

PREDICTION OF STUDENT SUCCESS THAT ARE GOING TO ENROLL IN THE HIGHER TECHNICAL EDUCATION

SHIV KUMAR GUPTA¹, SONAL GUPTA² & RITU VIJAY³

¹Principal, Engineering in Jaipur National University, Jaipur, Rajasthan, India

²SAP Expert, Enterprise Resource Planning Consultant, Nashua, USA

³Associate Professor, Head of Department, Banasthali University, Rajasthan, India

ABSTRACT

Data warehousing is one of the important approaches for data integration and data pre-processing. This study explores the socio-demographic variables (age, gender, ethnicity, education, work status and disability) and study environment that may influence persistence or dropout of the students. It examines to what extent these factors, i.e. enrolment data help us in pre identifying successful and unsuccessful students. Based on a data mining techniques such as feature selection, classification trees and logistic regression the most important factors for student success and a profile of the typical successful and unsuccessful students are identified. The empirical results show the following, the most important factors separating successful from unsuccessful students are: ethnicity, course programme and course block; among classification tree growing methods Classification and Regression Tree (CART) was the most successful in growing the tree with an overall percentage of correct classification of 60.5%; both the risk estimated by the cross-validation and the gain diagram suggests that all trees, based only on enrolment data, are not quite good in separating successful from unsuccessful students, and the same conclusion was reached using the logistic regression. The case study is to build a data warehouse for a university student enrolment prediction data mining system. This data warehouse is able to generate summary reports as input data files for a data mining system to predict future student enrolment.

KEYWORDS: Data Warehousing, Information Integration, Star Schema, Dimensional Modelling, Data Mining, Study Outcome, Persistence, Dropout, Classification Tree, Logistic Regression+