

การวิเคราะห์การทำนายการลาออกกลางคันของนักศึกษา ระดับปริญญาตรี โดยใช้เทคนิควิธีการทำเหมืองข้อมูล

ชณิดาภา บุญประสม^{1*} และ จรรย์ แสนราช²

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อ 1) วิเคราะห์หาปัจจัยที่เกี่ยวข้องในการลาออกกลางคันของนักศึกษาระดับปริญญาตรี 2) สร้างเคราะห้โมเดลสำหรับการทำนายการออกกลางคันของนักศึกษาระดับปริญญาตรี และ 3) เปรียบเทียบประสิทธิภาพการจำแนกข้อมูลของโมเดลด้วยเทคนิควิธี Decision Tree, K-Nearest Neighbors, Naive Bayes โดยใช้ข้อมูลจากฐานข้อมูลงานทะเบียนของมหาวิทยาลัยราชภัฏอุบลราชธานี ของนักศึกษาระดับปริญญาตรี ระหว่างปีการศึกษา 2558-2560 มีจำนวน 11 แอททริบิวต์และ 13,729 ชุดข้อมูล เมื่อนำมาวิเคราะห์ค่าน้ำหนักของแอททริบิวต์ด้วยวิธีการ Information Theory พบว่า 1) มีปัจจัยที่เกี่ยวข้องในการลาออกกลางคันของนักศึกษาจำนวน 8 ปัจจัย 2) นำปัจจัยที่ได้มาทำการสร้างเป็นโมเดลทดสอบผลลัพธ์ด้วยวิธีการ 10-Fold Cross Validation และวัดประสิทธิภาพด้วยค่า Accuracy เพื่อหาวิธีการที่มีความถูกต้องมากที่สุด 3) ผลการเปรียบเทียบประสิทธิภาพการจำแนกข้อมูลพบว่าโมเดลที่สร้างด้วยเทคนิควิธี Naive Bayes มีประสิทธิภาพสูงสุดมีค่าเฉลี่ยความถูกต้อง 93.58 % มากกว่าเทคนิควิธี Decision Tree มีค่าเฉลี่ยความถูกต้อง 93.52 % และเทคนิควิธี K-Nearest Neighbors มีค่าเฉลี่ยความถูกต้อง 87.95 % และมีปัจจัยที่เกี่ยวข้องสูงสุด 5 อันดับ ได้แก่ การกู้ยืมกองทุนเพื่อการศึกษา, สาขาวิชา, เกรดเฉลี่ย, อาชีพของมารดา และอาชีพของบิดา

คำสำคัญ: การลาออกกลางคัน, การทำเหมืองข้อมูล, ต้นไม้ตัดสินใจ, เคเนียร์เรสเนเบอร์, การเรียนรู้ naïf เบย์

¹ อาจารย์ประจำสาขาวิศวกรรมซอฟต์แวร์ คณะวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏอุบลราชธานี

² ผู้ช่วยศาสตราจารย์ ภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

* ผู้มีพันธะประสานงาน โทร. 08-3932-0845 อีเมล: chanidapa.b@ubru.ac.th



Predictive Analytic for Student Dropout in Undergraduate Using Data Mining Technique

Chanidapa Boonprasom^{1*} and Charun Sanrach²

Abstract

The purposes of this research were 1) to analyze the factors that involved with the dropout of undergraduate students 2) to propose a model for predicting the dropout of undergraduate students 3) to compare the performance of 3 different classification techniques, including Decision Tree, K-Nearest Neighbors, and Naive algorithms. The data was collected from the undergraduate student's registration database of Ubon Ratchathani Rajabhat University during the academic years from 2015 to 2017. The dataset has 11 attributes and 13,729 records. The data were analyzed using the Information theory selection method. The results showed that 1) there are 8 factors that influencing student's dropout 2) Those factors were used to build models with the different techniques, Moreover, the cross-validation with 10 folds method was used to evaluate the best prediction accuracy of each technique. 3) the result suggested that the Naive Bayes model has the best performance among all techniques. It has the average accuracy of 93.58 %, which are higher than Decision tree and K-Nearest Neighbors which have the average accuracy of 93.52 % and 87.95 %, accordingly. The findings also indicated that students' decision to dropout was significantly influenced by the student loan, major of study, grade point average, and the occupation of their parents.

Keywords: Data Mining, Decision Tree, K-Nearest Neighbors, Naive Bayes, student dropout

¹ Lecturer, Department of Software Engineering, Faculty of Computer Science, Ubon Ratchathani Rajabhat University

² Assistant Professor, Department of Computer Studies, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok

* Corresponding Author Tel. 08-3932-0845 e-mail: chanidapa.b@ubru.ac.th