A NOVEL APPROACH FOR PREDICTING STUDENT'S PERFORMANCE USING DATA MINING TECHNIQUE

P.P.V.Manikanta-Department of CSE, SRM University, Amaravati, Andhra Pradesh.
Koneru GnanDeep-Department of ECE, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh
Devendra Sai Mupparaju-Department of CSE, SRM University, Amaravati, Andhra Pradesh.

Abstract — In this paper, a mechanism in place to choose students who will do well academically based on objective factors used in the admissions process. admission based on certain pre-admission criteria (high school grade average, Scholastic Achievement Admission Test score, and General Aptitude Test score). The results also show that Scholastic Achievement Admission Test score is the pre-admission criterion that most accurately predicts future student performance. Therefore, this score should be assigned more weight in admissions systems.

INTRODUCTION

Currently in Malaysia, the lack of existing system to analyse and monitor the student progress and performance is not being addressed. There are two main reasons of why this is happening. First, the study on existing prediction methods is still insufficient to identify the most suitable methods for predicting the performance of students in Malaysian institutions. Second is due to the lack of investigations on the factors affecting students' achievements in particular courses within Malaysian context. Therefore, a systematically literature review on predicting student performance by using data mining techniques is proposed to improve students' achievements. The main objective of this paper is to provide an overview on the data mining techniques that have been used to predict students' performance. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a student's data. We could actually improve students' achievement and success more effectively in an efficient way using educational data mining techniques.

LITERATURE SURVEY

Usamah et al. (2013) stated that student's performance can be obtained by measuring the learning assessment and co-curriculum [2]. However, most of the studies mentioned about graduation being the measure of student's success. Generally, most of higher learning institutions in Malaysia used the final grades to evaluate students' performance. Final grades are based on

course structure, assessment mark, final exam score and also extracurricular activities [2]. The evaluation is important to maintain student's performances and the effectiveness of learning process. By analysing students' performance, a strategic program can be well planned during their period of studies in an institution [3]. Currently, there are many techniques being proposed to evaluate students' performance. Data mining is one of the most popular techniques to analyse students' performance.

METHODS

In order to build the predictive modelling, there are several tasks used, which are classification, regression and categorization. The most popular task to predict students' performance is classification. There are several algorithms under classification task that have been applied to predict students' performance. Among the algorithms used are Decision tree, Artificial Neural Networks, Naive Bayes, K-Nearest Neighbor and Support Vector Machine. Next, the specific application of data mining techniques grouped by algorithms in predicting student performance will be described in the next section. 3.2.1. Decision Tree Decision Tree is one of a popular technique for prediction. Most of researchers have used this technique because of its simplicity and comprehensibility to uncover small or large data structure and predict the value [8, 9, 13]. Romero et al. (2008) said that the decision tree models are easily understood because of their

reasoning process and can be directly converted into set of IF-THEN rules [22]. As shown in Table 2, there are approximately ten (10) papers that have used Decision Tree as their method to evaluate students' performance. Examples of previous studies using Decision Tree method are predicting drop out features of student's data for academic performance [8], predicting third semester performance of MCA students [32] and also predicting the suitable career for a student through their behavioural patterns [18]. The students' performance evaluation is based on features extracted from logged data in an education web-based system. The examples of dataset are student's final grades [23], final cumulative grade point average (CGPA) [3] and marks obtained in particular courses [22]. All these datasets were studied and analysed to find out the main attributes or factors that may affects the students' performance [28, 13]. Then, the suitable data mining algorithm will be investigated to predict students' performance [25]. Mayilvaganam and Kalpanadevi (2014) have compared the classification techniques for predicting students' performance in their study [12]. Meanwhile, Gray et al. (2014) investigated the accuracy of classification models to predict learners' progression in tertiary education [36]. 3.2.2. Neural Network Neural network is another popular technique used in educational data mining. The advantage of neural network is that it has the ability to detect all possible interactions between predictors variables [36]. Neural network could also do a complete detection without having any

doubt even in complex nonlinear relationship between dependent and independent variables [29]. Therefore, neural network technique is selected as one of the best prediction methods. Through the meta-analysis study, eight (8) papers have been published using Neural Network method. The papers present an Artificial Neural Network model to predict students' performance

This paper has reviewed previous studies on predicting students' performance with various analytical methods. Most of the researchers have used cumulative grade point average (CGPA) and internal assessment as data sets. While for prediction techniques, the classification method is frequently used in educational data mining area. Under the classification techniques, Neural Network and Decision Tree are the two methods highly used by the researchers for predicting students' performance.
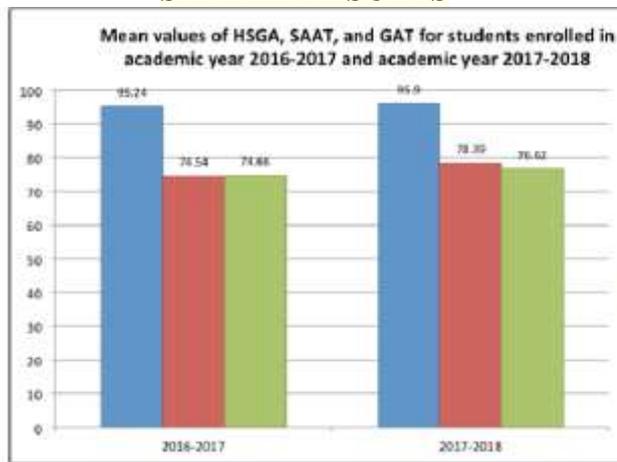
SAMPLE RESULTS

FIGURE 1 Mean values of HSGA, SAAT, and GAT for students en...
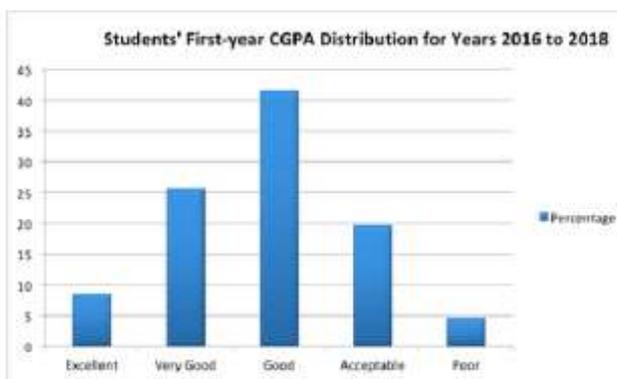

FIGURE 2 Students' first-year CGPA distribution for academic ...

CONCLUSION

In this, to help the educators and learners improving their learning and teaching process.

REFERENCES

[1] D. M. Anupama Kumar, Appraising the significance of self-regulated learning in higher education using neural networks, International Journal of Engineering Research and Development Volume 1 (Issue 1) (2012) 09–15.

[2] Z. K. Papamitsiou, V. Terzis, A. A. Economides, Temporal learning analytics for computer based testing, in: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, ACM, 2014, pp. 31–35.

[3] B. K. P. Conrad Tucker, A. Divinsky, Mining student-generated textual data in moocs and quantifying their effects on student performance and learning outcomes, in: 2014 ASEE Annual Conference, Indianapolis, Indiana, Indianapolis, Indiana, 2014, https://peer.asee.org/22840.

[4] C. Romero, S. Ventura, P. G. Espejo, C. Hervas, ´ Data mining algorithms to classify students, in: Educational Data Mining 2008, 2008.

[5] B. M. Bidgoli, D. Kashy, G. Kortemeier, W. Punch, Predicting student performance: An application of data mining methods with the educational web-based system Lon Capa, in: Proceedings of ASEE/IEEE frontiers in education conference, 2003.

[6] V. Oladokun, A. Adebanjo, O. Charles-Owaba, predicting students' academic performance using artificial neural network: A case study of an engineering course, The Pacific Journal of Science and Technology 9 (1) (2008) 72–79.

[7] V. Ramesh, P. Parkavi, K. Ramar, Predicting student performance: a statistical and data mining approach, International Journal of Computer Applications 63 (8) (2013) 35–39.

[8] A. Bogar´ın, C. Romero, R. Cerezo, M. Sanchez-Santill ´ an, ´ Clustering for improving educational process mining, in: Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, ACM, 2019