A Review on Predication of Students Performance Using Data Mining

Ms. A.V. Kadu and Dr. G.R. Bamnote

1M.E Scholar, Department of Computer Sc. & Engineering, PRMIT&R, Badnera
2Professor, Department of Computer Sc. & Engineering, PRMIT&R, Badnera

Abstract—The achievement of the good honours which is in the Undergraduate degrees which is the important in the context of Higher Education i.e (HE), and both for students and for the institutions that host them. We attempt to use data mining techniques to predict students’ outcomes based on early module performance and other student characteristics. In this paper, we work on whether the data mining can be used to highlight performance problems early in that and propose remedial actions. Many more are the some of the methods may also form the basis for recommender systems in that may guide students towards their module choices to increase the good outcomes their chances is necessary for good outcome. We use data collected through the admission process and through the students’ degrees. In this paper, we predict good honours outcomes based on data at admission also on the first year results of modules. To validate the proposed results modules, we evaluate the system whether the data relating to students with different characteristics from the different schools and the collages. Many universities have specific targets for students achieving good honor degrees.

Index Terms—Classification; Data Mining; Performance Prediction; Recommender System.

I. INTRODUCTION

Many universities have specific targets for students achieving good honor degrees. Achievement in terms of good honours is often reported in league tables. For example, the University has complete guide reports good honours as “The graduate’s student percentage of are achieving a first or second class honours degree”. The achievement of good honours in Undergraduate degrees is important in the context of Higher Education (HE), both for students and for the institutions that host them. The Guardian League Tables utilizes a value-added score that compares student’s individual degree results with their entry qualifications. It is also important for students to achieve a good degree as this can impact on their employment prospects. The explores the socio-demographic variables (gender, age, education, work status, ethnicity, and disability) and study environment. The databases in the most organization hold too much data and information that it become complicated and difficult to analyze those data manually. To overcome the human is a limitation on data are handling in the manual way, Data Mining is the suitable techniques to be used for conducting the analysis of data process. It examines to what extent these factors, i.e. enrolment data help us in pre identifying unsuccessful and successful students. It is therefore in the interest of both students and Universities to identify students at risk of not obtaining a good honours degree so that early intervention may improve their outcome. Universities have large numbers of module choices, and it is challenging for students to familiarize themselves with the entire make appropriate choices and possibilities. Appropriate choices may lead to higher student satisfaction and/or better performance. Universities are working in a very dynamic and powerfully viable environment today. They are gather large amount of volumes of data with reference to their students in electronic ways. However, they are data rich but information poor which results in unreliable decision making. The biggest challenge is the effective transformation of large volumes of data into knowledge to improve the quality of managerial decisions. Knowledge discovery in databases (KDD) refers to the discovery of interesting knowledge from the large volumes of data. It is therefore also important to predict overall outcome in specific modules and outcome, given particular module choices. It can be the basis for a recommender system to student’s aid in module selection. Recommender systems are currently considered as an advisable automated solution for assisting students with their choices.
In this we attempt to use techniques of data mining to predict students’ performance outcomes based on student characteristics and other early module performance. If our methods are successful for predicting the more general problem of student good and poor honours performance, we can then produce more granular predictions at the a module level and those would form the basis for a recommender system. We hope to uncover early indicators of poor performance that may be used to target remedial action for the concerned students. The aim to investigate the features which are available in that may be used for prediction, as well as the classifiers having a type of that may produce the best results. There is variety of Data Mining methods that available for data experiment such as clustering, association rule, and classification. This used the classification techniques to develop a model for predicting the Student good and poor Performance. The aims to do comparative analysis of Adaptive neuro fuzzy inference system (ANFIS). selected algorithms; the parameters that determine the Student good and poor Performance in all course at UniSZA are identified. The predicting the mark of students at the end of a university degree. To predict students’ performance which is good or poor at an early stage of the degree program helps universities not only to focus more on bright students. A number of theoretical models have been developed to explain what keeps students on a course. Identified variables that may impact attrition and have been included in theoretical models of dropout. These variables were classified into five constructs factors: motivation, technological support, academic integration, individual background, and social integration. The analysis is achieved by using historical data from the Data Warehouse of a specific University. The methods used are fairly general and can be used in any Higher Education (HE) institution. The results are highlight groups of students at considerable risk of obtaining poor outcomes.

II. REVIEW WORK

Students are accordingly evaluated through the three commonly known traffic signal lights (green, amber and red). The signal’s color indicates the risk level for the student relative to their counterparts. The algorithm used analyses both the student’s module performance and their help seeking. Another aspect often addressed is the prediction of student retention. Zahyah Alharbi, James Cornford, Liam Dolder and Beatriz De La Iglesia [13] combined social network analysis with data mining techniques to identify potentially nonsuccessful and successful students at the beginning of their studies. The aim of their study was to improve the classification accuracy of educational data when social network analysis is included behavior. Purposes, we applied the above series of steps on the second dataset. We utilized data mining techniques to predict students’ future grades by using two main attributes: the difficulty of each module and the level of a student’s knowledge before taking the module. When we began with by using only the attributes that related to student demographics. The assessment of each independent attribute using a Feature Selection ranking algorithm based on a Person chi-square test with significant. The data mining can be used to highlight performance problems early on and propose remedial actions. Some of the methods may also form the basis for recommender systems that may guide students towards their module choices to increase their chances of a good outcome. The primary goal was to predict students that are at a high risk of not achieving a good honours degree, to identify this as early as possible in year 1 so that interventions can be proposed.

In the last 15 years educational data mining emerged as a new application area for data mining, becoming well established with its own journal. It provided a survey of educational data mining extended their survey covering the latest development there are an increasing number of data mining applications in education, from enrollment management, academic performance, graduation, gifted education, web-based education, retention and other areas. This that reviews only research where the main focus is on study outcome, i.e. successful or unsuccessful course completion. Zlatko J. Kovacic [14] the number of predictor variables is not so large and selection of the subset of variables for further analysis which is the main purpose of applying feature selection to data is not required. It could be also used as a pre-processor for predictive data mining to rank predictors according to the strength of their relationship with dependent or outcome variable. The process no specific form of relationship, neither linear nor nonlinear is assumed and logistic regression the most important factors for student success and a profile of the typical unsuccessful and successful students is identified. The
explores the socio-demographic variables i.e. age, gender, ethnicity, education, work status, and disability and study environment. To improve the model, more attributes could be included to obtain prediction models with lower misclassification errors. The model in this case would not be a tool for pre-enrolment, i.e. early identification of ‘at-risk’ students. There are various previous studies conducted to predict the student good and poor honours performance using Data Mining techniques. The implementation of Data Mining techniques in educational field is giving additional insights in making better decisions and solutions for every issue arise. Azwa Abdul Aziz, Nor Hafieza Ismail and Fadhilah Ahmad [1]. The Waikato Environment for Knowledge Analysis. The Waikato Environment for Knowledge Analysis. The open source tool is used for classification model development. Classification is one of the most commonly applied technique in that predicts group exists in the data set. Classification technique is used by in the educational field to better understand students. The educational field that involves Data Mining techniques is rapidly increasing. Applying Data Mining techniques in an educational environment are known as Educational Data Mining that aims to discover hidden knowledge and patterns about students. The amount of data stored in an educational database at is increasing rapidly by the times.

III. PROPOSED WORK

In this paper we attempt to use data mining techniques to predict students’ outcomes based on early module performance and other student characteristics. If our methods are successful for predicting the more general problem of student good honours performance, we can then produce more granular predictions at the module level and those would form the basis for a recommender system. The analysis did not uncover specific problem modules as the poor performers seemed to do poorly across the board and on all modules in relation to their peers. The aim to investigate the available features that may be used for prediction, as well as the type of classifiers that may produce the best results. The analysis could also be used to influence admission policies given the characteristics of predicted poor performers. The hope to uncover early indicators of poor performance that may be used to target remedial action for the concerned students. That’s why we develop a predication system by which Students Performance Using Data Mining.

• Proposed diagram:

![Proposed Framework of Students Performance Prediction.](image)

Step1: Goal Identification
The goal is to develop the Students Performance prediction model using selected classification techniques and Data mining techniques which are Naïve Bayes, Decision Tree, and Rule Based. The accuracy values of the three models are compared and the model with the highest accuracy is selected as a Students Performance predictive.

Step2: Data collection and integration
In our system we are using student performance prediction. The data contain the information about students. All the students’ information is combined in a single file using the students’ matrix number as a primary key during the integration process. The six parameters were selected in this research; gender, race, hometown location, family income, university entry mode.

Step3: Data Preprocessing
The Data Pre-processing stage was performed to improve the quality of the data set by removing the missing or incomplete values in the data set. For the first batch data set, 69 records were removed from the 245 total data and only 176 records are ready for the
DM process later. After the pre-processing process applied on the second batch data set, 29 records are eliminated from the data set of 252 records which left only 223 clean records. The overall, the students’ data contains 98 missing values in various parameters from 497 records are ignored from the data set. The total numbers of records are reduced to 399.

Step4: Data Transformation
The selected parameters are categories into two; independent and dependent parameter. Independent parameter became input of the model used in methods equation or rules to predict the dependent parameter as an output. The independent parameters are student’s gender, race, family income, hometown location, and, university entry mode while a dependent parameter The Data Transformation stage will transform the numerical values.

IV. CONCLUSION
The primary goal of this work was to predict students i.e., that are at a high risk of not achieving a good honours degree, but more importantly, to identify this as early as possible in year 1 so that interventions can be proposed. The next stage will be to recommend strategies based on this and measure performance improvements.

The next stage will be to recommend strategies based on this and measure performance improvements. We have been able to achieve this goal with reasonable accuracy by using classification models to highlight the students that are predicted to be low achievers with high probability. It may also be possible to incorporate data on engagement (e.g. attendance monitoring, library loans) which is becoming available in the data warehouse to see its impact on prediction accuracy.

REFERENCES


