Available online @ www.actasatech.com

*acta*SATECH **4**(2): 61 - 78 (2013)

Research

Investigating the Effect of Students Socio-Economic/Family Background on Students Academic Performance in Tertiary Institutions using Decision Tree Algorithm

***Adeyemo, A.B. & Kuyoro, S. O.**

Department of Computer Science,
University of Ibadan,
Ibadan, Nigeria

*Correspondence author <afolashadeng@gmail.com>

Abstract

The causes of the difference in the academic performance of students in tertiary institutions has for a long time been the focus of study among higher education managers, parents, government and researchers. The cause of this differential can be due to intellective, non-intellective factors or both. From studies investigating student performance and related problems it has been determined that academic success is dependent on many factors such as; grades and achievements, personality and expectations, and academic environments. This work uses data mining techniques to investigate the effect of socio-economic or family background on the performance of students using the data from one of the Nigerian tertiary institutions as case study. The analysis was carried out using Decision Tree algorithms. The data comprised of two hundred forty (240) records of students. The academic performance of students was measured by the students' first year cumulative grade point average (CGPA). Various Decision Tree algorithms were investigated and the algorithm which best models the data was used to generate rule sets which can be used to analyze the effect of the socio-economic background of students on their academic performance. The rules generated can serve as a guide to educational administrators in their planning activities.

Keywords: Socio-Economic, Intellective, Family Background, Academic Performance.

Introduction

The differential students' performance in tertiary institutions has been and is still a source of great concern and research interest to the higher education managements, government and parents because of the importance that education has on the national development (Asikhia, 2010). Academic institutions are increasingly required to monitor their performance and the performance of their students. This gives rise to a need to collate, analyze and interpret data, in order to have evidence to inform academic policies that are aimed at, for example, improving student retention rates, allocating teaching and support resources, or creating intervention strategies to mitigate factors that may affect student performance adversely. There are a number of reasons for this: 1) Tertiary education should aim to maximize the potential of each student. Therefore, a careful examination of student outcomes against some benchmark or expected outcome may provide evidence as to whether student potential is being realized. Such insights may also help the institutions to prioritize scarce resources, to focus them on specific problem areas, 2) Institutions have an obligation to deliver value for money to the bodies that fund them, 3) Institutions are often judged by the quality of the awards that they provide; for example the more honours level graduates a course provides, the better the course is perceived to be. This provides additional incentive for institutions to take proactive steps to support students.

The observed poor performance of students in tertiary institutions has been partly traced to poor academic background. But it is worthwhile to note that other socioeconomic factors can also influence the performance of students in tertiary institution. Various factors that may likely influence the performance of a student in addition to academic background include: age on admission, parental educational background, family income level, social status, parents' occupation, education sponsor, cohabitation status of parents, family size, position in the family and time lapse before gaining admission. Bakare (1975) summarized the factors and variables affecting students' performance into the intellective and non-intellective factors, emphasizing that the intellectual abilities were the best measure. He categorized causes of poor academic performance into four major classes namely: Causes resident in society, Causes resident in school, Causes resident in the family and Causes resident in the student. Anderson et al., (1994) studied the effect of factors such as gender, student age, and students' high school scores in mathematics, English, and economics, on the level of university attainment. According to the study, students who received better scores in high school also performed better in university. Also men had better grades than women and choose to drop from school less often. McKenzie and Schweitzer (2001) investigated academic, psychosocial, cognitive and demographic predictors of academic performance to improve interventions and support services for student at risk of academic problems. They recommended implementing stringent record keeping procedures at the university level to enable researchers to fully examine the relationship between age, previous academic performance and university achievement. Golding and Donaldson (2006) stated that the use of performance in first year computer science course is a possible factor, which may determine academic performance. They also showed that gender and age have no significant correlation as predictive factors. Superby et al. (2006) and Vandamme et al. (2007) studied correlations of various parameters such as attendance, estimated chance of success, previous academic experience and study skills. They found out that changing process factors during a student's stay at the university plays a large part in academic performance. In addition they experimented on predicting students' performance using decision tree, neural network**s** and linear discriminant analysis. The rates of prediction obtained were not particularly good due to the difficulty to classify students into 3 groups, namely, high risk, medium risk and low risk, before the first university examinations.

In this paper, the socioeconomic conditions of students as an influencing factor on their performance in tertiary institutions, along with their previous academic records which had been used in previous studies, were used to propose a data mining model for analyzing and predicting the academic performance of students in higher institutions with a view to quantifying the effect of these socio-economic factors on their University studies. For this work, two hundred and forty (240) records of students from a University in Nigeria (Babcock University) from 2002 to 2009 were used. Data mining techniques have been applied in many application domains such as banking, fraud detection, network intrusion detection and telecommunications (Hans and Kamber, 2003). Data mining methodologies have also being used to enhance and evaluate the higher education tasks. Many researchers have proposed some methods and architectures for using data mining for higher education. Delavari and Beikzadeh (2004) proposed a model for using data mining in a higher educational system to improve the efficiency and effectiveness of the traditional processes. Kalles and Pierrakeas (2004) in an effort to analyze students' academic performance through the academic years, as measured by the students home work assignments, had attempted to derive short rules that explain and predict success or failure in the final exams using different machine learning techniques (decision trees, neural networks, Naive Bayes, instance-based learning, logistic regression and support vector machines) and compared them with genetic algorithm based induction of decision trees. Delavari et al (2005) proposed an analysis model and used it as a roadmap for the application of data mining in higher educational system. The model allows the decision makers to better predict which students are less likely to perform well in that specific course, or those who are less likely to be successful in it. Adeyemo and Kuye (2006) presented an evaluation of the factors that contribute to the academic performance of students admitted into the university. The variables of interest were the entry qualification and admission mode and how these factors affect the academic performance of the students. Student admission data obtained from a case study department in one of the university's in Nigeria was used. The results indicated that the observed performance of student whose admission into the case study department is through the University Matriculation Examinations (UME) depends more on their respective Senior School Certificate Examination (SSCE) performance than their entry scores in the UME examination used as the basis for their admission. Osofisan and Olamiti (2009) investigated the academic background in relationship with the performance of students in a computer science programme in a Nigerian university. Results indicated that the grade obtained from Senior Secondary Certificate Examination (SSCE) in mathematics is the highest determinant used by the C4.5 learning algorithm in building the model of the students' performance. Also, the result indicated that even if a student does not finish his programme in the normal number of (four) academic sessions for whatever reasons he would still graduate with minimum of second class lower if he took further mathematics at SSCE examination. Students who spend more than four academic sessions in the programme and did not take further mathematics at SSCE examination are more likely to graduate with class below second class lower. Other studies tried to identify the significant factors that can influence tertiary students' academic performance in a more detailed way. Many studies included a wide range of potential predictors, including personality factors, intelligence and aptitude tests, academic achievement, previous college achievements, and demographic data and some of these factors seemed to be stronger than others; but there is no consistent agreement among different studies. However, all studies show that academic success is dependent on many factors, where grades and achievements, personality and expectations, as well as sociological background all play significant roles.

Materials and Methods

Data Formatting

Data mining is an integral part of Knowledge Discovery in Databases (KDD), which is the overall process of converting a series of transformation steps, from data pre-processing to post-processing of data mining results. The data pre-processing has to do with gathering or collection of data, and data cleaning through data transformation. During data selection, the relevant data is gathered. Once the data has been assembled, its quality must be verified. Incomplete (lacking certain attributes of interest, or containing only aggregate data), noisy (containing errors, or outlier values that deviate from expected), and inconsistent (for example, discrepancies in the codes used to categorize items) data are common. Data cleaning routines attempt to clean the data by filling in missing values; smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. Finally, the cleaned data are transformed into a format suitable for data mining.

The data gathering process for this study involves the collection of the raw data about the students from students' records. Academic records collected include students SSCE results, post-UME score (entry qualification into the University) and first year CGPA. Students socio-economic records which were obtained from administrative forms filled by the students on admission into the University include their gender, age on entry, mother's educational qualification, father's educational qualification, sponsor, family size, student's position in the family, types of secondary school attended, location of school, residence location, mother's occupation, father's occupation, marital status of parents, social class and year lapse before gaining admission.

The data used was collected from the records of students admitted in Babcock University in Nigeria from year 2002 to 2009. Students whose records are incomplete were filtered out from the dataset. After the data cleansing, two hundred and forty (240) complete records of students were selected for the purpose of this study and a data model for storage and analysis of the data was formulated. Eighteen (18) attributes composed of thirteen (13) categorical and five (5) continuous attributes were selected including the target/class variable: The students 100 level CGPA was used as the class variable while other variables were used as the predictor variables. The data format is presented in Table 1.

Table 1: Student data format

S/N	Variable Name	Variable description/format	Variable Type
1.	Age on entry	Students age on admission	Continuous
2.	Gender	Male or Female	Categorical
3.	Total SSCE results score	A1-6 B2-5 B3-4 C4-C6-3 D7-2 E8-1 F9-0 **Subjects considered** Eng, Math, Chemistry, Biology, Agriculture or Food&Nutrition and 1 Nigerian language	Continuous
4.	Post-UME score	Students Post-UME score	Continuous
5.	Year lapse before admission	Years between completion of O/L and admission into university	Continuous
6.	Social class	Upper, Middle, Lower	Categorical
7.	Mother's educational qualification	Primary, SSCE, 1st degree, 2nd degree, PhD	Categorical
8.	Father's educational qualification	Primary, SSCE, 1st degree, 2nd degree, PhD	Categorical

S/N	Variable Name	Variable description/format	Variable Type
9.	Marital status of parents	Married, Divorced, Separated, Widowed	Categorical
10.	Mother's occupation	Government worker, Private, Self employed	Categorical
11.	Father's occupation	Government worker, Private, Self employed	Categorical
12.	Family size	Total number of children in family and parents	Continuous
13.	Student's position in the family	1st born, last born, only child, others	Categorical
14.	Type of secondary school attended	Private, State, Federal	Categorical
15.	Location of secondary school	Rural, Semi-Urban, Urban	Categorical
16.	Residence location	Rural, Semi-Urban, Urban	Categorical
17.	Sponsor	Parents, Scholarship, Self, Others	Categorical
18.	100 level CGPA	A: 4.5-5.0, B+:3.5-4.49, B: 3.0-3.49, C: 2.5-2.99, D:2.0-2.49, E: 1.0-1.99, F:<1.0	Categorical

Data Processing

Decision Tree algorithms were used to analyze the data because they can be used to classify mixed data types, the algorithm is easy to understand and the results are very easy to interpret. Decision tree rules like decision trees are also easy to explain because the rules are in IF-THEN format. The data mining software used to generate the decision tree algorithms for this study are See5 for Windows and Waikaito Environment for Knowledge Analysis (WEKA). See5 for Windows can only generate decision trees based on C5 algorithm while WEKA allows many algorithms giving room for comparison to determine the better classifier among those used in this study.

See5 classifiers are expressed as Seetrees (decision trees) and can also directly generate rule sets otherwise referred to as See5 Rules. The software automatically generates training and test data from the total record of data input into the software. There are several classifiers available in WEKA but CART (SimpleCart), REPTree and C4.5 (J48) was used for the purpose of this study. The RIpple DOwn Rule Learner (Ridor) which is a rule based learner was also used. Attribute importance analysis was carried out to rank the attributes by significance using Information gain. Correlation-based Feature Subset Selection (CfS) and Consistency Subset Selection (CoE) filter algorithm were used to rank and select the attributes that are most useful. Both selection techniques use the BestFirst-D1-N5 searching technique. The F-measure and the AUC which are well known measures of probability tree learning were used as evaluation metrics for models generated by the classifier software.

Decision Tree Analysis

See5 can generate single decisions tree and boosted decision trees. The single decision tree generated showed a large error rate both for the training and test data, therefore the boost tree was used to reduce the error rate. The Boost decision tree generated 10 trials of the single decision tree in order to reduce the error rate. The 10 trials generated a maximum of 71 leaves and minimum of 52 leaves. The performance of the classifier constructed at each trial is presented in Table 2. Each trial result is summarized on a separate line, while the line labeled boost shows the result of voting all the classifiers. The decision tree constructed on Trial 0 is identical to singe decision tree. Some of the subsequent trees were generated by paying more attention to certain cases have relatively high overall error rates. When the trees are combined by voting, the final predictions have a lower error rate of 7.1% on the training cases and 7.3% on test cases. Table 3 presents the attribute usage statistics for the best decision tree trial.

Table 2: See5 Boost Decision Tree Summary

	Training		**Test**	
Trials	**Size**	**Errors**	**Size**	**Errors**
0	71	64(26.7%)	71	41 (27.3%)
1	60	83(34.6%)	60	52(34.7%)
2	61	86(35.8%)	61	58(38.7%)
3	63	85(35.4%)	63	52(34.7%)
4	65	81(33.8%)	65	47(31.3%)
5	67	89(37.1%)	67	57(38.0%)
6	59	78(32.5%)	59	52(34.7%)
7	54	88(36.7%)	54	55(36.7%)
8	52	98(40.8%)	52	67(44.7%)
9	56	91(37.9%)	56	58(38.7%)
boost		17(7.1%)		11(7.3%)

Table 3: See5 boost decision tree Attribute usage

Attributes	Percentage of usage
Parents Marital status	100%
Sponsor	97%
Gender	96%
Family size	89%
Mother's education	88%
Secondary School Location	86%
Father's occupation	84%
Age on entry	80%
Total SSCE score	75%
Post UME score	74%
Student's Position in the family	74%
Secondary School Type	67%
Residence Location	46%
Father's education	42%
Social Class	40%
Year lapse before admission	33%
Mother’s occupation	30%

Four decision tree algorithms in WEKA were used to analyze the data set. These are the CART algorithm, the REPTree algorithm, the WEKA J48 which implements the C4.5 algorithm and the RIpple DOwn Rule Learner (Ridor).

Table 4 presents the summary of the results of the CART, REPTree and C4.5 algorithms.

Table 4 Classifiers comparison using F-Measure and AUC

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
CART	0.400	0.236	0.326	0.400	0.346	0.622
REPTree	0.500	0.180	0.515	0.500	0.478	0.768
C4.5(J48)	0.713	0.099	0.733	0.713	0.691	0.939

The Prediction Accuracy of the algorithms can be determined by comparing their True Positive (TP) rate, False Positive (FP) rate, F-Measure and AUC. The True Positive (TP) rate is the proportion of cases which were classified as the actual class, that is, how much part of the class was captured. It is equivalent to Recall. In the confusion matrix, this is the diagonal element divided by the sum over the relevant row. The False Positive (FP) rate is the proportion of cases which were classified as the one class, but belong to a different class. In the matrix, this is the column sum of the class minus the diagonal element, divided by the rows sums of all other classes. The Precision is the proportion of the cases which truly have the actual class among all those which were classified as the class. In the matrix, this is the diagonal element divided by the sum over the relevant column. The F-Measure is simply 2*Precision*Recall/(Precision+Recall), a combined measure for precision and recall. The Receiver Operating Characteristic (ROC) is the graphical display of TPR versus FPR while the AUC represent the Area under the ROC curve. These measures are useful for comparing classifiers based on the accuracy.

Rule Generation

The See5 decision tree results can also be presented in the form of rules which are easier to understand and use. Each rule consists of:

1. A rule number that serves only to identify the rule
2. Statistics (*n*, lift *x*) or (*n*/*m*, lift *x*) that summarize the performance of the rule
3. *n* is the number of training cases covered by the rule and *m* shows how many of them do not belong to the class predicted by the rule. The rule's accuracy is estimated by the Laplace ratio *(n-m+1)/(n+2)*. The lift *x* is the result of dividing the rule's estimated accuracy by the relative frequency of the predicted class in the training set
4. One or more conditions that must all be satisfied for the rule to be applicable
5. Class predicted by the rule
6. A value between 0 and 1 that indicates the confidence with which this prediction is made, and
7. Default class that is used when none of the rules apply.

Most of the rules indicate that students whose parents are the sponsor for their education end up with at least C grade. Also students whose parents are self employed especially the mothers have at least C grade. Some of the rules generated are presented here:

Rule 1: (3/1, lift 20.6)

age on entry > 18
Gender = Female
Mother's education = 1stdegree
Residence Location = Semiurban
Sponsor = Parents
-> class A [0.600]

Rule 2: (2, lift 2.3)

age on entry > 19

Parents marital status = Separated

-> class C [0.750]

Rule 3: (8/2, lift 2.2)

Year lapse before admission > 1

Parents marital status = Married

Secondary School Location =

Semiurban

Sponsor = Parents

-> class B+ [0.700]

Rule 4: (4, lift 3.8)

Gender = Male

total SSCE score > 29

Mother’s education = 1stdegree

Father's occupation = Government

Secondary School Type = Private

Secondary School Location = Urban

Sponsor = Parents

-> class B [0.833]

Rule 5: (6/1, lift 3.4)

Gender = Female

Mother’s education = 1stdegree

Father's education = 1stdegree

Parents marital status = Married

Father's occupation = Government

Secondary School Location = Semiurban

-> class B [0.750]

Rule 6: (2, lift 2.8)

Father's education = 2nddegree

Parents marital status = Divorced

Father's occupation = Private

-> class C [0.750]

Rule 7: (2, lift 6.7)

Gender = Female

Father's occupation = Private

Sponsor = Scholarship

-> class D [0.750]

Rule 8: (1, lift 16.0)

Father's education = 2nddegree

Parents marital status = Divorced

Father's occupation = Selfemployed

-> class E [0.667

The RIpple DOwn Rule Learner (Ridor) was used to generated rules similar to those generated by the See5 algorithm rule generator. Fifteen (15) Ridor rules were generated. The longest rule consists of 6 attributes and the shortest rule has 2 attributes. The default rule is 100LCGPA = B+.

Comparison of Algorithms

Table 5 presents the comparison among classifiers used in the study.

Table 5: Classifiers comparison using classification accuracy rate

Classifier	**CART**	**REPTree**	**C4.5 J48**	**SeeRule**	**SeeTree**	**Boost SeeTree**	**Ridor**
Accuracy	0.40	0.50	0.71	0.65	0.73	0.93	0.48

Classification and Regression Tree (CART) has the lowest accuracy rate of 0.40 followed by REPTree with the accuracy rate of 0.50. As shown in the table, Boost decision tree with 0.93 accuracy level is considered to be better than other classifiers followed by the single decision tree with accuracy rate of 0.73. C4.5 decision tree shows better accuracy than other classifiers generated by WEKA. Comparing the two rulesets (See5 Rule and Ridor), the See5 rule show better accuracy than Ridor.

Table 6 presents the results of the attribute importance analysis using CfS and CoE which demonstrates attribute that have strong correlation with students' first year CGPA. They both evaluate the worth of an attribute by measuring the information gain with respect to the class. Table 7 presents the Attribute importance ranking using 10 fold cross-validation for CfS and CoE.

Table 6: Attribute importance ranking using Information gain

Attribute	**Information gain**	**Cfs**	**CoE**
Parents Marital Status	0.0664	4	5
Mother's education	0.0612	2	3
Student's position in the family	0.0534	7	8
Sponsor	0.0525	11	12
Father's education	0.0502	3	4
Father's occupation	0.0496	6	7
Gender	0.0491	1	1
Mother's occupation	0.0448	5	6
Secondary school type	0.0444	8	9
Secondary school location	0.0434	9	10
Residence location	0.0389	10	11
Social class	0.0238		2
Family size	0		
Age on entry	0		
Year lapse before admission	0		
Total SSCE score	0		
Post UME score	0		

CfS-Correlation based Feature Subset Selection evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them.

CoE-Consistency Subset Selection evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes.

Table 7: Attribute importance ranking using 10 fold crossvalidation for CfS and CoE

	Fold %	
Attribute	**Cfs**	**CoE**
Parents Marital Status	100%	100%
Mother's education	100%	100%
Student's position in the family	100%	100%
Sponsor	100%	100%
Father's education	100%	100%
Father's occupation	100%	100%
Gender	100%	100%
Mother's occupation	100%	100%
Secondary school type	100%	100%
Secondary school location	100%	100%
Residence location	100%	100%
Social class	0%	100%
Family size	0%	0%
Age on entry	0%	0%
Year lapse before admission	0%	0%
Total SSCE score	0%	0%
Post UME score	0%	0%

CfS-Correlation based Feature Subset Selection evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them.

CoE-Consistency Subset Selection evaluates the worth of a subset of attributes by the level of consistency in the class values when the training instances are projected onto the subset of attributes.

The attribute importance ranking in Table 6 is similar to Table 7. It shows the percentage of fold in the selection of each attribute. CfS and CoE selected similar attributes with the addition of Social class for CoE. CfS ranked 11 attributes and CoE ranked 12 attributes as having better predictive ability compare to other attributes. These attributes are similar. Six out of them, namely, parent marital status, mother's education, father's occupation, mother's occupation, father's education and sponsor are related to the students' parental information while five other variables ranked by both CfS and Coe are students' position in the family, gender, secondary school type and location, Coe added social class as important attribute in determining students performance. This clearly indicates that majority of the attributes influencing the prediction of the students' academic performance according to this finding has to do with parents, education sponsor and home environment in general.

Discussion of Results

The attribute usage (Table 3) shows that parents marital status, sponsor, mother's education and father's occupation rank higher than other socio-economic factors. The percentage usage for marital status is 100%, sponsor 97%, mother's education 88% and father's occupation 84% indicating that parental background and education sponsor contribute immensely to the academic performance of students. Age on entry, secondary school location and total SSCE score also shows more than 30% contribution to the performance of students.

The Boost decision tree shows that all the attributes have considerable effect on the performance of students with the parental information and education sponsor ranking highest. All the attributes contributed at least 30% to the outcome of the student grade class. This shows that parental background and education sponsor are important factors to be considered for the good performance of students in tertiary institutions.

FRAMEWORK FOR STUDENTS PERFORMANCE PREDICTION SYSTEM

The rules inferred from the boosted decision tree generated by C5 (See5) algorithm which gave the best overall performance in the modeling of the data set used in this study can be used as the basis for designing a Student's Academic Performance Prediction software (Decision Support System) for higher education institutions. The prediction system will maintain a historical database from which a dataset of students and respective grade can be obtained. The rules will be generated by the software's Pattern discovery module which will use the training data as input for generating decision trees from which the rules can be obtained. This system will then make prediction based on these rules. The proposed framework for this application is shown in Figure 1.

CONCLUSION

This study identified parental conditions such as parents' education and occupation, marital status and sponsor as influencing socioeconomic factors affecting the performance of students in tertiary institutions. This we think is due to the fact that students whose parents' occupation are very demanding and barely have time for the students will have an unsound academic background which will influence the performance at the tertiary institution. Also, except in rare cases students whose parents are unable to provide adequately for the sponsoring of the education are distracted and perform poorly. Therefore, the parents socio-economic background determine to a very great extent the academic achievement and overall success of their children and parents are urged to provide adequate resources and the needed encouragement for the students.

The findings of this study agree with previous studies such as (Bakare, 1975; Anderson et al., 1994; Superby et al, 2006; Vandamme et al., 2007, McKenzie&Schweitzer, 2001, Adeyemo&Kuye, 2006 and Osofisan&Olamiti, 2009). These previous studies had focused majorly on past academic background of students while this present study on the other hand investigated family and socioeconomic factors of students in addition to previous academic records of students as predictive attributes of students' academic performance in tertiary institutions. The result corroborates the findings of Bakare 1975 that factors affecting students' academic performance can be intellective or non-intellective, that is, the determinant factors of students' performance can be traced to the society, school, family, or student.

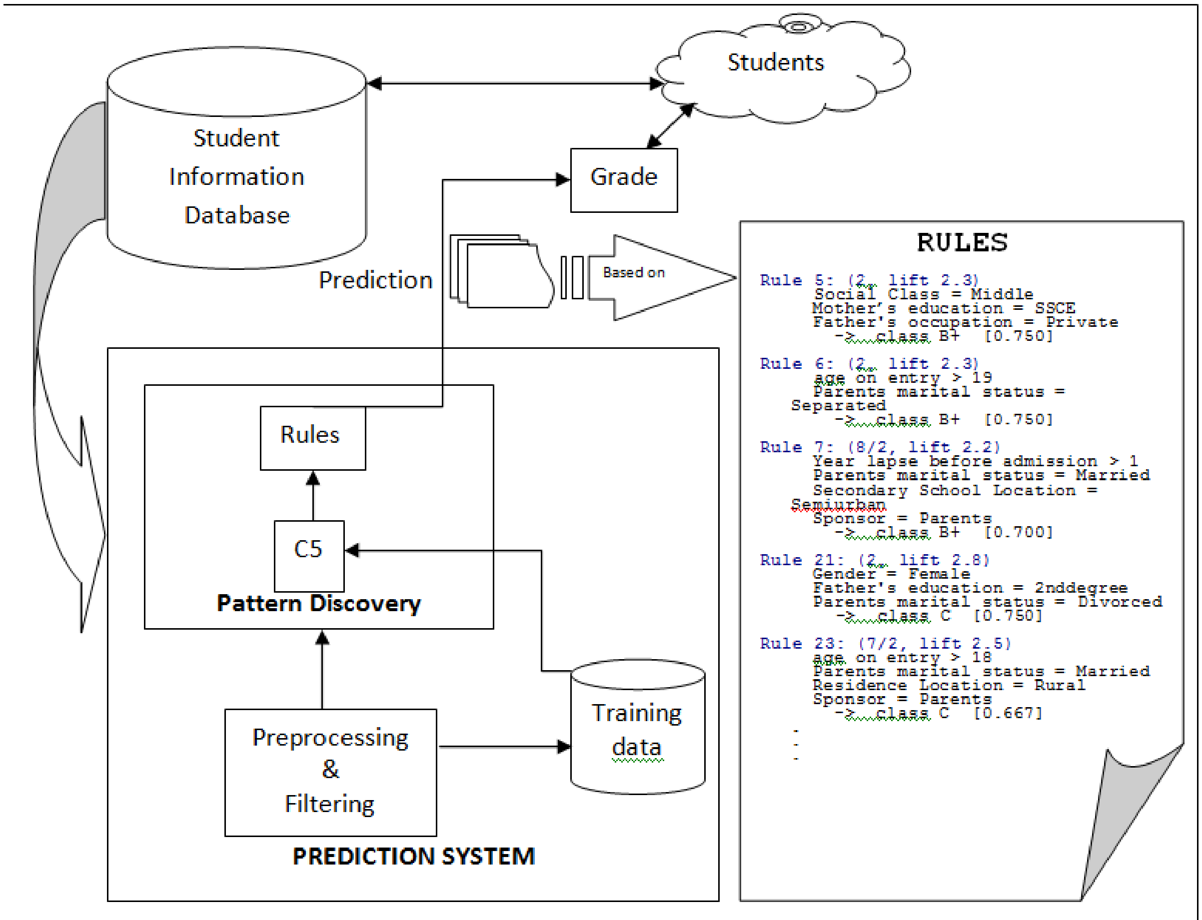

Figure 1: Students' Academic Performance Prediction System Architecture

Many parents may not be aware of the influence of various family background factors on the academic achievement of their children. It is recommended that educationists, leaders and higher education managements should try to create awareness in parents about the importance of the family background on academic achievement motivation which can improve the children's performance. Parents need to be informed that they can contribute to the education of their children through encouragement, provision of funding and other resources, and active assistance among other strategies. Higher education managements can use this as a basis for organizing seminars or counseling sessions for students experiencing problems in their academics and parents.

REFERENCES

Adeyemo, A. B and Kuye G., 2006. Mining Students' Academic Performance using Decision Tree Algorithms. *Journal of Information Technology Impact* 6(3): 161-170.

Anderson, G., Benjamin, D., and Fuss, M. 1994. The Determinant of Success in University, Introductory Economics Courses. *Journal of Economic Education*. 25:99-120.

Asikhia O.A. 2010 Students and Teachers' Perception of the Causes of Poor Academic Performance in Ogun State Secondary Schools [Nigeria]: Implication for Counseling for National Development. *European Journal of Social Sciences*. 13(2):229-242.

Bakare, C.C. 1975. Some Psychological Correlates of Academic Success and Failure. *African Journal of Educational Research*.

Delavari N, Beikzadeh M. R. 2004 A New Model for Using Data Mining in Higher Educational System, *5th International Conference on Information Technology based Higher Education and Training*: ITEHT '04, Istanbul, Turkey.

Delavari N, Beikzadeh M. R, Amnuaisuk S. 2005. Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System. *6th Annual International Conference*: ITEHT Juan Dolio, Dominican Republic.

Golding, P. and O. Donaldson, 2006. Predicting academic performance. *Proceedings of the 36th ASEE/IEEE Frontiers in Education Conference T1D-21*, San Diego, CA.,1-6.

Han J, Kamber M. 2003. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, New Delhi.

Kalles D., Pierrakeas C. 2004, *Analyzing student performance in distance learning with genetic algorithms and decision trees*, Hellenic Open University, Patras, Greece.

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

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten 2009. *The*

WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

McKenzie, K. and R. Schweitzer, 2001. Who succeeds at University? Factors predicting academic performance in first year Australian university students. *Higher Educ. Res. Dev.*, 20: 21-33.

Osofisan, A. O. and Olamiti, A. O. 2009. Academic Background of Students and Performance in a Computer Science Programme in a Nigerian University. *European Journal of Social Sciences*. 9(4): 564-572.

Quinlan J.R. 1993 C4.5: *Programs for Machine Learning*. Morgan Kaufmann, San Mateo California.

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

Superby, J.F., J.P. Vandamme and N. Meskens, 2006. Determination of factors influencing the achievement of the first-year university students using data mining methods. *Proceedings of the 8th international conference on intelligent tutoring systems, Educational Data Mining Workshop, (ITS'06),* Jhongali, Taiwan, 37-44.

Vandamme, J.P., N. Meskens and J.F. Superby, 2007. Predicting academic performance by data mining methods. *Educ. Econ*., 15: 405-419.