International Journal of Computer and Information Technology (ISSN: 2279 – 0764)
Volume 02– Issue 04, July 2013

Framework of Recommendation System for Tertiary Institution

Kuyoro 'Shade O., Awodele Oludele, Okolie Samuel O.
Department of Computer Science
Babcock University
Nigeria

Goga Nicolae
University of Groningen, The Netherlands or
Politenica Bucharest

Abstract— **Understanding the reasoning behind variation in student academic performance in tertiary institutions has been the concern of many researchers for decades. Numerous studies have used traditional statistical methods to identify factors that affect and predict student performance. Machine learning has been successfully applied to so many domains, thus recently, researchers are employing this paradigm for modeling student academic performance and other related problems in higher education. This work focus at addressing the following: proposing optimal algorithm suitable for predicting students academic performance; designing a framework of intelligent recommender system that can predict students' performance as well as recommend necessary actions to be taken to aid the students and identifying background factors that affect students' academic performance in tertiary institution at the end of first year. This research used ten classification models and a multilayer perceptron -an artificial neural network function-generated using Waikato Environment for Knowledge Analysis (WEKA). Each model was built in two different ways: the first was built using the 10-fold cross validation, and the second using holdout method (66% of the data was used as training and the remaining as test). Purposive and selective sampling techniques were used in selecting one thousand five hundred (1,500) enrolment records of students admitted into computer science programme Babcock University between 2001and 2010. Results of the classifiers were compared using accuracy level, confusion matrices and speed of model building benchmarks. The random tree identified as optimal in this work is incorporated into designing a framework of intelligent recommender system. The work shows that identifying the relevant student background factors can be incorporated to design a framework that can serve as valuable tool in predicting student performance as well as recommend the necessary intervention strategies to adopt.**

***Keywords-* decision trees, neural networks, background factors, educational planning activities, machine learning, intelligent recommender system.**

I. INTRODUCTION

Higher education systems all over the world nowadays are challenged by the new information and communication technologies.[1] With the increasing competition among higher education institutes, most are focusing on how to increase student retention rates and number of completions. University performance is one of the means of measuring its quality and reputation [2]; thus higher education institutions are becoming more interested in predicting the paths of students, and identifying which students will require assistance in order to graduate. [3] Higher learning institutions encounter many problems which keep them away from achieving this objective. Some of these problems stem from knowledge gap. Knowledge gap is the lack of significant knowledge at the educational main processes such as counseling, planning, registration, evaluation and marketing. For instance, many learning institutions do not have access to the necessary information to counsel students, thus they are unable to give suitable recommendation to the students.

Also, there is a growing interest and concern in many countries about the problem of school failure and the determination of its main contributing factors. This problem has been referred to as "the one hundred factors problem". [4]. Studies have shown that not all students who enroll as freshmen will complete their studies. [5]. The study of factors that influence the academic performance of students in higher education has a long history. The philosophy behind these studies is to understand what is related to, or predicts, poor academic performance and to use this information to design appropriate interventions [6]. A plethora of factors have been found to predict or influence retention and performance, falling into a number of broad categories: including gender, personality factors, intelligence and aptitude tests, academic achievement, previous college achievements, and demographic data, student background and demographics, prior educational achievement and level, psychosocial factors and approaches to study, and institutional and course factors.[7][8] However, factors found to be predictive in some studies are not always predictive in others [9], due in part to the ways in which different studies are designed. Indeed, even in the same study with the same methodology, results for student cohorts sampled at different universities have differed [10] and in general the results of particular studies cannot be generalised to other environments [7]. Machine learning is considered the most suitable technology in giving additional insight into educational entities such as; student, lecturer, staff, alumni and managerial behavior. It acts as an active automated assistant in helping them make better decisions on their educational activities. So far, there are limited numbers of research that examine the influence of family factors on first year academic performance in tertiary institution. In this work, focus is on how to design a predictive framework to serve as a basis for predicting the performance of students at the end of their first year in tertiary institution based on their family background factors and previous academic achievements using machine learning algorithms.

The study develops a framework of predictive system by identifying the most suitable algorithms for predicting first year student performance. The model uses family background factors and previous academic achievement before admission into tertiary institutions characteristics such as SSCE score, UME score, mother's educational qualification, father's educational qualification, sponsor, family size, student's position in the family, mother's occupation, father's occupation, marital status of parents, and average family income to identify significant factors that impact the decision of persistence.

II. DESIGN OF EXPERIMENT

The dataset for the purpose of this study was obtained from Babcock University, Nigeria' Students Record Systems. Both the institution and the students' records used in this work were selected using purposive and selective sampling techniques.[11] The focus is on students admitted into computer science programme of the institution between 2001 and 2010. Within the period of 2001 and 2010, a total of 2042 students were admitted into computer science programme out of a total of 6758 students admitted into Babcock University. As it is common to freshmen, some of these students transfer to other departments while some change institution, few has no complete record in the students record, thus only 1500 complete records of the enrolled students were involved in this study. According to literature, there are diverse factors that can be considered as family background factors. [12][13][14][15] Background factors have been described as the total of a person's experience, knowledge and education, upbringings, training, family socio-economic status and values. Factors such as family income, education qualifications, occupation, gender, religion, even ethnicity have been identified as background factors. The background information used in this study was extracted from the enrolment forms that are given to students to fill as part of entrance registration requirements. The identified background factors include: mother's educational qualification, father's educational qualification, sponsor, family size, mother's occupation, father's occupation, marital status of parents, and average family income. Also the variables relating to their previous academic achievement of students before entering into tertiary institutions were extracted from these forms; these include SSCE grade in English, Mathematics, Physics, Chemistry, Biology and one other relevant subject, UME score. The students' cumulative grade point average (CGPA) at the end of the first session was extracted from Students Record. The data extraction was carefully done, avoiding incomplete record. After the data has been collected, the data was thoroughly cleaned by smoothing noisy data. All inconsistencies were resolved. The SSCE grade was ranked to generate total SSCE score for each student; also the first year cumulative grade point average was grouped into different classes for easy identification. All other variables were grouped appropriately as shown in Table 1. Data repository that interfaces with WEKA was created for the data collected. The fields selected for the model include gender, average family income, mother's qualification, father's educational qualification, parents' marital status, mother's occupation, father's occupation, family size, ethnicity, religion, education sponsor, age the student entered the university, secondary school certificate examination grade (SSCE grade), University Matriculation examination score (UME), first year cumulative grade point average (CGPA). The selection is based on what has been identified in the literature, opinion of some experts in the field of education and the available information about the students. The CGPA is the response/class variable while other variables are predictors. Each variable is categorized appropriately to accommodate all the available information. Data repository was created to store the dataset.

TABLE I. DATA FORMAT

S/N	Variable Name	Variable format	Variable Type
1.	Gender	Male, Female	Categorical
2.	Average Family Income		Continuous
3.	Mother's educational qualification	No formal education, Primary, SSCE, 1st degree, 2nd degree, PhD	Categorical
4.	Father's educational qualification	No formal education, Primary, SSCE, 1st degree, 2nd degree, PhD	Categorical
5.	Marital status of parents	Married, Divorced, Separated, Widowed	Categorical
6.	Mother's occupation	Unemployed, Government worker, Private, Self employed	Categorical
7.	Father's occupation	Unemployed, Government worker, Private, Self employed	Categorical
8.	Family size		Continuous
9.	Ethnicity	Yoruba, Hausa, Igbo, Others	Categorical
10.	Religion	Christianity, Islam, Traditional, Others	Categorical
11.	Sponsor	Parents, Scholarship, Self, Others	Categorical
12.	SSCE Grade Score	A1-8, B2-7, B3-6,C4-5,C5-4,C6-3,D7-2,D8-1,F9-0	Continuous
13.	UME Score		Continuous
14.	Age on entry		Continuous
15.	Current CGPA	A: 4.5-5.0, B+:4.0-4.49, B: 3.5-3.99, C+: 3.0-3.49, C: 2.5-2.99, D:2.0-2.49, E: 1.0-1.99, F:<1.0	Categorical

In machine learning field, there are three basic phases: pre-processing, processing and post-processing in setting up a model. The processing phase involves the model building as shown in Figure 1. The stage involve environment setup, the tool used, and the evaluation of the model

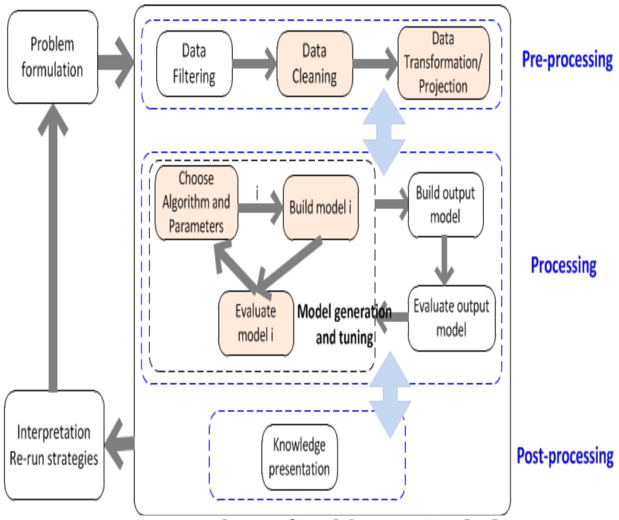

Figure 1: Phases of model generation [16]

Waikato Environment for Knowledge Analysis (WEKA) was used to build software tool for all experiments. WEKA is an open source machine learning package developed in Java; it is a collection of machine learning algorithms for data mining tasks. It contains tools for data preprocessing, classification, regression, clustering, association rules, and visualization. There are many machine learning algorithms implemented in WEKA including Bayesian classifiers, Decision Trees, Rules, Functions, Lazy classifiers and miscellaneous classifiers. It is also well-suited for developing new machine learning schemes. It has a GUI Chooser from which any one of the four major WEKA applications can be selected. For the purpose of this study the Explorer application was used. This study deals with classifier algorithms, functions (multilayer perceptron) and uses the default options presented by Weka for each algorithm. All the machine learning techniques that are used in this study are implemented in WEKA so that they will be easily and fairly compared to each other. The latest Windows version of Weka 3.7.7 has been used.

Though, many learning algorithms are available for use, however, according to the no-free lunch theorem –"There is no universally best classification method, which can achieve superior performance to all the others on any problem"; there is no "winner" method, which achieves superior performance to all the other methods, on all problems. Thus, in the experiment, ten classification algorithms available in WEKA and multilayer percetron, an artificial neural network function were used in this study. The ten classification algorithms comprise five rule induction algorithms –JRip, OneR, ZeroR, PART, and Decision table; and five decision tree algorithms - Random forest, Random tree, J48, Decision stump and REPTree. These classification algorithms have been selected because they are considered as "white box" classification model, that is, they provide explanation for the classification and can be used directly for decision making. Each classifier belongs to a different family of classifiers implemented in WEKA. Random forest, Random tree, J48, Decision stump and REPTree relate to Decision Trees, JRip, OneR, ZeroR, PART, and Decision table belong to Rules, the multilayer perceptron belong to neural networks. Since they are from different classifiers family, they yielded different models that classify differently on some inputs. Attribute importance analysis was carried out to rank the attributes by significance using Information gain and gain ratio attribute evaluators. Ranker's Search method was used to achieve this.

III. EXPERIMENTAL RESULTS

The attributes relating to students' family background factors and previous academic achievement were considered. The attributes used in this study was ranked in order of importance using information gain and gain ratio measures. The outcome is presented in Table 2 and Figure 2. The ranking of both attribute evaluators was done using ranker search method. Among the fourteen attributes used in this study, it was discovered that students JAMB Score, Age on entry, Father's occupation, Mother's occupation are the best five attributes. The outcome of both evaluators is similar as shown in Figure 2.

TABLE II. ATTRIBUTES RANKING USING INFORMATION GAIN AND GAIN RATIO

S/N	Attribute	Information Gain		Gain Ratio	
		Value	Rank	Value	Rank
1	Gender	0.0389	10	0.0453	11
2	Age on entry	0.1689	5	0.0951	5
3	Ethnicity	0.0609	8	0.0478	10
4	Religion	0.0277	14	0.0673	9
5	Family Size	0.1465	6	0.0745	7
6	Sponsor	0.064	7	0.0681	8
7	Father's education	0.0313	12	0.044	12
8	Mother's education	0.0359	11	0.0424	13
9	Father's Occupation	0.4343	2	0.1088	2
10	Mother's Occupation	0.374	3	0.1063	4
11	Parent's marital status	0.0588	9	0.0761	6
12	Monthly Family Income	0.0293	13	0.0397	14
13	Jamb score	0.8013	1	0.1658	1
14	SSCE Score	0.2164	4	0.1072	3

Attribute ranking (with respect to the class attribute) according to information gain and gain ratio criteria show that students JAMB Score, Age on entry, Father's occupation, Mother's occupation are the best attributes. These attributes outperform other attributes in their contribution to the outcome of students' first year performance in tertiary institution as shown in Figure 2.

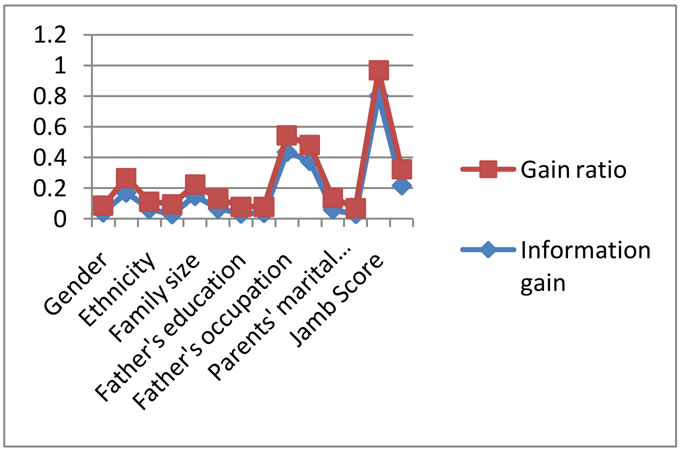

Figure 2: Information gain and gain ratio of the attributes

The outcome of both 10-fold cross validation and hold-out method is similar for all the classifiers. Random tree outperform all other classifiers on both counts. Random forest, Reptree, J48, JRIp, PART, Decision table and multilayer perceptron perform well with the lowest accuracy for both hold-out and 10-fold cross validation being 63.6%. Decision stump, OneR and ZeroR slightly fall behind in accuracy. But overall random tree gives accuracy of 96.07% for 10-fold cross validation and 85.69% for holdout method which outperform all other classifiers used in the course of this study.

TABLE III. CLASSIFICATION ACCURACY ON 10-FOLD CROSS VALIDATION AND HOLD OUT METHODS

S/N	Classifier	10-fold	Hold-out
1	Random forest	95.87	85.29
2	Random tree	96.07	85.69
3	REPTree	74.87	69.22
4	C4.5(J48)	80.6	77.03
5	Decision stump	39.2	34.90
6	JRip	88.87	72.94
7	OneR	45.33	44.90
8	PART	80.4	76.67
9	Decision table	63.6	63.92
10	ZeroR	36.53	34.90
11	Multilayer Perceptron (MLP)	87.13	77.25

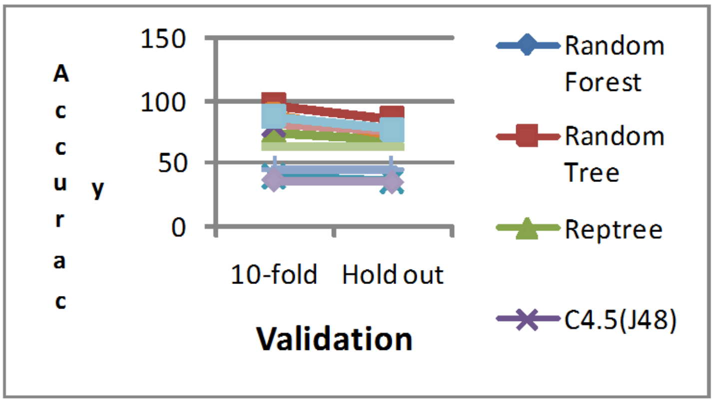

Figure 3: Prediction accuracy for classifiers

The disparity between time taken to build multilayer perceptron and other classification algorithms is very wide as shown in figure 4. Multilayer perceptron consumes much computer resources. Other classifiers took considerable time to execute and consume less system resources. Considering the time taken for building the models in relation to the accuracy level and the performance of the model, it can be established that random tree takes very short time and outperform all other classifiers in this study. Therefore, it can be deduced that random tree according to the outcome of this study is a very good classifier for predicting student first year academic performance in relation to other algorithms used in this study.

TABLE IV. TIME TAKEN (SECONDS) TO BUILD THE ALGORITHMS

S/N	Classifier	10-fold	Hold-out
1	Random forest	0.89	0.22
2	Random tree	0.05	0.01
3	REPTree	0.56	0.08
4	C4.5(J48)	0.27	0.11
5	Decision stump	0.05	0.02
6	JRip	4.42	2.64
7	OneR	0.03	0.03
8	PART	1.09	1.87
9	Decision table	1.25	1.51
10	ZeroR	0.01	0.01
11	Multilayer Perceptron (MLP)	337.08	279.5

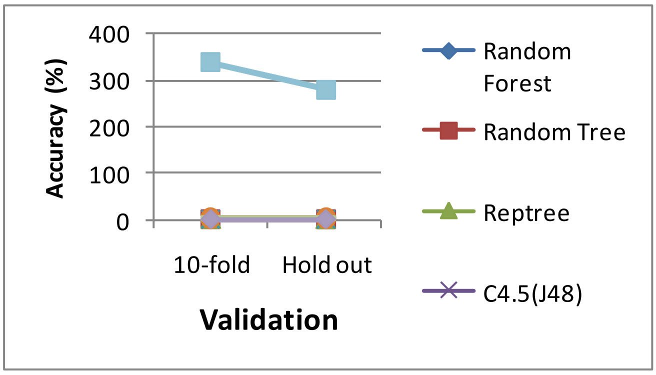

Figure 4: Time taken (seconds) to build the classifier algorithms

Based on all the benchmarks used to measure the algorithms employed in this work, it is discovered that random tree performance is better than all other algorithms. Random forest performance is also considered acceptable since there is a better algorithm with higher performance rate in this domain, we will focus on designing our predictive system on the most suitable algorithm which is random tree in this domain.

IV. OVERVIEW OF EXISITING SYSTEMS

A. *Initial Student Model (ISM).*

Tsiriga [17] developed Initial Student Model (ISM) to set initial values for all aspects of student models using an innovative combination of stereotypes and the distance weighted K-nearest neighbor algorithm. In attempt to solve the challenges associated with Web tutoring, it necessary to develop a personalized environment for each student, thus making initialization of individual student knowledge important. A student is first assigned to a stereotype category concerning his /her knowledge level of the domain being taught. Then, the model of the new student is initialized by applying the distance weighted k-nearest neighbor algorithm among other students that belong to the same stereotype category with the new student. ISM was implemented in a Web-based Intelligent Computer Assisted Language Learning (ICALL) System ISM has been applied in a language learning system used as a test-bed, it was found to be successful at providing sufficiently accurate initial student models, given the fact that very little is known about new students. The ISM exploits the fact that Web-based systems have a large number of users; machine learning reasoning mechanism that is based on recognized similarities between users was used.

B. *Student Relationship Management (SRM).*

Piedade [18] proposed SRM model which supports concepts and techniques to promote an effective student-institution relationship through the closely monitoring of the students and their academic activities. The study found that the technological support to SRM was insufficient at that time. This model was further improved in the latter works of these authors.[19][20][21][22] The SRM system architecture aggregates four components: The Data Acquisition and Storage component, responsible for storing the students' data, gathered from different data sources in a data warehouse. The Data Analysis component, responsible for obtaining knowledge about the students, through appropriates data analysis tools (OLAP, data mining). The Interaction component, responsible for maintaining an effective relationship with the students and the Assessment component, responsible for the carried out actions assess [18]. To validate the SRM system a prototype was implemented and a set of application cases tested. The prototype shows the SRM system feasibility; and the application case shows the SRM system utility to the SRM concept and practice support. The data analysis process allows, among others, the students characteristics identification related with the academic failure and success. One of the behaviors that led to the failure was students' not attending classes. So, it was essential to fights against this trend. One of the actions to develop it was to alert in a regular basis the students' for his/her regular presences in classes through automatic email alert messages. This action integrates the SRM practice, was supported by a Web-SRM application. After that, this system was evaluated by the reaction of students to classes. In mean more students attends the classes comparatively with the same period of the previous curricular year.[22] The SRM concept and the SRM practice implementation, supported by the SRM system, to a large extent create an advantage towards the students' success promotion.

C. *Intelligent Recommendation System for Student Relationship Management (IRSSRM).*

Kanokwan [23] built on [18] to propose an intelligent recommendation system framework for student relationship management which can assess the performance of students and provide appropriate recommendations for their choice of courses and subjects. The Student Relationship Management, similar to customer relationship management in business, aims at developing and maintaining close relationship between institute and the students by supporting the management processes and monitoring the students' academic activities and behavior. The framework is divided into 3 phases: data preparation, data analysis and predictive modeling. The main techniques used are Clustering, Association Rule, Classification using Fuzzy Logic and Rough Set. Further works of these authors in [24][25[26] refine this framework to Hybrid Recommendation System framework to support student relationship management. The component of the framework is increased to four phases: Data Preprocessing, Data Analysis, Intelligent Prediction Models and Model Validation. In the data analysis phase various improvements have been done to include decision tree, neural networks, and support vector machine. Though the work is still on-going, it is planned that the new intelligent Recommendation Models will form an integral part of an online system for a private university in Thailand. After proper evaluation of the developed system, the system will be made available for the use of new students who will access the online-application in their course selection during the enrolment process. The system will also make prediction of the Year 2 and subsequent years' results, which can be used by the counselors, staff and university management to provide supports for students who are likely to need help with their studies. This information will enable the university to better focus on the utilisation of their resources. In particular, this could be used to improve the retention rate by providing additional supports to the group of students who may be at risk.

D. Recommendation system using pattern discovery.

Vialardi [27] proposed a recommendation system based on C4.5 algorithm using pattern discovery module to offer students key element to base their decision on in the enrolment process. The objective of the system is to be able to predict the failure or success of a student in a course using a classifier obtained from the analysis of a set of historical data related to the outcome of other students who took the course in the past. The pattern discovery used offer two types of information: the student can infer what his/her performance will likely be in a particular course and the university have information about type of students and their likely performance in one or more courses which can be a useful input in curriculum modification.

E. Faculty Support System.

Shana [28] proposed a framework called Faculty Support System that enables faculty to analyze their student performance in a course. The framework uses open source analysis software that can dynamically update itself whenever there is change in analysis result. The component of the framework includes both the client and server side. The sever side include student database and analysis model (data selection and transformation, association analysis and prediction, and rule database) while the client side include factor analysis and prediction. The client component is used by all non technical faculties and the analysis component (server side) is controlled by technical experts in knowledge mining where they perform the analysis and load the rules into a rules database which is implemented by the client component. The empirical studies on 182 students taking 'C' programming course identified two data mining techniques that generate rules with considerable accuracy. Supervised association rule mining is used to identify the factors influencing the result of students and C4.5 decision tree algorithm to predict the result of student. The FSS can be integrated into any student management system or operated as a standalone system to enable concerned faculty take effective measures to improve academically weaker students in their courses. This work only concentrates on identifying factors that contribute to success or failure of students in a particular subject and predict result.

The existing systems majorly focus on predicting students' performance in a particular course or subject; considering students' factors from admission to graduation. These systems were developed to aid students decision-making either about their course of study, a particular subject or course registration. According to no-free lunch theorem, the system that was able to predict students' performance in a particular course may not do so for another course. It is thus necessary to design a system that can predict the overall performance of students at the end of each academic session.

V. THE PROPOSED SYSTEM -INTELLIGENT RECOMMENDATION SYSTEM (IRS)

The proposed system aims to aid the higher education management in identifying students' likely performance from enrolment to enhance how the students will be treated in order to achieve optimal academic performance. Unlike the existing systems which are students oriented, this system is basically administrative-oriented. The IRS framework revised the work of [25][26], the Hybrid Recommendation System framework for student relationship management. The authors focus on course registration and proposed hybrid system to consider students' performance at different academic level. This work unlike [25][26], considered models comparison and chose the optimal algorithm. The proposed IRS aggregates four components: The Data Acquisition and Storage component responsible for storing the students' data, gathered from different data sources in a data warehouse. The Model building component, responsible for obtaining knowledge about the students, through appropriates classification models (decision trees, neural network, and rule induction). The Intelligent Recommender System component, responsible for mapping the pattern in the rules generated with the new student data to predict likely performance and the Recommendation component, responsible for recommending necessary actions to be carried out on individual student based on the prediction from the Intelligent recommender system. The proposed framework is presented in Figure 5.

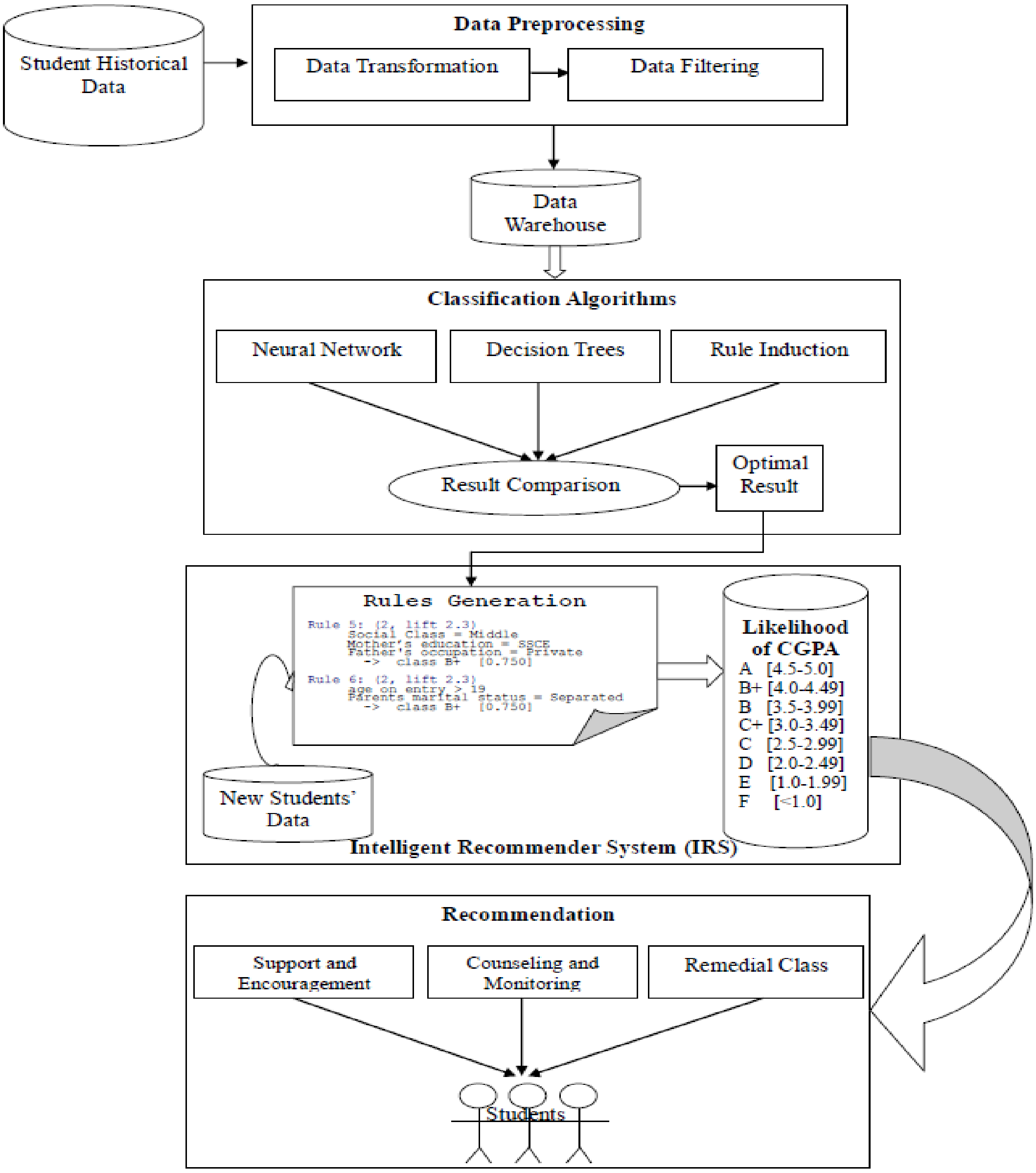

Figure 5: Framework of Intelligent Recommender System

VI. Implementation of the Intelligent Recommender System

The goal of recommender system is to provide an environment that allows timely assessment to support decision-making. This intelligent recommender system is in two-fold. It predicts the outcome of a particular student and recommends the necessary action to be taken based on the prediction. Administrator inserts the likely students' background factors and the system displays both the predicted grade and the recommendation for such student. This system is limited in that it cannot exhaust all the available rules to match all the real world situation, so a back-end for updating the rules as the situation arises has been incorporated into the system. The stated objectives of this work ends with the designing of intelligent recommender system framework, but it is expedient to implement a part of the system to see how effective the system will be if fully implemented. To fully implement this system, an in-depth understanding of Web Ontology Language (OWL) and a larger dataset that will cover all tertiary education institutions will be needed to fully cover the wide range of students' background to generate rules that can exhaustively represent real world situations. There will also be need to compare more than ten classification algorithms. Thus, this system considering all these limitations has been implemented using Netbeans IDE environment for application development, MySQL for database and Java programming language. Different interfaces of students background factors and the associated predictions and recommendations are depicted in Figures 6 to 11.

Figure 6: Intelligent Recommender system Welcome page

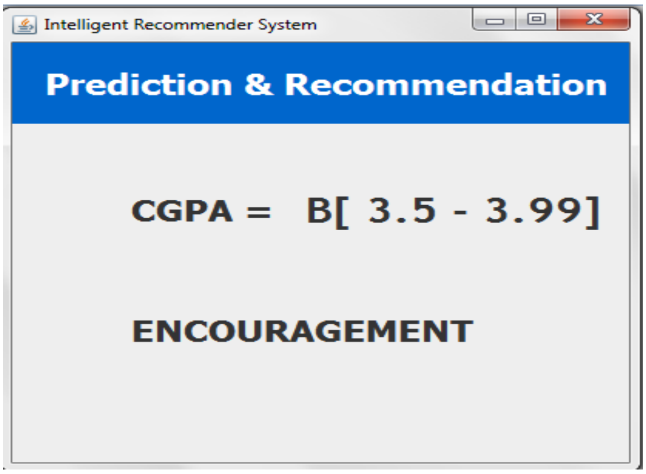

Figure 9: Prediction and Recommendation page

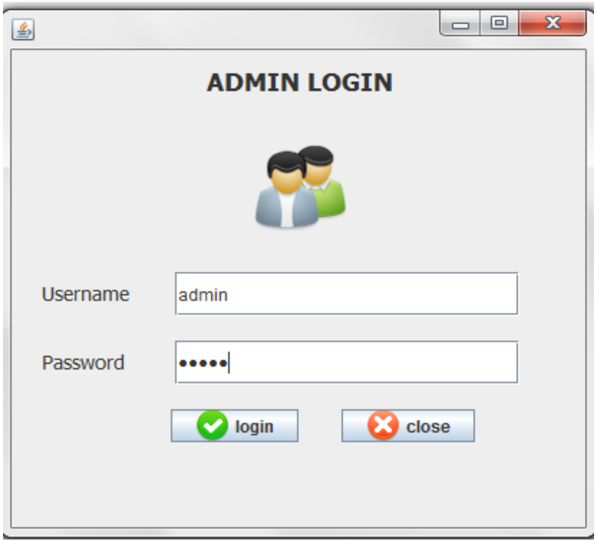

Figure 7: Admin log in page for rules updating

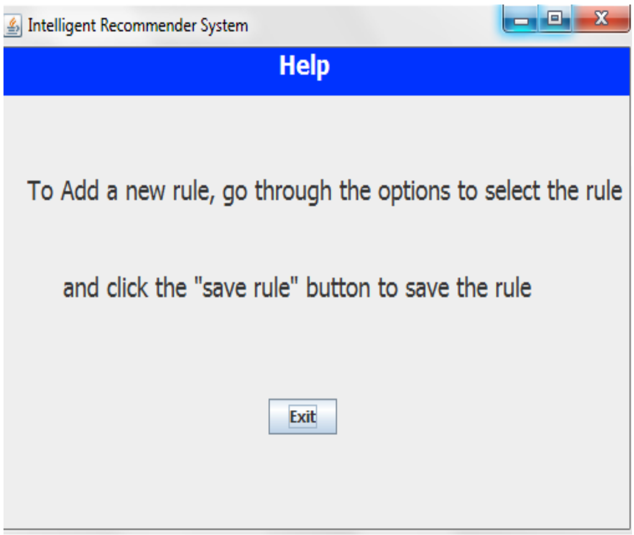

Figure 10: Help Window

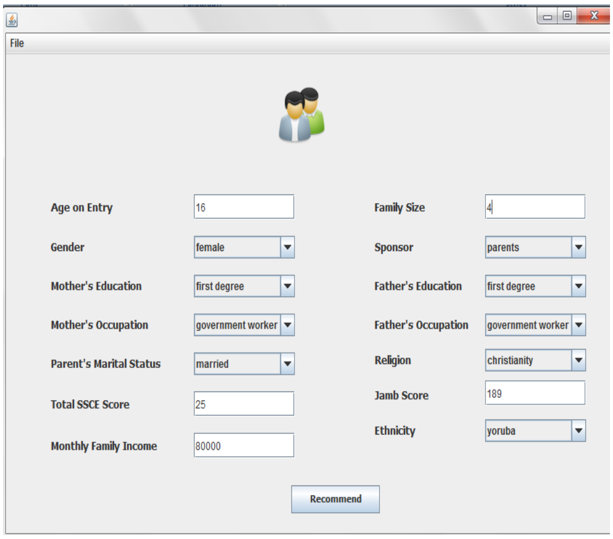

Figure 8: Input page

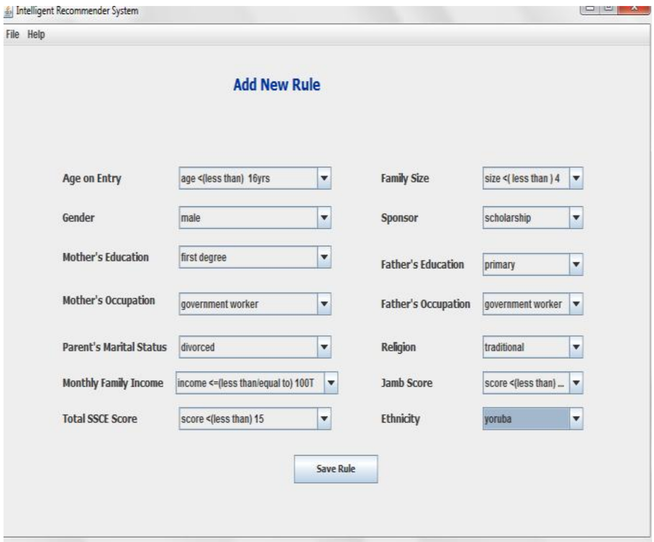

Figure 11: Add New Rule

VII. CONCLUSION AND RECOMMENDATION

The process of developing the intelligent recommender system has three major phases as follows:

• Identifying relevant background factors that impact students' academic success. For student academic success, the first year CGPA was used as the response variable, while identified background factors and previous academic achievements measured by SSCE and UME score were used for the retention model.

• Identifying the most suitable classifier for predicting student academic success using classification models based on 10-fold cross-validation and holdout methods.. Ten classifiers and multilayer perceptron were built using WEKA computing tool. Results obtained from the models were compared based on accuracy level, confusion matrices and speed of model building. It was found that random tree outperform other classifiers on all counts.

• Results of phases one and two were incorporated to develop the framework of intelligent recommender system, using the rules obtained from random tree as basis.

Thus, this research paves the way for future research to use additional significant inputs that identified background factors; larger dataset that will cover all tertiary education institutions will be needed to fully cover the wide range of students' background to generate rules that can exhaustively represent real world situations.

The related works presented in [25][26][27[28] deal with students course registration and performance in a particular course, with different student populations, different input features, and different methodologies. Hence, it is hard to make a direct comparison between the accuracy of the developed framework presented in this work and accuracy of the other developed frameworks. Moreover, this research incorporated rules obtained from the optimal algorithm to develop a system that can predict student performance as well as recommend necessary intervention strategies. To the best of the researchers' knowledge, the presented method of incorporating optimal model to develop intelligent recommender system that predict performance and recommend strategies is new to the field of modeling student performance.

Future studies could use the same procedure and models with a larger dataset and more identified background factors of student at the first year academic level. This study focused on determining to what extent family background factors and previous academic achievement affect students' first year academic performance in tertiary institution; the specific machine learning algorithm best model the student academic performance and designing a framework for predictive system be developed from these rules. This study could be extended by adding more precollege inputs to the model. Predicting student academic success at an early stage (i.e. at the end of first session) would be effective to enhance targeted students' performance. Also, it would assist in gearing intervention programs towards particular groups of students in order to address their needs. Since this work address first year academic level, models can be built for other academic level.

In most of related literature decision trees proved its superiority among other techniques. It is also preferable regarding its good predictive ability. Many other machine learning can be employed using larger dataset. There will also be need to compare more than ten classification algorithms. One of limitations was in the data set used which was limited to only first year students. Also, to fully implement the framework of intelligent recommender system designed in this study, there is need for an in-depth understanding of Web Ontology Language (OWL) and a larger dataset that will cover all tertiary education institutions will be needed to fully cover the wide range of students' background to generate rules that can exhaustively represent real world situations.

REFERENCES

[1] Boufardea E. and Garofalakis J., A Predictive System for Distance Learning Based on Ontologies and Data Mining, COGNITIVE 2012: The fourth International Conference on Advanced Cognitive Technologies and Applications, 151-158

[2] Jusoff K., S. A. A. Samah, and P. M. Isa, "Promoting university community's creative citizenry," in Proceeding of world academy of science, 2008, pp. 1-6.

[3] Luan Jing, Data Mining Applications in Higher Education, SPSS Exec.Report; 2004. http://www.spss.com/home_page/wp2.htm

[4] Marquez-Vera C., Romero C. and Ventura S. Predicting School Failure Using Data Mining. Proceedings of the 4th International Conference on Educational Data Mining, pp 271-276, 6 Jul 2011

[5] Bean J. P., "Nine themes of college student retention," in College Student Retention: Fomula for student success, A. Seidman, Ed. America: CT:Praeger, 2005, p. 29.

[6] McKenzie, K., & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. Higher Education Research and Development, 20(1), 21-33.

[7] Scott D. and Smart W. (2005) What factors make a difference to getting a degree in New Zealand? Education Counts. Tertiary Sector Performance Analysis & Reporting, Ministry of Education, New Zealand.

[8] Wimshurst, K and Allard, T (2008) Personal and institutional characteristics of student failure. Assessment and Evaluation in Higher Education, 33 (6), 687-698.

[9] DeBerard, M. S., Speilman, G. I., & Julka, D. L. (2004). Predictors of academic achievement and retention among college freshmen: A longitudinal study. College Student Journal, 38(1), 66-80.

[10] Vandamme JP, Meskens N, Superby JF. Predicting Academic Performance by Data Mining Methods, Education Economics, Vol. 15, Issue 4; December 2007. p. 405-419

[11] http://changingminds.org/explanations/research/sampling/choosing_sampling.htm Retrieved 10-10-2012

[12] Bryan, E., & Simmons, L.A. (2009). Family involvement: impacts on post-secondary educational success for first-generation Appalachian college students. *Journal of College Student Development*, 50(4), 391-405.

[13] Bradbury, B.L., & Mather, P.C. (2009). The integration of first-year, first-generation college students from ohio appalachia. National Association of Student Personnel Administrators, 46(2), 258-281. Bryan, E., & Simmons, L.A. (2009). Family involvement: impacts on post-secondary educational success for first-generation Appalachian college students. Journal of College Student Development, 50(4), 391-405.

[14] Gofen, A. (2009). Family capital: how first-generation higher education students break the intergenerational cycle. Family Relations, 58(1), 104-120.

[15] McCarron, G.P., & Inkelas, K.K. (2006). The Gap between educational aspirations and attainment for first-generation college students and the role of parental involvement. Journal of College Student Development, 47(5), 534-549.

[16] Lemnaru C. (2012). *Strategies for dealing with Real World Classification Problems*, (Unpublished PhD thesis) Faculty of Computer Science and Automation, Universitatea Technica, Din Cluj-Napoca

[17] Tsiriga V. and Virvou M., 2004. A Framework for the Initialization of Student Models in Web-based Intelligent Tutoring Systems. User Modeling and User-Adapted Interaction 14: 289-316.

[18] Piedade, M.B. and M.Y. Santos. 2008. "Student Relationship Management: Concept, Practice and Technological Support". In Proceedings of the IEMC-Europe 2008- IEEE International Engineering Management Conference (Estoril, Portugal, June 28-30) pp2-5.

[19] Piedade, M.B. and M.Y. Santos. 2009a. "An Application of the Student Relationship Management Concept". In Proceedings of the ICEIS 2009 - 11th International Conference on Enterprise Information Systems (Milan, Italy, May 6-10) INSTICC Press 2009, pp360-367.

[20] Piedade, M.B. and M.Y. Santos. 2009b. "Business Intelligence Supporting the Teaching-Learning Process". In Proceedings of the 5th WSEAS International Symposium on Data Mining and Information Processing (Budapest, Hungary, September 3-5). WSEAS Press, pp 292-297.

[21] Piedade, M.B. and M.Y. Santos. 2009c. "Business Intelligence in Higher Education Institution: Managing the Relationships with Students". In Proceedings of KDIR - International Conference on Knowledge Discovery and Information Retrieval (Funchal, Madeira, Portugal, October 6-8). INSTICC Press 2009, pp 297-302.

[22] Piedade, M.B. and M.Y. Santos. 2010. "Promoting Students Success with a Business Intelligence System". In Proceedings of CSEDU - 2nd International Conference on Computer Supported Education. (Valencia, Spain, April 7-10). INSTICC Press 2010, pp 46-52.

[23] Kanokwan K., Chun C. F. and Wudhijaya P. 2009. An intelligent recommendation system for student relationship management. The 8th International Conference on e-Business (iNCEB 2009) October28th-30th, 2009 pp87-91

[24] Kanokwan K., Chun C. F. and Tuchtawan C. 2010. Developing an intelligent recommendation system for a private university in Thailand. Issues in Information Systems Vol. XI, No1 pp467-472.

[25] Kanokwan K., Tuchtawan C. and Chun C. F. 2012. Decision Tree Modelling for an intelligent recommendation system supporting SRM for Universities in Thailand. International Conference on Computer and Information Technology (ICCIT'2012) June16-17,2012 pp81-85.

[26] Kanokwan K., and Chun C. F. 2012. Neural Network Modeling for an Intelligent Recommendation System Supporting SRM for Universities in Thailand. WSEAS TRANSACTIONS on COMPUTERS Vol. 11, Issue 2 pp34-44.

[27] Vialardi C., Bravo J., Shafti L. and Ortigosa A. (2009). Recommendation in Higher Education Using Data Mining Techniques, Educational Data Mining, pp190-199, 2009

[28] Shana J. and Venkatachalam T. 2012. A framework for Dynamic Faculty Support System to Analyze Student Course Data. International Journal of Emerging Technology and Advanced Engineering, Vol. 2 Issue 7 pp478-482

[29] Kuyoro S. O. 2010, Investigating the Effect of Students Socio-Economic/Family Background on Students Academic Performance in Tertiary Institutions Using Decision Tree Algorithms. Department of Computer Science, University of Ibadan. Unpublished MSc Thesis

[30] Quinlan J.R. 1996. Improved use of continuous attributes in C4.5. Journal of Artificial Intelligence Research, 4:77-90.