Data Warehousing and Data Mining

Project Synopsis

# Team Members

* Ahana Agrawal-190911048
* Sanskar Mangla-190911029
* Abheesht Roy-190911066

# Title

Predicting Student Performance Using classification technique

# Problem Statement

Data mining is the knowledge of attaining useful information from a huge data set. There has been a rapid increasing of data in educational environment, educational data mining has emerged to develop methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings in which they learn.

# Abstract

In this project we aim to analyse, evaluate, and predict students’ performance based on their academic record using a concept of educational data mining. The data was collected from the student performances dataset available on kaggle.com. The data include marks secured by the students in high school from the United States. The pre-processing, processing, and experimenting will be conducted using RapidMiner tool. In this paper a classification task was used for the prediction. Classification is one of the data mining techniques which is used to classify each data items to a predefined set of class label. We can use models like the Random Forest Classifier for this classification problem.

# Introduction

In these days of the educational system of higher learning, any university database contains huge amounts of data of students with different courses and subjects upper and different years of intake. The information keeps increasing rapidly so we need to find out valuable information which can be used to determine the student success and the factors involved. Educational data mining is used to gain knowledge from it by using one of the statistical techniques, and data mining tools.

Data Mining is a well-suited technology that can be utilized by lecturers, students, managers, and other supporting staff. And it is a useful tool for decision making on their educational activities. It can be used to analyse the student’s academic data to promote and enhance the quality of the higher education system. The management can use the classification technique using a good tool to enhance the outcomes of the courses. The knowledge can be utilized for a better understanding of the student performance in the course faculty performance, and it can be used in the faculty and managerial decisions for the necessary action to be taken. Using such knowledge, the management can amend their policies and strategy for better decision making.

There have been previous works in this field, but we aim to increase the number of factors that are being taken and take factors that may indirectly affect a student’s performance. The result will help us determine how a student’s performance varies based on various circumstances in his life as well as his previous academic record. These findings can be used to implement some good policies. A lecturer can report the students with less CGPA. Courses with a higher number of failures lead to having pending courses at a final year students can be observed by the decision-makers or Head of Departmental. The lecturer can focus on such a category of students that needs more academic assistance.

# Methodology

We will be using RapidMinerwhich is a Data Science software that is developed by the RapidMiner Company and provides an integrated condition for data preparation, deep learning, predictive analysis, text mining, and machine learning. It is used for business and commercial services. RapidMiner is also used in research, education, rapid prototyping, training, application development and supports all steps of the machine learning techniques and process including data training, model validation, Result visualization, and optimization. We will be using operators like Read Excel, filter the decision tree and Random Forest algorithm.

Next, we will be using a classification technique for the prediction. Classification is one of the data mining techniques which is used to classify each data item to a predefined set of class labels. Each tuple that constitutes the training set is referred to as a class. These tuples can also be referred to as samples, data points or objects. We can use build-in models like Random Forest Classifier for this classification problem.

# References

P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April 2008, EUROSIS, ISBN 978-9077381-39-7.

Salisu, Sani & Usaman, Babandi. (2020). Data Mining: Predicting of Student Performance Using Classification Technique.