gathered (3) to utilize the data model created from data mining techniques applied in the development of ExAid Prototype.

Literature Review

This section discusses the previous study and literature specifically education data mining, student performance predictors and algorithms used. This particular part of the research will be valuable because this will help the researcher the appropriate data mining algorithms to be used.

The student performance is of great concern in the educational institutes where several factors may affect the performance. To be able to predict the student performance three required components are: (1) parameters which affect the student performance, (2) data mining methods and (3) data mining tool. These parameters may be psychological, personal, and environmental. The study aims to maintain the education quality of institute by minimizing the diverse effect of these factors on student's performance. In his paper, Prediction of student Performance is done by applying Naïve bayes and J48 decision tree classification techniques WEKA tool. By applying data mining techniques on student data we can obtain knowledge which describes the student performance. This knowledge will help to improve the education quality, student's performance and to decrease failure rate. All these will help to improve the quality of institution [6].

Also, data Mining techniques was used for the Prediction of Student Performance based the literature review presented by [6]. The data mining algorithm Naïve Bayes and J48 are identified to be the most accurate data mining techniques for the prediction of student performance [17]. However, the algorithms CHAID and CART were applied on student enrolment data of information system students of open polytechnic of New Zealand to get two decision trees classifying successful and unsuccessful students. The accuracy obtained with CHAID and CART was 59.4 and

60.5 respectively [16]. Obtaining the university students data like attendance, class test, seminar and assignment marks from the students' database, to predict the performance at the end of the semester using three algorithms ID3, C4.5 and CART and shows that CART is the best algorithm for classification of data. The CHAID prediction model was useful to analyze the interrelation between variables that are used to predict the outcome on the performance at higher secondary school education. Finally, classification techniques such as Naïve Bayes, J48, Decision Tree and CHAID analysis will be useful in the conduct of the data mining stage of this study as previous study suggests [9].

The application of decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the NaïveBayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models [2]. However, obtained the university students data like attendance, class test, seminar and assignment marks from the students' database, to predict the performance at the end of the semester using three algorithms ID3, C4.5 and CART and shows that CART is the best algorithm for classification of data [16].
In addition to this, a study was conducted on the student performance by selecting 200 students from BCA course. By means of ID3, C45 and Bagging they find that SSG, HSG, Focc, Fqual and FAIn were highly correlated with the student academic performance [1].

Decision tree model predicted the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. They used 12 predictive variables and a 4-class response variable for the model construction. Three different classification methods namely ID3, C4.5, and the NaïveBayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models with the predictive accuracy of 38.33% for four-class response variable [2]. In a case study on educational data mining to identify up to what extent the enrolment data can be used to predict student's success. The algorithms CHAID and CART were applied on student enrolment data of information system students of open polytechnic of New Zealand to get two decision trees classifying successful and unsuccessful students. The accuracy obtained with CHAID and CART was 59.4 and 60.5 respectively [16]. Also in another study which obtained the university students data like attendance, class test, seminar and assignment marks from the students' database, to predict the performance at the end of the semester using three algorithms ID3, C4.5 and CART and shows that CART is the best algorithm for classification of data [17].

The CHAID prediction model was useful to analyze the interrelation between variables that are used to predict the outcome on the performance at higher secondary school education. Finally, classification techniques such as Naïve Bayes, J48, Decision Tree and CHAID analysis will be useful in the conduct of the data mining stage of this study as previous study suggests [9]. However, the Figure 1 shows the sample of Data Mining Techniques used for Prediction of Student Performance based the literature review [4]. Based on this table Naïve Bayes and J48 is the most accurate data mining techniques for the prediction of student [18].

Table 1: Sample of data mining techniques used for prediction of student performance

Author	year	DM technique	Accuracy
Saurabh Pal et. al	2011	Naïve bayes	Not assigned
Anwar M. A	2012	Apriori algorithm	Not assigned
Vaibhav P. Vasani et. al	2014	Naïve Bayes	86.4%
		J48	95.9%
Sembiring et al.	2011	Naïve bayes	82.4%
		K-Means	93.7%
		DecisionTree	80.2%
Edin Osmanbegovet.al	2012	Naïve Bayes	76.48%
		Multilayer Perception	71.2%
		J48	73.98%

Fig 1: Sample of data mining techniques used for prediction of student performance Baradwaj (2011)

As mentioned in the above literature, educational data mining has been recognized and utilized by different educations institutions in support with development of curriculum, student retention and attrition, learning preference, tracing students' academic performance and educational institution strategic management decision making. In addition to this, educational data mining has been recognized and significant in predicting student performance and learning outcome. Also on the previous literature mentioned there are several Data Mining techniques utilized to produce prediction model and give best results.

From these specific studies, the researcher found out that the student performance could depend on diversified factors such as demographic, academic, psychological, socio-economic and other environmental factors. It was learnt that the variation in the predictive accuracy could be correlated with the nature of student data set and utilization of number of records, predictive variables and class values of response variable. In addition, several data mining techniques were identified to be utilized for this study. It is found out that classification data mining techniques are widely used in educational data mining specifically in student performance prediction. The use of k-means clustering algorithm to predict student's learning activities simple linear regression analysis it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance [3] [16].

Several data mining algorithms were stated to be useful and reliable in predicting student performance. Classification methods namely ID3, C4.5, and the NaïveBayes were utilized in different studies. One of the study indicated that outcome of their results indicated that Decision Tree model had better prediction than other models. A case study on educational data mining to identify up to what extent the enrolment data can be used to predict student's success [19]. Also, the algorithms CHAID and CART were applied on student enrolment data of information system students to get two decision trees classifying successful and unsuccessful students. A study obtained the university students data like attendance, class test, seminar and assignment marks from the students' database, to predict the performance at the end of the semester using three algorithms ID3, C4.5 and CART and shows that CART is the best algorithm for classification of data. The CHAID prediction model was useful to analyze the interrelation between variables that are used to predict the outcome on the performance at higher secondary school education [17] [9].

Finally, classification techniques such as Naïve Bayes, J48, Decision Tree and CHAID analysis will be useful in the conduct of the data mining stage of this study as previous study suggests. Decision Tree inducers ID3 and J48 was utilized in different studies specifically in student performance prediction and has given best results.

The literature review has produced best results in finding out the data mining algorithms. Also, a conceptual framework was developed using several theoretical frameworks reviewed in previous related studies conducted. The framework consisted of two stages the Data Mining Stage and Decision Support Stage. At the Data Mining Stage CRISP-DM was used having six phases business understanding, data understanding, data preparation, modeling, evaluation and deployment while at the Decision Support Stage which consists of five main phases data management, model management, knowledge management and decision making.

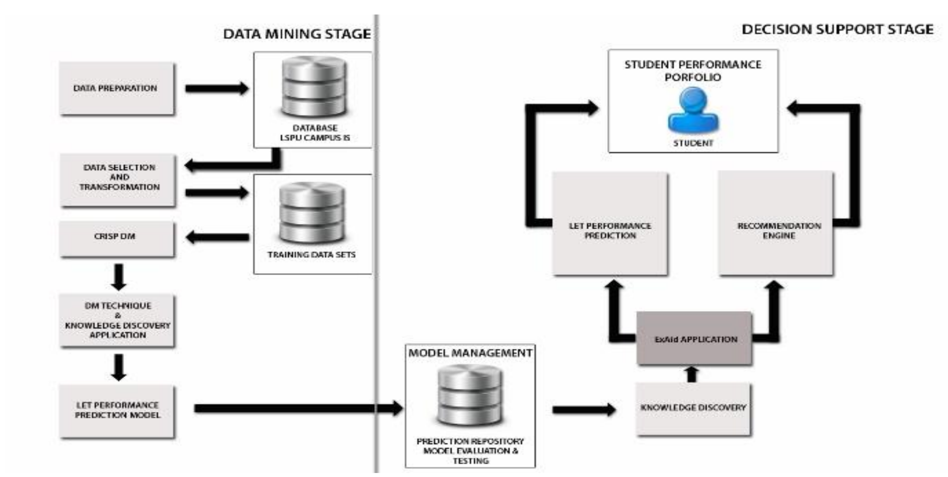

Fig 2:-Conceptual Framework

Materials And Method:-

Methods and Techniques Used

The study will make use of variety of data mining algorithm specifically classification based on the literature review decision tree, J48, CHAID, PART and k-means are data mining algorithms that specifically gives best results to determine student performance, prediction outcome, recommendation and personalization.

Instrument of the Study

This study will use the acquired basic data and information coming from an Automated Campus Information System - Student Academic Record – Registrar, LET Results and survey results. Academic Records of Bachelor of Secondary Education Graduates of 2013 – 2015.

Development Methodology

The researcher developed a research model to guide with the conduct of the research, significant phases are illustrated in Fig. 2. The figure shows three phases of the study, the first phase of the research, predictors identified based on the literature review and exhaustive research. Identification of predictors and attributes is vital to the knowledge extraction specifically academic records such as General Education, Professional Education and Major Core Course grades will be captured and process. Data on the following attributes such as self-review, peer study, ask questions, notes during session, give ideas, Pre LET Results will be gathered thru survey [8]. Using different data mining techniques and knowledge discovery application, data models will be generated using these predictors correlating historical academic records of students and LET Results. This phase is vital in the design and development of the Personalized Learning Management System (PLMS) to handle major functionalities. The input variable consists of predictors and attributes needed to examine and process. The researcher will use CRISP–DM procedures consisting of business understanding, data understanding, data preparation, modeling, evaluation and deployment. The output of the system will consist of data models for the design and development of ExAid: A Personalized Learning Management System with Performance Prediction and Recommendation Engine Capabilities.
The second phase of the research, is the design and development of Personalized Learning Management System (PLMS), input variable such as knowledge requirements – the data model generated from the database of the LSPU Campus Information System and survey results. Also experts and advisers were consulted for the execution of the design and development of the system. Functional and Nonfunctional requirements are identified based on fact finding techniques conducted and IT experts consultation, considering the following functionalities: (**a**) LET Performance Prediction based on Academic Records (General Education, Professional Education and Major Core Course) (**b**) pre LET Examination (Questions based on the CTE prescribe Pre LET Reviewer) with automatic feedback (**c**) LET Performance Prediction based on pre LET Examination Results and Academic Records. (**d**) recommended area to be focused to or area of deficiency (**e**) provide a learning environment specifically for the user based on the recommendation engine (**f**) formulate a shortest learning path for the personalized learning environment. Last phase of the research, is the integration of a Recommendation Engine injected on the Personalized Learning Environment and Shortest Learning Path Capabilities.

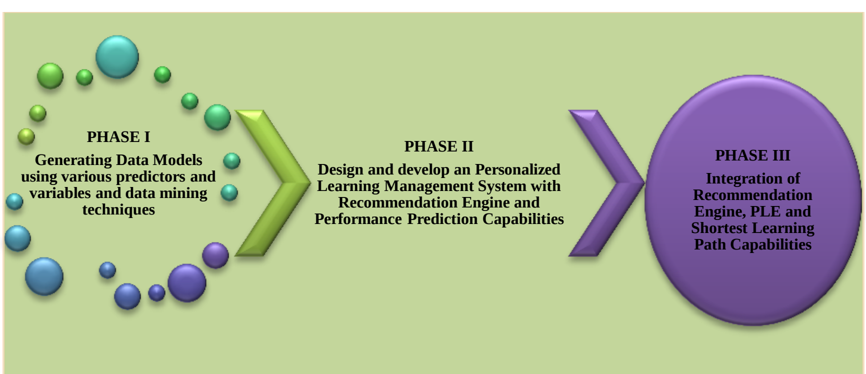

Fig 2:-Research Model

Data Model Generation

The CRISP-DM methodology will be utilized in this study [12]. It remains the most popular methodology for analytics, data mining, and data science projects [5]. CRISP-DM: Cross-Industry Standard Process for Data Mining is a standardized approach to data mining and process model that describes commonly used approaches that data mining experts use to resolve problems and challenges during the process. The CRISP-DM methodology is the life cycle of a data mining project consists of six phases, shown in Fig. 2. The sequence of the phases is not rigid. Moving back and forth between different phases is always required. The outcome of each phase determines which phase, or particular task of a phase, has to be performed next. The arrows indicate the most important and frequent dependencies between phases. The outer circle in the figure symbolizes the cyclical nature of data mining itself. Data mining does not end once a solution is deployed. The lessons learned during the process and from the deployed solution can trigger new, often more-focused research questions. The subsequent data mining processes will benefit from the experiences of previous ones.

The initial phase of CRISP-DM is business understanding which focuses on understanding the project objectives and requirements from a business perspective. Organization and university process assessment was conducted specifically college admission process. Literature consultation and relevant research was carried out to translate the information into knowledge and converting this knowledge into a data mining problem definition, and formulate a preliminary plan designed to achieve the objectives. In this study, research related to data mining for educational environment were sought and trace the appropriate data mining techniques the previous study utilized to be adopted on this research.

The main objective of this study is to develop a framework to be used in the design and development of ExAid Prototype that would help examinee improve licensure examination performance through prediction and recommendation capabilities of the personalized LMS. Specifically it aims to (1) identify the predictors significant to the prediction of the LET performance of the students (2) evaluate the models using cross-validation and misclassification errors to decide which model outperforms other models in term of classification accuracy (3) utilize the data model created from data mining techniques applied in the development of ExAid Prototype (4) evaluate is the design of the ExAid prototype based on the evaluation of administration.
The second phase is data understanding and it starts with an initial data collection. This phase proceeds with familiarization with the data, identifying data quality problems, discovering first insights into the data and detecting interesting subsets to form hypotheses for hidden information.

This phase was performed in parallel with the initial and the subsequent phase. The university database contains student information from admission to graduation. The database is in MySQL has been used for almost 10 years, outsourced and not well maintained. The database possesses flaws and few bugs.

This phase covers all activities to construct the final dataset from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection, as well as transformation and cleaning of data for modeling tools. In this study, data preparation was performed parallel to business and data understanding.

Capturing data from the University Campus Information System from five tables namely classcards, gradesheet, name, persons and student. With classcards table having 92,769 instances, gradesheet table having 705,247 instances, name table having 31,282 instances, person table having 21,541 instances and student table with 14,698 instances.

In accomplishing the preprocessing, data integration was performed combining data from multiple sources to form a coherent data store. Cross Tabulation Query was used to extract data and join operations using SQL Code, linking the five tables to generate the need data. The data quality problem has been identified, SQL code was utilized to parse attributes, and data was further process on the next phase.

Data cleaning will be accomplished to check missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data. It will performed as an iterative two-step process consisting of discrepancy detection and data transformation.

Data transformation routines convert the data into appropriate forms for mining. Codes will be assigned to appropriately describe the data. Table I and II demonstrates the sample data converted to excel file, converted from its original state which is SQL format to Comma Separated Value. Inner join was performed to produce the needed data it will be converted to Attribute-Relation File Format (ARFF) for WEKA Tool. WEKA Tool will be utilized to train and test data as well as SPSS and Rapid Miner, these knowledge discovery applications are best known for EDM. Column of values are transformed into many columns in the new view.

Table I:-The Symbolic Attribute Description

ATTRIBUTE	DESCRIPTION	VALUES
GenEd (Predictor)	This is the General Weighted Average of the student in his general education courses taken from the academic record.	E, VS, S, FS, VG, G, FG, FR, P, F
ProfEd (Predictor)	This is the General Weighted Average of the student in his professional education courses taken from the academic record.	E, VS, S, FS, VG, G, FG, FR, P, F
MajorCore (Predictor)	This is the General Weighted Average of the student in his professional education courses taken from the academic record.	E, VS, S, FS, VG, G, FG, FR, P, F
SelfReview	This tells if the student conducted self-review	Y,N
MBResult	This tells the score in the Pre LET Exam	VG, G, F
LETPerf	This is the LET Performance of the student.	PASSED, FAILED

Data reduction is used to obtain a reduced representation of the data while minimizing the loss of information content. In this study, various data reduction was done while maintaining its integrity and preserving the quality of information.

Datasets tend to expose new issues and challenges while in the process of transformation. With the goal in mind, it is important to choose the right data mining algorithms, techniques and tools which are expected to give best results with data mining.

Table II:-The Numerical Value Of Predictors Values

For GenEd, ProfEd, and MajorCore		
VALUE		**GRADE/NUMERICAL EQUIVALENT**
E	EXCELLENT	1.0
VS	VERY SATISFACTORY	1.25
S	SATISFACTORY	1.5
FS	FAIRLY SATISFACTORY	1.75
VG	VERY GOOD	2.00
G	GOOD	2.25
FG	FAIRLY GOOD	2.5
FR	FAIR	2.75
P	PASSED	3.00
F	FAILED	5.00
For MBResult		
VG	Very Good	100-150
G	Good	50-99
F	Fair	0-49

Descriptive data summarization techniques was used to identify the typical properties of the data and highlight which data values should be treated as noise or outliers and it provides the analytical foundation for data preprocessing. The basic statistical measures for data summarization was utilized which includes mean, weighted mean, median, and mode for measuring the central tendency of data, range, quartiles, and standard deviation for measuring the dispersion of data – to evaluate the correlation of the data.

Predictors and attributes have been identified based on survey and literature review however the researcher will continue to search for the appropriate variables. The tables below presents the variables – predictors and attributes - the Symbolic Attribute Description and numerical value of predictors.

Data gathered will be transformed in this format to be able to process in the knowledge discovery application – WEKA Tool, Rapid Miner and SPSS which ever gives best rules which will be consulted to domain experts.

During the modeling phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. The models are classified in two main categories; predictive and descriptive models. Predictive data mining will be used, which analyzes data in order to construct one or a set of models, and attempts to predict the behavior of new data sets.

Data mining can be divided into two tasks: predictive tasks and descriptive tasks. The ultimate aim of data mining is prediction; therefore, predictive data mining is the most common type of data mining and is the one that has the most application to businesses or life concerns. Predictive data mining has three classification mathematical, distance and logic solution. These stages are elaborated done with predictive stages such as data processing, prediction and deployment which explains a more complete scenario of all the aspects of data mining [5].

Classification is a model finding process that is used for portioning the data into different classes according to some constrains. In other words we can say that classification is process of generalizing the data according to different instances. Several major kinds of classification algorithms including C4.5, k-nearest neighbor classifier, Naive Bayes, SVM, Apriori, and AdaBoost [4]. Also according to him classification algorithms can be implemented on different types of data sets like data of patients, financial data according to performances. On the basis of the performance of these algorithms, these algorithms can also be used to detect the natural disasters like cloud bursting, earthquake, etc. as well as student performance.

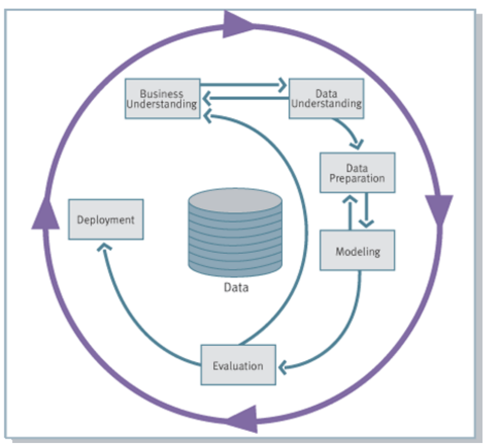

Fig 3:-Steps of CRISP-DM Methodology (Adopted from Chapman et al., 2000)

Data Mining Models

This study compared the statistical predictive algorithms which include decision tree and its inducers, J48, Naïve Bayes, and PART. Each of the result of the data mining model will be evaluated to determine the best data model.

Decision Tree

A decision tree is a classifier expressed as a recursive partition of the instance space. The decision tree consists of nodes that form a rooted tree, meaning it is a directed tree with a node called “root” that has no incoming edges. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test node. All other nodes are called leaves (also known as terminal or decision nodes). In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. In the simplest and most frequent case, each test considers a single attribute, such that the instance space is partitioned according to the attribute’s value. In the case of numeric attributes, the condition refers to a range. Each leaf is assigned to one class representing the most appropriate target value. Alternatively, the leaf may hold a probability vector indicating the probability of the target attribute having a certain value. Instances are classified by navigating them from the root of the tree down to a leaf, according to the outcome of the tests along the path.

The ID3 algorithm is considered as a very simple decision tree algorithm [15]. ID3 uses information gain as splitting criteria. The growing stops when all instances belong to a single value of target feature or when best information gain is not greater than zero. ID3 does not apply any pruning procedures nor does it handle numeric attributes or missing values.

The J48 is one of the decision tree induction algorithm. It is an open source Java implementation of the C4.5 algorithm in the WEKA data mining tool. This algorithm was developed by Ross Quinlan. C4.5 algorithm creates a decision tree which can be used for classification based the value which are presented on dataset. The following steps are used while the decision tree is constructed on J48 classification algorithm; (1) In general the tree is constructed in a top-down recursive divide-and-conquer manner, at start, all the training examples are at the root, attributes are categorical (if continuous-valued, they are discretized in advance), examples are partitioned recursively based on selected attributes test attributes are selected on the basis of a heuristic or statistical measure (e.g., information gain) (2) Conditions for stopping partitioning are as follows; all samples for a given node belong to the same class, there are no remaining attributes for further partitioning – majority voting is employed for classifying the leaf, there are no samples left.

Results And Discussion:-

The intensive literature review has given the research directions and elucidations. It was found out that most related research uses decision tree specifically J48 and ID3. The 2 algorithms was found out to give best results in correlating the predictors. The Methodologies in the development of the system as well as the execution of the conduct of research has cleared the research path. The formulated conceptual framework has also provided a blueprint in the conduct of the research.

Conclusion And Recommendation:-

This study aims utilize data mining techniques and algorithms in the development of a personalized learning management system with the capability of shortest learning path. Providing the LET takers an avenue to be trained and assure LET examination success.

Although J48 anf ID3 was found out to be useful and signification in developing the model. It would also be helpful to test other algorithms to be able to discover hidden patterns and knowledge.

References:-

1. A. K. Pal and S. Pal, "Analysis and Mining of Educational Data for Predicting the Performance of Students", (IJECCE) International Journal of Electronics Communication and Computer Engineering, Vol. 4, Issue 5, pp. 1560-1565, ISSN: 2278-4209, 2013.
2. AI-Radaideh,Q. A., AI-Shawakfa, E.M., and AI-Najjar, M. I., "Mining Student Data using Decision Trees", International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.
3. Ayesha, S. Mustafa, T. Sattar, A.R., Khan, M.I., "Data mining model for higher education system", Europen Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010.
4. Brijesh Kumar Baradwaj , Saurabh Pal , "Mining Educational Data to Analyze Students‟ Performance" ,(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011
5. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining.* AAAI/MIT Press, 1996.
6. G. Kaur and A. Chhabra, "Improved J48 Classification Algorithm for the Prediction of Diabetes," Int. J. Comput. Appl., vol. 98, no. 22, pp. 13–17, 2014.
7. G. Piatetsky, 'The CRISP-DM process model', 2014. [Online]. Available: (http://www.kdnuggets.com/2014/10/crisp-dm-top-methodology-analytics-data-mining-data-science projects.html) [Accessed 10-Aug-2015].
8. I. Tarun, B. Gerardo, and B. Tanguilig, 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), 203-207, 2014.

9. M. Ramaswami and R. Bhaskaran , “A CHAID Based performance Prediction model in Educational Data Mining”, IJCSI Vol. 7 issue no. 1 January 2010. ISSN 1694-0784.
10. Philippine Business for Education ‘Let Schools Performance Briefer (2009-2013 LET)’, 2014.
11. Philippine Business for Education ‘Manifesto for Teacher Quality’, 2015.
12. P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth : **CRISP-DM 1.0: Step-by-step data mining guide**. SPSS Inc., 2000.
13. Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, “Mining student data using decision trees”, International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.
14. R. Baker and K. Yacef,. “ The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3-17, 2009.
15. R. Quinlan, “Induction of Decision Trees”, Kluwer Academic Publishers-Plenum Publishers. Print ISSN 0885-6125, **Online ISSN** 1573-0565, Kluwer Academic Publishers-Plenum Publishers, Volume 1, Issue 1, pp 81–106 March 1986.
16. Syed Tahir Hijazi and S.M.M. Raza Naqvi, ‘Factors Affecting Students’ Performance: A Case of Private Colleges’. Bangladesh e-Journal of Sociology: Volume 3, Number 1, January 2006).
17. S. K. Yadav, B.K. Bharadwaj and S. Pal, “Mining Educational Data to Predict Student’s Retention: A Comparative Study”, International Journal of Computer Science and Information Security (IJCSIS), Vol. 10, No. 2, 2012.
18. V. Vasani and R. Gawali, “Classification and performance evaluation using data mining algorithms”, International Journal of Innovative Research in Science, Engineering and Technology ,(An ISO 3297: 2007 Certified Organization), Vol. 3, Issue 3, March 2014.
19. Z. J. Kovacic, “Early prediction of student success: Mining student enrollment data”, Proceedings of Informing Science & IT Education Conference, 2010.