ชื่อเรื่อง การพยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูล
ผู้วิจัย ทิพย์หทัย ทองธรรมชาติ
ประธานที่ปรึกษา ดร.พรรณี สิทธิเดช
กรรมการที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.เกตุจันทร์ จำปาไชยศรี
ประเภทสารนิพนธ์ วิทยานิพนธ์ วท.ม., สาขาวิชาวิทยาการคอมพิวเตอร์, มหาวิทยาลัยนเรศวร, 2554
คำสำคัญ ผลสัมฤทธิ์ทางการเรียน การพยากรณ์ การทำเหมืองข้อมูล

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเทคนิคเหมืองข้อมูลมาประยุกต์ใช้ในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิต การดำเนินงานวิจัยได้ทำการศึกษา 1) ปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ และนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ 2) เทคนิคการคัดเลือกตัวแปรระหว่างเทคนิคการคัดเลือกชุดตัวแปรแบบอัตราส่วนเกนร่วมกับวิธีการค้นหาแบบจัดลำดับเปรียบเทียบกับเทคนิคค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันและสัมประสิทธิ์อีต้า และ 3) พยากรณ์ผลสัมฤทธิ์ทางการเรียนด้วยเทคนิคเหมืองข้อมูล 4 เทคนิค ได้แก่ เทคนิคเคเนียเรสเนเบอร์ (K-Nearest Neighbor) เทคนิคเนอีฟเบย์ (Naïve Bayes) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับ (Back-propagation Neural Network)

ข้อมูลที่ใช้ในงานวิจัยในครั้งนี้ คือ ข้อมูลภูมิหลัง และข้อมูลผลการเรียนรายวิชาที่อยู่ในแผนการเรียน ชั้นปีที่ 1 และ ปีที่ 2 ของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ และสาขาวิชาเทคโนโลยีสารสนเทศ ที่ศึกษาด้วยหลักสูตรปรับปรุง พ.ศ. 2548 คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร จำนวนทั้งสิ้น 489 คน จำแนกเป็นข้อมูลชุดฝึกสอน (Training Data) ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2551 – 2552 จำนวน 236 คน ข้อมูลชุดทดสอบ (Testing Data) เป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2553 จำนวน 126 คน และข้อมูลชุดตรวจสอบ (Validation Data) ซึ่งเป็นผู้สำเร็จการศึกษาประจำปีการศึกษา 2554 จำนวน 127 คน

จากการทดลองพบว่า ปัจจัยที่ส่งผลต่อผลสัมฤทธิ์ทางการเรียนของนิสิตสาขาวิชาวิทยาการคอมพิวเตอร์ และนิสิตสาขาวิชาเทคโนโลยีสารสนเทศ มีความแตกต่างกัน ชุดตัวแปรสำคัญที่คัดเลือกโดยพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน และสัมประสิทธิ์อีต้าเมื่อนำมาใช้จะให้ผลการพยากรณ์สูงกว่าชุดตัวแปรที่ได้จากแบบอัตราส่วนเกนร่วมกับวิธีการค้นหา

แบบจัดลำดับ และพบว่า ข้อมูลภูมิหลังไม่อยู่ในตัวแปรสำคัญที่คัดเลือกทั้ง 2 วิธี และในขั้นตอนการพยากรณ์นั้น ผลการเปรียบเทียบประสิทธิภาพพบว่า เทคนิคเนอีฟเบย์และเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับให้ค่าความถูกต้องในการพยากรณ์สูงใกล้เคียงกันและสูงกว่าเทคนิคเคเนียเรสเนเบอร์และเทคนิคต้นไม้ตัดสินใจ ผู้วิจัยจึงได้นำผลการพยากรณ์ด้วยเทคนิคเนอีฟเบย์ร่วมกับผลการพยากรณ์ด้วยเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับที่เป็นประเภทกลุ่มมาแปลงค่าเป็นค่าเกรดแล้วคำนวณค่าเฉลี่ยเพื่อเสนอผลการพยากรณ์เป็นผลสัมฤทธิ์ทางการเรียนแบบเกรดเฉลี่ย ซึ่งวิธีการนี้สามารถให้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ที่ต่ำกว่าการใช้เทคนิคเนอีฟเบย์ และเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับเพียงอย่างเดียว เมื่อได้รูปแบบแนวทางการพยากรณ์ที่ดีที่สุดในข้อมูลชุดฝึกสอนแล้ว จึงนำหลักการเดียวกันมาใช้ทดลองกับชุดข้อมูลทดสอบ และชุดตรวจสอบ ได้ผลการพยากรณ์ที่มีประสิทธิภาพที่ดีสอดคล้องกันกับชุดฝึกสอน

Title	LEARNING ACHIEVEMENT PREDICTION USING DATA MINING TECHNIQUES
Author	Tiphathai Thongthammachart
Advisor	Punnee Sittidech, Ph.D.
Co - Advisor	Assistant Professor Katechan Jampachaisri, Ph.D.
Academic Paper	Thesis M.S. in Computer Science, Naresuan University, 2011
Keywords	Learning Achievement, Prediction, Data Mining

ABSTRACT

This research is to apply data mining techniques to predict learning achievement of students. The study aims 1) to find major factors affecting academic achievement of students in the field of computer science and in the field of information technology 2) to compare the variables selection techniques between Gain Ratio Attribute Evaluators Technique with Ranker and Person Correlation Coefficient with Eta Coefficient, and 3) to apply four data mining techniques; the K-nearest Neighbor, Naïve Bayes, Decision Tree and Back-propagation Neural Network Techniques, to predict the students' achievement.

The data used in this research include the students' background information and studied grades in the first and second years of Bachelor's degree students who studied in Computer Science program and Information Technology program at Faculty of Science, Naresuan University, using the adjusted curriculum of the year 2005. The overall data of 489 records includes 236 records of training data, 126 records of testing data, and 127 records of validation data of students who graduated in the year 2008-2009, graduated in the year 2010, and graduated in the year 2011, respectively.

The experiment results show that the major factors affecting academic achievement of students in the field of computer science and in the field of information technology are different. The significant variables selected by the Person Correlation with Coefficient Eta Coefficient can be used to predict learning achievement with higher

accuracy than Gain Ratio Attribute Evaluators Techniques with Ranker. It is found that the students' background information is not included in the key factors in both techniques. For prediction experiment process among 4 data mining techniques, Naïve Bayes and Back-propagation Neural Network Techniques have comparable accuracies in prediction and higher perform than K-nearest Neighbor and Decision Tree. Finally, we then combined the better two methods' results by transforming learning achievement predicted class into numeric predicted learning achievement and then computed grade point average. This final result has the lowest RMSE (Root Mean Squared Error). The proposed method was applied to the test data set and the validation data set, which have consistent performances compare to the training set's results.