Research on a Predictive Learning Effect Support System for Programming Courses

[1]Chorng-Shiuh Koong, [2*]Yu-Ming Ding

Introduction

In the modern field of information technology, programming languages have become an indispensable compulsory course in many university computer science programs. However, for freshmen, learning a programming language is often a significant challenge (Amalina et al., 2019). It is not only a fundamental and practical skill but also a key aspect of future professional development. During the teaching process, it is evident that instructors often struggle to simultaneously address the needs of each student (Cheah, 2020), especially those whose performance is near the passing threshold or significantly below it. This challenge may result in diligent students who aspire to pass the course but ultimately fail to reach their academic goals. Hence, we strongly recognize the necessity for a system capable of early identification of students facing potential difficulties, serving as a remedial tool in the educational process.

Research Objectives

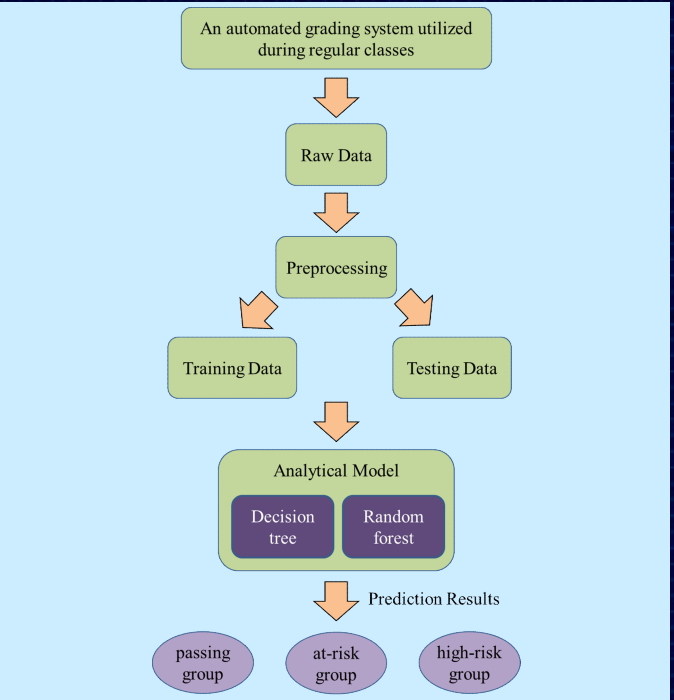
To address the challenges of students learning programming languages, we developed a "Learning Performance Prediction System." The system employs decision trees and random forest models adjusted for the dataset. This system, implemented a few weeks into the course, offers early predictions of learning outcomes by analyzing students' behavioral data during in-class activities and post-lecture exercises. Key points of our analysis include:

1. Predicting students' final semester grades, categorizing them into two groups: below 60 points (failing group) and 60 points or above (passing group).
2. Predicting students' final semester grades, categorizing them into three groups: above 60 points (passing group), 40-59 points (at-risk group), and below 40 points (high-risk group).

Methodology

The subjects of the experiment were sophomore students majoring in Computer Science at a Taiwanese university over the past three years, totaling 150 students. The data consisted of students' records of routine practice within the system. The data from the first and second years (2020 and 2021) were used as training data for the model, while the data from the third year (2022) served as testing data. The collected data included seven chapters with a total of 150 questions, comprising 23 analysis fields. Some descriptions of the data fields are as follows:

1. Submission deadlines for each chapter.
2. Answering rates for each chapter.
3. Assignment scores for each chapter.
4. Midterm exam scores.
5. Semester grades.



Results

The analysis results indicate that the decision tree achieved an accuracy of 84% for predicting the 2 groups classification, while the random forest reached an accuracy of 86%. For the 3 groups classification, both the decision tree and random forest exhibited an accuracy level of 84%. This predictive data serves as a valuable reference for future classes, enabling the identification and guidance of students in the danger zone and high-risk categories based on their regular performance data. Moreover, this predictive information is crucial for counseling students and enhancing teaching effectiveness.

	Two groups	Three groups
Decision tree	84%	84%
Random forest	86%	84%

Conclusion

In the face of the formidable challenge of learning a programming language, we have developed a "Learning Performance Prediction System" to empower educators in early identification and assistance of students. In system testing, the analysis results demonstrated exceptional accuracy. We believe that such a system will increasingly play a significant role in future teaching, continuously providing students with an improved learning environment.