Student Grade Analysis and Prediction Using Machine Learning Algorithms

Syed Abubakar S[1] Dr. A. Rengarajan[2]
[1]Student [2]Professor
[1]Department of Master of Computer Application
[1,2]Jain University Bangalore, India

Abstract— The potential to predict the overall performance aim of the students is very essential to improve their training skills. It has come to be a useful information that can be used for various purposes. For example, a blueprint can be utilized for the development of a standard education. This paper proposes the utility of records mining techniques to predict the closing grades of students primarily based on their important data. In the experimental studies, three data mining strategies had been operating on two academic datasets related to arithmetic lesson and Portuguese language lesson. The results validated the success of data mining getting to know the methods during the prediction of student's performances.

Keywords: Data Mining, Student Performance Prediction, Classification

I. INTRODUCTION

Students overall performance is an important section in high level institutions. This is due to the fact that one of the standards for a higher rank institution is primarily based on its outstanding report of educational achievements [1]. There are a lot of definitions on students overall performance in the earlier literature. Students overall performance can be obtained via measuring the teaching evaluation and co-curriculum [2]. However, most of the research stated about graduation being the measure of student's success.

Recently, on-line tutoring have increased, and student's computerized data has come to big information size. This tends to achieve the policies and predictions about the students by way of processing academic information with data mining techniques. All kinds of records about the student's social class environment, studying environment, or course notes can be used for prediction, which have an effect on the success or failure of a student.

Generally, maximum tutoring institutions uses the final grades to consider the student's achievement. These final grades are based on the sequence procedure, evaluation marks, final examination rating and additionally extracurricular things to do [2].The assessment is necessary to keep student's performances and the value of tutoring process. By studying students' performance, an approach can be properly consider throughout their duration of learning in an institution [3]. Currently, there are many strategies being proposed to consider college students performance. Data mining is the most well-liked method to examine student's performance.

II. LITERATURE REVIEW

The idea draws the technique of ANN Network is used for the prediction of academic result of students. The grade points (CGPA) is used as the resulting criterion. The data needed for the task is accumulated from electrical department of Teknologi MARA University, Malaysia. Result of the models have been measured with the use of the Pearson's correlation of coefficient R and Mean Square Error (MSE). The effects from the learning about the confirmed subjects have robust have an impact on in the CGPA [4].

This usually make use of two kind of techniques for the normal prediction of the student's achievements over a huge data. These techniques are Learning Analytics and Predictive Analytics. Learning Analytics offers mainly on the information series and information pre-processing section the place the required facts for constructing the prediction model is accumulated from the schools, they have used MySQL server for storing the massive amount of data. Then in the information pre-processing step like data cleaning, data transformation etc. [5]

The make use of a selection tree model for predicting the educational overall achievements of students. The data required for setting up the blueprint is collected and records pre-processing is carried out the place where the continuous values are converted to discrete values and the null values are ignored. Then with the usage of CART algorithm to the records the decision tree prediction model is constructed. The students with negative overall performance are evicted. [6]

There are two most concentrating approaches one is Clustering and another is regression. The role of clustering is to provide the data pre-processing and the role of regression is for making the predictions. In clustering comparable data's are grouped in one, even though where the data classification is done and the regression model is used for doing the predictions. [7]

III. EXISTING SYSTEM

Education is a key aspect for a long-term progress in achieving. During the final decades, the Portuguese schooling stage has improved. Where, the information maintained the Portugal at Europe's tail stop due to its excessive scholar failure and losing out rates. For example, in the year 2006 the student's leaving schooling rate in Portugal was once 40% for 18 to 24 yr. olds, whilst the European Union common price was once simply 15%. In a particular failure of the core lessons of Mathematics and Portuguese (the native language is very serious, when considering that they grant proper education for the success in the other faculty lectures (e.g. physics or history).

IV. PROPOSED SYSTEM

In this, the process of analyzing the records from the two Portuguese secondary schools. Two kinds of sources have been used in the process, the one is the mark reviews and

another one is questionnaires. Since in the earlier records i.e. the grades and number of absences have been available, it is used to allow the group of school associated attributes, few demographic and social attributes (e.g. alcohol Intake, mother's qualification, student's age).The intention is to predict student's success and if required to know the key attributes that have an effect on academic role of success and failure.

The two main training classes Mathematics and Portuguese will be formed with the below three data modelling goals:

1) Binary classification (pass/fail);
2) Classification with 5 degrees (from I very good or excellent to V - insufficient);
3) Regression, with the output that lie between zero (0%) and twenty (100%).

There are three inputs for these approaches, (e.g. there may be the school assigned grades or may be not) and the four data modelling algorithms will be examined. Moreover, an explanatory evaluation will be carried out over the good models, in order to know the most applicable features.

V. DATA PREPROCESSING

In this process, two datasets have been used to predict student performances. Both datasets from the secondary school of two Portuguese schools. In the dataset process, two grading systems are used: five-level grading and binary grading systems. The dataset contains the attributes to analyze the student data that can predict the status of the individual student how well the student can perform in the examine process. [8]

Attribute	Description (Domain)
sex	student's sex (binary: female or male)
age	student's age (numeric: from 15 to 22)
school	student's school (binary: *Gabriel Pereira* or *Mousinho da Silveira*)
address	student's home address type (binary: urban or rural)
Pstatus	parent's cohabitation status (binary: living together or apart)
Medu	mother's education (numeric: from 0 to 4[a])
Mjob	mother's job (nominal[b])
Fedu	father's education (numeric: from 0 to 4[a])
Fjob	father's job (nominal[b])
guardian	student's guardian (nominal: mother, father or other)
famsize	family size (binary: ≤ 3 or > 3)
famrel	quality of family relationships (numeric: from 1 – very bad to 5 – excellent)
reason	reason to choose this school (nominal: close to home, school reputation, course preference or other)
traveltime	home to school travel time (numeric: 1 – < 15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour or 4 – > 1 hour).
studytime	weekly study time (numeric: 1 – < 2 hours, 2 – 2 to 5 hours, 3 – 5 to 10 hours or 4 – > 10 hours)
failures	number of past class failures (numeric: n if $1 \leq n < 3$, else 4)
schoolsup	extra educational school support (binary: yes or no)
famsup	family educational support (binary: yes or no)
activities	extra-curricular activities (binary: yes or no)
paidclass	extra paid classes (binary: yes or no)
internet	Internet access at home (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
romantic	with a romantic relationship (binary: yes or no)
freetime	free time after school (numeric: from 1 – very low to 5 – very high)
goout	going out with friends (numeric: from 1 – very low to 5 – very high)
Walc	weekend alcohol consumption (numeric: from 1 – very low to 5 – very high)
Dalc	workday alcohol consumption (numeric: from 1 – very low to 5 – very high)
health	current health status (numeric: from 1 – very bad to 5 – very good)
absences	number of school absences (numeric: from 0 to 93)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final grade (numeric: from 0 to 20)

VI. OBJECTIVES

The main objective is to help students to know their capabilities and also their weaknesses so that they can make use of those capabilities and work even more on them to get great opportunities and also to make them know their weaknesses so that they can work hard to achieve their expected result. School faculties can analyse the results of the student who are really worked hard to pass and the students who are going to fail. This analysis can assign the

faculties to concentrate on the low performing students so that the students can perform well in upcoming examination.

VII. CONCLUSION AND FUTURE SCOPE

The performance of students benefits to predict the lectures and the learners getting to know the improving skill for learning and teaching process if needed. The paper has been evaluated in the earlier research on predicting students overall performance with a range of analytical methods. Many researchers have used cumulative grade factor

common (CGPA) and internal evaluation as datasets. The paper points the utility of data mining methods to predict the final grades of students primarily on using the data. Three regularly used classification techniques (decision tree, random forest, and naive Bayes) had been used in the process. In conclusion, analysis on predicting student's performance has urged to similarly lookup further researches in our environment. In the future, unique function determination techniques can be used. In addition to different classification algorithms can additionally be utilized on the datasets.

REFERENCES

[1] M. of Education Malaysia, National higher education strategic plan (2015).
URL http://www.moe.gov.my/v/pelan-pembangunan-pendidikan-malaysia-2013-2025

[2] U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.

[3] Z. Ibrahim, D. Rusli, Predicting students academic performance: comparing artificial neural network, decision tree and linear regression, in: 21st Annual SAS Malaysia Forum, 5th September, 2007.

[4] Pauziah Mohd Arsad, Norlida Buniyamin and Jamalul-lail Ab Manan. "A Neural Network Students' Performance Prediction Model (NNSPPM)" IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), 26-27 November 2013.

[5] Midhun Mohan M G, Siju K Augustin and Dr. Kumari Roshni V S "A BigData Approach for Classification and Prediction of Student Result Using MapReduce", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10-12 December 2015

[6] Madhav S. Vyas and Reshma Gulwani. "Predicting Student's Performance using CART approach in Data Science", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017

[7] Alana M. de Morais, Joseana M. F. R. Araújo and Evandro B. Costa. "Monitoring Student Performance Using Data Clustering and Predictive Modelling", 978-1-4799-3922-0/14/$31.00 ©2014 IEEE

[8] Cortez P, Silva A. Using data mining to predict secondary school student performance. In: Brito A, Teixeira J,editors. Proceedings of 5th Annual Future Business Technology Conference. tpPorto: EUROSIS-ETI; 2018. pp. 5-12