Abstract

Student information systems hold a lot of information that can be mined for useful patterns. This work aims to build a prediction model to predict the result of students in ‘C’ Programming course by analyzing the factors that affect the performance of students. We applied feature selection techniques to select the most relevant academic and non-academic factors. The
model is implemented using various classification algorithms and it is found that Naïve Bayes classification model gives the highest accuracy of 82.4%. Decision tree based algorithm also showed considerable accuracy of 80.2%. The model was trained using 182 records from student dataset collected from the college with 20 attributes within the year 2008 to 2010. The model was validated using test records. It would predict the class label ‘Result’ as categorical value, Pass or Fail. Such a prediction model would help the faculty in early identification of ‘at risk’ students and thereby take timely and proactive measures to improve their performance.

Reference

- Jiawei Han and Micheline Kamber, 2009. Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers.
- Anne F. Maben, 2005, Chi-square test adapted from Statistics for the Social Sciences.

Index Terms

Computer Science  Data Management

Key words

Feature Selection  Classification  Prediction

Data mining