Predicting Students' Educational Performance

Folorunsho O. S
Washington University of Science and Technology
Vienna, VA, USA.

Ayinde A. Q
Northcentral University
Scottsdale, AZ, USA.

Yusuf A.S
New York Institute of Technology
Old Westbury, NY, USA

Abstract— **Data mining has drawn significant attention in numerous industries due to the big data stored in different repositories using structured and unstructured file formats. There is an imminent need to turn such massive data into meaningful information and knowledge. Appropriate mining techniques are applied to data stored and pulled from the different repositories to extract knowledge [5]**
Several data mining techniques require intensive and extensive use of algorithms, models, and methods to process and analyzed store data to read the patterns in the data set. During the process of pattern discoveries, the knowledge extracted is stored as a knowledge base environment that leadership, analyst, and researcher can use for data discoveries [1]
Students' grades will be predicted using an appropriate classifier to analyze educational data from the exam and record database of Osun State University, Osogbo, Osun State, Nigeria. This paper is partitioned into five sections namely pattern discovery, literature review, research methodology, results and discussions and the final step is conclusion.

Keywords— *Data mining; algorithms; knowledge base; discoveries, education*

1. INTRODUCTION

Over the years, the success rate of students enrolling in Electrical, and Electronics has dropped below average. The school management was looking at how to improve the course curriculum and ensure that the new practical works aid easy learning for the students. A descriptive analysis was conducted for the over five hundred students in the Electrical and Electronic department to determine which courses the students found difficult to pass.

The result of the analysis revealed that most students that passed EEE 201 course with B- failed the core Electrical and Electronic courses. The prerequisite courses for EEE 201 are MTH 101 and MTH 102. The student that passed MTH 101 and MTH 102 with an A grade tend to pass the core Electrical and Electronic course.

This paper predicts the success rate of students enrolling in Electrical and Electronic Engineering based on their MTH 101 and MTH 102 grades.

2. LITERATURE REVIEW

Researchers conducted a study to determine the performance of student in high school based on some science process skills. The performance was linked to school, student, and socio-economic factor [6]. A study on the performance of senior high school student in India was conducted. The student population was evenly partition into male and to establish an accurate method to measure cognition based on prognostic value, demographic, and personality contribute to the success of high education learning. Data mining selection had been based on cluster analysis method to analysis

growing population using incremental learning classifiers to analyze cluster groups and random data [4].

3. RESEARCH METHODOLOGY

In this research, Cross-Industry Standard Process for Data Mining (CRISP-DM) was adopted as a research methodology. During the data modeling phase of the CRISP-DM, an open-source software tool, WEKA, will be applied to the data stored in the development environment for this research. The Knowledge Flow Interface of WEKA will be used to implement the predictive strength of the classifiers in this research. The statement of the problem must be addressed at the business understanding phase. Data points and databases that house the data used in this research must be identified during the data understanding and preparation phase. The scope of students' educational results is limited to 100, 200, and 500 levels. During the Data Preprocessing Phase, student data collected Osun State University Exam and Record database are subjected to an extract, transform and load process. The 786 students' data were used in the research stored at the development environment, described by eleven data points (Gender, Age, MTH 101 Grade, MTH 102 Grade, Religion, EEE 201 Grade, student matriculation number, State of Origin, 200 level CGPA, 500level GPA, and Student Final Year Grade). The data has been preprocessed to ensure that they are clean and noise-free so that WEKA can analyze the data optimally.

The predictive model is built at the data modeling phase, where the algorithms are calibrated to work optimally using the data stored in the development environment. The student's final year grade is the target variable, and the predicting variables are part of the eleven abovementioned variables. The student final year grade, is categorized into five, as stated in Table 1. During the Modeling Phase, the decision tree and naïve Bayes algorithms were applied and calibrated to work incrementally.

4. RESULT AND DISCUSSIONS

Table 1
Performance Evaluation of the Classifiers

	First Class		Second Class Upper		Second Class Lower		Third Class	
	NB	DT	NB	DT	NB	DT	NB	DT
True Positive Rate	0.85	0.12	1.0	0.88	0.65	0.07	0.97	0.6
False Positive Rate	0.94	0.23	0.97	0.84	0.99	0.91	1.0	0.84
Precision	0.95	0.97	0.70	0.93	0.87	0.91	0.98	0.72

NB: Naïve Bayes DT: Decision Tree

During the data modeling phase, the data was partitioned into training and testing sets, 80 percent of the data was for training, and 20 percent was for testing. The key performance indices (KPI) used to measure the performance of the algorithms are True Positive Rate, False Positive Rate, and Precision. The performance evaluation of the algorithms is presented in Table 1. To ensure that the predictive model is working accurately, a cross-validation process was conducted to ensure that the algorithms are working optimally with high precision [2].

Based on the result in Table 1, the Naïve Bayes precision rate and True Positive Rate are higher than Decision Tree. Future research efforts should be directed at achieving a higher precision rate when the population size to be analyzed has increased. For scalability purposes, the algorithms must be configured to support incremental learning.

5. CONCLUSION

Both classifiers performed optimally, but Naïve Bayes' precision is higher than Decision Tree's. Naïve Bayes precision rate is an average of 80 percent, which means its False Positive rate is at 20 percent average. The school leadership can adopt this predictive model to determine the graduation rate of the students enrolling in Electrical and Electronics at their institution. The predicting variables that form part of the prerequisite and core courses to be taken by the students will be reviewed by the leadership and departmental chair on the best method to teach the core courses. This model will be used as a decision-making tool to improve the teaching and practical pattern. Also, it will be used to track the student's performance from the 100 level to the 500 level.

REFERENCES

[1] Ayinde, Abiodun., Adetunji, Abigail., Bello, Muriana., Odeniyi, Olubunmi. (2013). Performance Evaluation of Naive Bayes and Decision Stump Algorithms in Mining Students' Educational Data. IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 4, No 1, July 2013 ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784

[2] Baker R.S.J.D. (2003). "Data Mining for Education. Encyclopedia of Education (3rd edition)", B.MCGAW, PETERSON, P., BAKER Ed. Elsevier, Oxford, UK, 2009.Forman, G. 2003. An extensive Empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3, pp. 1289-1305, 2003.

[3] Hijazi, S. & Naqvi, R. (2006). "Factors Affecting Student's Performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006

[4] Khan, Z. (2005). "Scholastic Achievement of Higher Secondary Students in Science Stream",

[5] Usama, Fayyad., Gregory, Piatetsky-Shapiro., & Padhraic, Smyth. (1997). "From Data Mining to Knowledge Discovery in Databases", American Association for Artificial intelligence, pp. 37-54, 1997.

[6] Walters, Y., & Soyibo, K. (2001). "An Analysis of High School Students' Performance on Five Integrated Science Process Skills", Research in Science & Technical Education, Vol. 19, No. 2, 2001, pp.133 – 145.