
A recommender for improving the student academic performance

Maria Goga\textsuperscript{a}, Shade Kuyoro\textsuperscript{b}, Nicolae Goga\textsuperscript{c,d,*}

\textsuperscript{a}Department of Science of Education, Universitatea Tehnica de Constructii Bucuresti, Bulevardul Lacul Tei 124, Bucuresti 020396, Romania
\textsuperscript{b}Department of Computer Science, Babcock University, Ilishan Remo Ogun State Nigeria, Nigeria
\textsuperscript{c}Department of Engineering in Foreign Languages, Politehnica Bucharest, Splaiul Independentei 313, Bucuresti, Romania
\textsuperscript{d}Molecular Dynamics Group, University of Groningen, the Netherlands

Abstract

There is a growing awareness among researchers about the apparent variations in the academic performance of students in tertiary institutions. Although, many studies have employed traditional statistical methods in identifying the factors responsible for the disparity, the statistical tool for setting a yardstick is yet to be established. Machine learning techniques have been employed as a paradigm in the modeling of students’ academic performance in higher learning. However, they could be the springboard for improving prediction of students’ academic performance. This work therefore aimed at designing a framework of intelligent recommender system, based on background factors, which can predict students’ first year academic performance and recommend necessary actions for improvement.

© 2015 The Authors. Published by Elsevier Ltd. Peer-review under responsibility of The Association “Education for tomorrow” / [Asociatia “Educatie pentru maine”].

Keywords: Recommender System, Family Background, Student Performance

1. Paper Rationale

The differential students’ performance in tertiary institutions has existed and is still a source of great concern and research interest to the higher education managements, government, parents and other stakeholders because of the importance of education to the national development. Academic institutions are increasingly required to monitor

* Corresponding author.
\textit{E-mail address:} goga.maria@yahoo.com
both their performance and that of their students. This gives rise to a need to extract useful information from the available students’ large datasets to inform academic policies on how best to improve student retention rates, allocate teaching and support resources, or create intervention strategies to mitigate factors that affect student performance adversely. Maximizing the potential of students, providing evidence of delivering value for money to the bodies that fund them, and performing up to expectation is very crucial to tertiary institutions. Most institutions are often judged by the quality of the awards they provide. For instance, the more honors level graduates a course provides, the better the course is perceived to be. This provides additional quest for institutions to take proactive steps to investigate students’ data with a view to finding useful information that can aid planning activities, decision making and students’ intervention strategies. It is necessary to carefully measure student outcomes or expected outcomes that may provide evidence as to whether a student potential is being realized against some benchmarks.

The focus of this work is on designing a framework of intelligent recommender system that can predict tertiary students’ first year academic performance as well as recommend necessary actions to be taken to aid the students, thus guiding the tertiary institution management in their decision-making about early intervention strategies. This paper will focus on the educational site of this research and less on the computational site of this work. For a detailed description of the computational site we refer to (Kuyoro, Oludele, Okulie & Goga, 2013).

2. Paper theoretical foundation and related literature

Student Performance and family background The literature is replete with various studies relating machine learning to university admission, student performance, and other related problems. From diverse literature, the observed poor performance of students in tertiary institutions has been partly traced to poor academic background and wide range of other predictors, including personality factors, intelligence, gender and aptitude tests, academic achievement, previous college achievements, and demographic data (Bean, 2005), (Bryan & Simmons, 2009), (DeBerard, Speilman, & Julka, 2004), (Kuyoro, 2010), (Marquez-Vera, Romero & Ventura, 2011), (McKenzie & Schweitzer, 2001), (Scott & Smart, 2005), (Vandamme, Meskens & Superby, 2007), (Vialardi, Bravo, Shafti & Ortizosa, 2009), (Wimshurst & Allard, 2008).

According to reviewed literature, the factors which affect student’s performance are the following: Age (Simpson, 2006), (Golding and Donaldson, 2006), (Vandamme, Meskens & Superby, 2007); Gender (Anderson, 1994), (Veitch, 2004), (Boero, Laureti & Naylor, 2005), (Herzog, 2006); Father’s education (Vandamme, Meskens & Superby, 2007), (Dekker, Pechenizky & Vleeshouwers, 2009), (Cortez & Silva, 2008); Mother’s education (Ishitani, 2003), Father’s occupation (Vandamme, Meskens & Superby, 2007); Mother’s education (Veitch, 2004), (Hijazi & Naqvi, 2006); Family income (Johnson, 1996), (Khan, 2005); Ethnicity (Woodman, 2001); Parents’ marital status (Hodge & Mellin, 2010); Family size (Hodge & Mellin, 2010), (Cortez & Silva, 2008); Religion (Ishitani, 2003); Sponsor (Hodge & Mellin, 2010); UME Score (Adeyemo & Kuye, 2006), (Osifisan & Olamiti, 2009); SSCE Grade (Oladokun, Adebanjo & Charles-Owaba, 2008); CGPA. This factor is included in almost all the reviewed studies.

Application of machine learning algorithms for student performance Many researchers have come to some interesting conclusions as to which of these predictors has impacted students’ academic performance in tertiary institutions. Though, literature shows that academic success is dependent on one factor or the other; there is no consistent agreement among different studies. (Adeyemo & Kuye, 2006) built decision tree algorithms to evaluate 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. (Quadri & Kalyankar, 2010) used decision trees to make important design decisions about the interdependencies among the properties of dropout students. The study provides examples of how machine learning algorithms can be used to improve the effectiveness and efficiency of the modeling process. (Yadav, Bhawar & Pal, 2012) examine the quality of the predictive models generated by the machine learning algorithms used (ID3, C4.5 and ADTree) using existing records of students. Some of the Machine Learning algorithms were able to establish effective predictive models from the existing student retention data. (Kanokwan, Chun & Wudhijaya, 2009) based on (Piedade and Santos, 2008) 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.

Tinto (1982) pointed out that background factors and a host of other factors contribute to the academic success of students in tertiary institutions. (Marquez-Vera, Romero & Ventura, 2011) corroborate this claim by tagging the issue of students’ academic success one-hundred factor problem. Many authors at one time or the other focus on different problem and come out with different conclusions. However, (Herrera, 2006) affirms that ‘different models which differentiate between students’ success should be constructed for each academic level.’ This is due to the fact that variables that affect success at one academic level will not necessarily affect it at a different academic level.

In accordance with this school of thought, though studies have been carried out at one time or the other showing that various predictors at various time and different locations contribute to the outcome of students, there exist some evidence that students’ background information contribute immensely to the early prediction of student success. Though none of the studies directly shows how family background factors relate to student performance, it is necessary to construct a model to capture students’ success at the first-year level. Moreover, existing recommender systems are doing this only for a particular course. Giving a good prediction for student academic performances for a particular course is no guarantee that it will give a good prediction for another course. Therefore it is necessary to design a recommender that can give a global prediction at the end of each student session.

This study serves as furtherance to other studies on machine learning algorithms and students’ academic performance by focusing on adopting machine learning algorithms to determining the influence of family background factors coupled with students’ previous academic achievement on students’ first year academic performance in tertiary institutions. The construction of a model of students’ academic success at the end of the first year in tertiary institution which can serve as building block of model for other students’ levels, and the designing a system that can predict the overall performance of students at the end of first academic session in tertiary institution is also necessary.

3. Methodology

We used purposing sampling for this research obtained from the student record files from Babcock University, Nigeria. Within the period of 2001 and 2010, a total of eleven thousand and twenty-seven students (11,027) students were admitted into Babcock University. Seven thousand, five hundred (7,500) complete records of the enrolled students representing 68% of the entire population were randomly selected for the purpose of this study.

Once a candidate enrolls at a university in Nigeria, it is given to him/her an enrolment form that is filled up. We get required information for this research from those enrolments student forms from Babcock University. We extracted information related to family background factors such as age, gender, parent occupations and education, parents marital status, religion, etc.; previous educational performances obtained from the scores in relevant subject matters before the university entry and entry examination score. Score of the students at the end of first session was also taken into account.

Based on the reviewed literature (see Section 2), what information was available and interview with the experts, the following variables were considered for the model: age, gender, marital status of the parents, mother and father educational level, occupations of the parents, SSCE score (Secondary School Certificate Examination Grade)UME score (University Matriculation Examination), CGPA score from first year (Cumulative Grade Point Average). The interview includes structured questions given to four people in the field of education (within and outside Nigeria) to find out the necessary factors that can be considered as background factors. This also includes the questions about what corrective measures can be adopted in improving students’ academic performance. The response of these experts was qualitatively analyzed, and this informs the selected students’ background factors and the suggested recommendations used in the recommender system.

For building the recommender, we used algorithms from WEKA and multilayer perceptron. The algorithms from WEKA are in number of ten and are used for classification while the perceptron represents an artificial neural network.
The classification algorithms are split in two categories: 1) induction (PART, OneR, Decision table and JRip) and 2) decision tree algorithms (REPTree, J48, Random Tree, Decision stump and Random forest). The advantage of using the algorithms from Weka consists in the fact that they can be used for recommendations and give insights in the classification made. The structure of the recommender system is described Figure 1.

4. Results

Relationship between family background and students’ academic performance All the respondents acknowledged that family background factors have large roles to play in the performance of first year students in tertiary institutions. It is emphasized that a harmonious home environment gives the students stability and this is always evident in academic performance. It is concluded from the responses of the participants that a troubled mind will find it difficult to concentrate on academic, which may eventually lead to poor performance.

Family background and its associated influence performance is cultural-regional dependent The respondents affirmed that family background and its associated students academic performance is culturally as well as regionally dependent. Nigeria is a country with diverse cultures and is divided into three main regions. The most widely considered region in terms of education is South-West (Yoruba) because education is highly valued both for male and female children. Family members take the schooling of their wards so important such that they go to a great length in ensuring that every child acquires education at all cost. The Igbo are mixed in their inclination towards education. They are involved in trading more than education. But whether trading or education, every child both male and female has to be gainfully involved. The third region, the Hausas, are the less considered when it comes to education. The female children are not expected to attend school. They are to stay home and bear children while the males do the struggling. Even the male children’s schooling is optional. So, education in this region is sparsely distributed. The respondents in their views observed that a student from the region that sees education as a must will be more motivated than those from regions that care less.

Rating on some background factors identified in literature The respondents were asked to rate some background factors identified in literature on a four-Likert scale based on their perspective about whether the factor is very influential, influential, less influential or has no influence at all on students academic performance. The factors identified in literature are 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 at which 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 respondents considered all the identified factors to be very influential. One respondent considered Average family income and student’ education sponsor to be very influential to students academic performance.

**Corrective measures that can improve students' academic performance** The participants agreed that corrective measures can be employed in improving students’ academic performance. It is pointed out that the corrective action has to be the joint efforts of teachers, parents/guardians, the higher education administration and the students. Some corrective measures that were identified include encouragement and motivations, closely monitoring the students and additional or extramural tutorials.

**Machine Learning algorithms results** Once the variables were established from literature and interview with experts in the field, machine learning algorithms for building the recommender were applied to the sample of the students records. The performance of classification models was compared using classification accuracy, time taken to build the model and the confusion matrices. Random tree gives accuracy of 99.908% for 10-fold cross validation and 99.8205% for holdout method which outperform all other classifiers; compared to other models relative to the accuracy level, random tree was found to take lesser time. The detailed accuracy further revealed the performance of each classifier, based on the true positive rate (TP rate), false positive rate (FP rate), precision, recall and other measures. The benchmarks (accuracy level, confusion matrices and speed) used in measuring reveal that random tree perform better than all other algorithms. Therefore the reason for choosing random tree is given by the fact that it has a better performance as compared with the other algorithms such as random forest - the second as overall
performance. Using random tree algorithm, we built an intelligent recommender system for predicting the performance of the students in the first year. The system uses the following two categories of information: 1) family factors and 2) previous educational performances before entry in a tertiary educational institution. Characteristics such as age, gender, marital status of the parents, mother and father educational level, occupations of the parents, SSCE score (Secondary School Certificate Examination Grade) UME score (University Matriculation Examination), CGPA score from first year (Cumulative Grade Point Average) were considered.

**Implementation of the Intelligent Recommender System** The recommender system is made: 1) to give an estimation of the academic performance of a particular student and 2) to give recommendations to help increasing his/her academic performance. One typical use case is the following: the user fills in the student existing information (1) family factors and 2) previous educational performances) and the recommender will give an estimation for the grades of the student and also possible recommendations to be followed. The recommender has also the possibility to update the existing rules if needed through a back-end. The main constitutive parts are depicted in Figure 1.

**5. Discussions**

**In Relation to Other Studies** Different authors of (Adeyemo&Kuye, 2006), (Osofisan&Olamiti, 2009), (Oladokun, Adebanjo& Charles-Owaba, 2008), (Cortez& Silva, 2008), (Bhardwaj & Pal, 2011), (Ghaleb&Qeethara, 2010), (Akinola, Akinkunmi&Alo, 2012) have in various ways pointed out that background factors, entry qualification, students’ performance in SSCE or a particular subject, parental information among many others as influencing factors of students’ academic performance in tertiary institution. (Adeyemo&Kuye, 2006) specifically identified performance in SSCE as a major determinant, while (Oladokun, Adebanjo& Charles-Owaba, 2008), (Osofisan&Olamiti, 2009), (Ghaleb&Qeethara, 2010), and (Akinola, Akinkunmi&Alo, 2012) corroborate this claim. (Cortez& Silva, 2008) identified student’s age, parents’ age and education as influencing factors of students’ academic performance in tertiary institution.

The findings of this study agrees with the previous studies in identifying students family background factors and previous academic achievements as predictive attributes of students’ academic performance in tertiary institutions at the end of their first year. The outcome corroborated the postulation of Tinto in 1982 that though other factors contribute to students’ integration to tertiary institution; students’ academic past is also very significant in determining academic success. Furthermore, this research incorporated rules obtained from the optimal algorithm to develop a system that can predict student performance as well as recommend necessary interventions. The existing frameworks presented in (Vialardi, Bravo, Shafti&Ortigosa, 2009), (Shana &Venkatachalam, 2012) deals with students course registration and performance in a particular course, with different student populations, different input features, and different methodologies. Hence, it is difficult to make a direct comparison between the accuracy of the developed framework presented in this work and accuracy of the other developed frameworks.

**Recommendations and Suggestions for Further Studies** 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 achievements affect students’ first year academic performance in tertiary institutions; identifying the specific machine learning algorithm that best model the student academic performance, and designing a framework for a predictive system that can be developed from the generated 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 addresses the first year academic level, models can be built for other academic levels.

In most of the related literature, the decision trees proved its superiority over other techniques. It is also preferable regarding its good predictive ability. In addition, 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.
6. Conclusions

A multi-dimensional methodological approach was employed. An in-depth interview technique was employed in identifying appropriate students' background factors. The identified background factors are 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. Next, 7,500 students’ enrolment records representing 68% of the admissions into Babcock University between 2001 and 2010 were randomly selected. The institution of study was purposively selected. Ten classifications trees (Random forest, random tree, J48, Decision stump, REPTree, JRip, OneR, ZeroR, PART, and Decision table) and a multilayer perceptron (artificial neural network) learning algorithms were used in generating models from Waikato Environment for Knowledge Analysis (WEKA) based on the students’ enrolment records. The following benchmarks were used in comparing the generated models: accuracy level, confusion matrices and the model building’s speed. A framework of an Intelligent Recommender System (IRS) was then designed based on the outcome and the IRS was developed using Netbeans IDE and Java programming language with MySQL as the database server. From the models comparison, the study showed that Random Forest, Reptree, J48, JRip, PART, Decision Table and Multilayer Perceptron performed well with the lowest accuracy being 96.78%, while Decision stump, OneR and ZeroR were slightly lower in accuracy. Overall, random tree gave accuracy of 99.908% thereby outperforming other classifiers. Furthermore, other benchmarks employed revealed that random tree outperformed other algorithms. Therefore, random tree is adopted as optimal algorithm in the domain of this study. The rules generated from the optimal algorithm forms the back-end of the IRS and informed the prediction of the students’ academic performance.

In conclusion, relevant students’ background factors can be incorporated into designing a framework that will serve as a valuable tool for predicting performance and recommend necessary intervention strategies. Also, the superiority of decision trees over other techniques was affirmed. Although, the IRS framework designed was targeted at predicting students’ first year academic success in tertiary institutions based on background factors, it can be adapted in designing frameworks for other levels of learning. It can also serve as the building block for designing a generic system that can predict tertiary students’ overall performance from the point of entry to graduation.

References:


Hodge A. E. and Mellin E. A. (2010). First-Generation College Students: The Influence of Family on College Experience submitting to the Journal of College Counselling publication in progress;


