

ระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์
มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล
A Prediction System for Undergraduate Student Dropout at Faculty of
Science, Buriram Rajabhat University using Data Mining Techniques

นนทวัฒน์ ทวีชาติ^{1*} อรยา เพ็งประจัญ¹ วิไลรัตน์ ยาทองไชย¹ และ ชุศักดิ์ ยาทองไชย¹
Nontawat Thaweechat^{1*}, Oraya Pengprachan¹, Wilairat Yathongchai¹,
and Chusak Yathongchai¹

สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์¹

Information Technology Program, Faculty of Science at Buriram Rajabhat University¹

E-Mail: tongnam33@gmail.com*, beawzi246@gmail.com, wilairat.bru@gmail.com, chusak.bru@gmail.com

Received: February 3, 2020

Revised: April 6, 2020

Accepted: May 17, 2020

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) สร้างแบบจำลองการทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี และ 2) พัฒนาระบบการทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี วิเคราะห์ข้อมูลโดยใช้ข้อมูลของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ที่เข้าศึกษาระหว่างปี 2556 - 2561 จำนวน 3,650 ชุดข้อมูล โดยเลือกใช้เทคนิคการจำแนกประเภท ด้วยเทคนิคต้นไม้ตัดสินใจ และใช้อัลกอริทึม J48 เพื่อสร้างแบบจำลองการทำนายการพ้นสภาพของนักศึกษา และทำการทดสอบประสิทธิภาพของแบบจำลองด้วยวิธีการทดสอบแบบไขว้พบ แบบ 10 ส่วน โดยใช้โปรแกรมเวก้า จากนั้นนำแบบจำลองมาใช้ในการพัฒนาระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี ด้วยภาษาพีเอชพี เอชทีเอ็มแอล จาวาสคริปต์ ซีเอสเอส และระบบจัดการฐานข้อมูลมายเอสคิวแอล ในรูปแบบเว็บแอปพลิเคชัน

ผลการศึกษาพบว่า แบบจำลองการทำนายการพ้นสภาพของนักศึกษา มีปัจจัยที่เกี่ยวข้อง 11 ปัจจัย คือ สาขาวิชาที่เรียน เกรดสะสม 6 เทอม เกรดเฉลี่ยจากโรงเรียนเดิม หลักสูตรที่จบจากโรงเรียนเดิม ขนาดโรงเรียน และทุนกู้ยืม โดยมีค่าความถูกต้องร้อยละ 95.57 ค่าความแม่นยำร้อยละ 95.40 และค่าความระลึกร้อยละ 95.60 ซึ่งจากการใช้งานระบบการทำนายการพ้นสภาพของนักศึกษาโดยอาจารย์ที่ปรึกษาและนักศึกษาพบว่า มีระดับความพึงพอใจในระดับมาก (\bar{X} = 3.71, S.D. = 0.60)

คำสำคัญ : ระดับปริญญาตรี, การพ้นสภาพ, เทคนิคเหมืองข้อมูล, การจำแนกข้อมูล, ต้นไม้ตัดสินใจ

ABSTRACT

This research aims to 1) create a prediction model for students' dropout in higher education, and 2) develop a prediction system for students' dropout in higher education. To analyze 3,650 datasets of undergraduate students at the faculty of science, Buriram Rajabhat University who were admitted between 2013 and 2018. To use the Weka software with the decision tree classifier and J48 algorithm to create the prediction model for students' dropout and evaluate the model using a 10-fold cross validation method. The prediction model had incorporated in the prediction system development for students' dropout in higher education. The development tools were PHP, HTML, JavaScript, CSS, and MySQL DBMS. The prediction system had developed in the form of a Web application.

The study results found that the prediction model for students' dropout had 11 associative factors: major, 6 GPAXs (6 academic terms), GPA from high school, major in high school, school size, and scholarship loan. The model had the values in 95.57% of accuracy, 94.50% of precision, and 95.60% of recall. The use of the prediction system by the advisors and the students found that in a high level of satisfaction (\bar{X} = 3.68, S.D. = 0.58).

Key Words : Undergraduate, Student drop out, Data mining technique, Classification, Decision tree

บทนำ

คุณภาพเป็นปัจจัยหลักที่สำคัญในการศึกษาระดับอุดมศึกษา การได้มาซึ่งคุณภาพต้องได้รับการวางแผน ติดตาม และควบคุมในแต่ละกระบวนการการศึกษาทั้งนี้เพื่อการปรับปรุงประสิทธิภาพของผู้เรียน โดยสิ่งที่บ่งชี้ถึงจุดอ่อนของคุณภาพการศึกษาก็คือ จำนวนการพ้นสภาพของผู้เรียนที่เพิ่มสูงขึ้น จึงมีงานวิจัยต่าง ๆ ให้ความสนใจในการทำนายจำนวนผู้เรียนที่จะพ้นสภาพ และปัจจัยที่มีผลต่อสถานการณ์การพ้นสภาพต้องใช้กระบวนการที่มีประสิทธิภาพ (Dekker *et. al.*, 2009) การใช้เทคนิคการทำเหมืองข้อมูลเพื่อทำนายการพ้นสภาพและระบุปัจจัยความสำเร็จของนักศึกษาวิศวกรรมไฟฟ้า กอตเซนทิส (Kotsiantis, 2009) รวมถึงการวิเคราะห์ปัญหาของนักศึกษาที่เลิกเรียนระดับอุดมศึกษา

โดยใช้วิธีการเหมืองข้อมูล (Jadrić et. al., 2010) ทั้งนี้เพื่อเพิ่มประสิทธิภาพในการบริหารจัดการด้านการศึกษา

คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ เป็นองค์กรแห่งการเรียนรู้ที่หลากหลายทางด้านวิทยาศาสตร์และเทคโนโลยีภายใต้พันธกิจในการผลิตบัณฑิตด้านวิทยาศาสตร์และเทคโนโลยีที่มีความรู้ มีคุณธรรมและมีคุณภาพตามมาตรฐานสากล มีการจัดการเรียนการสอนประกอบด้วย 2 ภาควิชา 12 สาขาวิชา คือ เคมี คณิตศาสตร์ เทคโนโลยีสารสนเทศ ชีววิทยา ภูมิศาสตร์และภูมิสารสนเทศ วิทยาการคอมพิวเตอร์ วิทยาศาสตร์การกีฬา วิทยาศาสตร์การอาหาร วิทยาศาสตร์สิ่งแวดล้อม วิทยาศาสตร์สิ่งแวดล้อม สถิติประยุกต์ และสาธารณสุขชุมชน (ประวัติและที่มาของคณะวิทยาศาสตร์, 2556) ช่วงเวลาที่ผ่านมาคณะวิทยาศาสตร์มีอัตราการพ้นสภาพของนักศึกษาในระดับที่เป็นปัญหา ซึ่งมาจากหลายปัจจัย เช่น ปัญหาครอบครัว ปัญหาทางเศรษฐกิจ ปัญหาการเรียน และปัจจัยที่มาจากตัวนักศึกษาเอง เป็นต้น ซึ่งปัญหาการพ้นสภาพหรือการออกกลางคันของนักศึกษาถือว่าเป็นปัญหาที่สำคัญของสถาบันการศึกษาซึ่งจะส่งผลกระทบต่อประสิทธิภาพในการจัดการศึกษา การบริหารจัดการงบประมาณขององค์กร นั่นคือ นักศึกษาที่ไม่สำเร็จการศึกษาตามหลักสูตรที่กำหนดไว้ ถือว่าเป็นความสูญเสียโอกาสในการผลิตบัณฑิตและการสูญเสียเศรษฐกิจครอบครัว มหาวิทยาลัย และประเทศชาติ หรือที่เรียกว่า “เกิดความสูญเสียในการลงทุนเพื่อการศึกษา” กล่าวคือ คณะและมหาวิทยาลัยย่อมเสียเวลาในการบริหารจัดการ เสียทรัพยากรในการลงทุนและเสียโอกาสในการสร้างคน ส่วนนักศึกษาเสียเวลา เสียค่าใช้จ่ายและประการสำคัญ คือ เสียขวัญและกำลังใจในการถอยหลังเพื่อไปเริ่มต้นใหม่ รวมทั้งภาครัฐก็จำเป็นต้องจัดสรรเงินงบประมาณเพื่อสนับสนุนการอุดมศึกษาเป็นจำนวนมากเช่นกัน ดังนั้น การพ้นสภาพการเป็นนักศึกษาจึงเป็นประเด็นปัญหาที่คณะและมหาวิทยาลัยต้องหาทางแก้ไข

จากปัญหาดังกล่าว งานวิจัยนี้จึงมีแนวคิดการนำเหมืองข้อมูล (Data mining) ด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree) มาช่วยสร้างกฎเพื่อพัฒนาระบบทำนายการพ้นสภาพของนักศึกษา คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ โดยนักศึกษาหรืออาจารย์ที่ปรึกษาสามารถใช้ระบบเพื่อการทำนายการพ้นสภาพเพื่อวางแผนในการจัดการศึกษา รวมถึงการติดตามดูแลเพื่อลดจำนวนการพ้นสภาพของนักศึกษาต่อไป

วัตถุประสงค์งานวิจัย

1. เพื่อสร้างแบบจำลองการทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล
2. เพื่อพัฒนาระบบทำนายการพ้นสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยเทคนิคการทำเหมืองข้อมูล

เอกสารและงานวิจัยที่เกี่ยวข้อง

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการในการตีความหมายข้อมูล จากการจัดเก็บข้อมูลอย่างง่ายมาสู่การจัดเก็บในฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้ จนถึงการค้าพบสารสนเทศที่ซ่อนอยู่ในข้อมูล เทคนิคการจำแนกประเภท (Classification technique) ถูกใช้ในการทำเหมืองข้อมูล ซึ่งใช้ชุดของตัวอย่างก่อนจัดประเภทเพื่อพัฒนาแบบจำลองที่สามารถจำแนกข้อมูลที่บันทึกไว้ขนาดใหญ่ได้ วิธีการนี้มักใช้ต้นไม้ตัดสินใจ (Decision tree) หรืออัลกอริทึมการจำแนกประเภทตามโครงข่ายประสาทเทียม (Neural network-based classification algorithms) กระบวนการจำแนกข้อมูลเกี่ยวข้องกับการเรียนรู้และการจำแนกในการเรียนรู้ข้อมูลการฝึกอบรม (Training data) จะถูกวิเคราะห์โดยอัลกอริทึมการจำแนก ส่วนการจำแนกประเภทของข้อมูลการทดสอบ (Test data) ถูกใช้เพื่อประเมินความถูกต้องของกฎการจำแนกประเภท หากความถูกต้องเป็นที่ยอมรับกฎสามารถนำไปใช้ข้อมูลใหม่ได้

ต้นไม้ตัดสินใจเป็นวิธีการในการจัดรูปแบบการจัดหมวดหมู่ในการจำแนก โดยแสดงภาพขั้นตอนที่นำไปสู่การจำแนกประเภท ต้นไม้ตัดสินใจเป็นวิธีการแสดงข้อมูลจากอัลกอริทึมการเรียนรู้ของเครื่อง (Machine learning algorithm) ที่รวดเร็วเมื่อเทียบกับเทคนิคอื่นและมีประสิทธิภาพในการแสดงโครงสร้างของข้อมูล โดยอัลกอริทึม J48 จะให้หลายตัวเลือกที่เกี่ยวข้องกับการตัดแต่งต้นไม้ซึ่งสร้างผลลัพธ์ที่กระชับและง่ายขึ้น ผลลัพธ์ที่ได้มีความยืดหยุ่นสามารถนำไปใช้โดยแปลงเป็นกฎได้ ให้เทคนิคการจำแนกประเภท (Classification technique) มักใช้ในการทำเหมืองข้อมูลซึ่งใช้ชุดของตัวอย่างก่อนจัดประเภทเพื่อพัฒนาแบบจำลองที่สามารถจำแนกข้อมูลที่บันทึกไว้ขนาดใหญ่ได้ วิธีการนี้มักใช้ต้นไม้ตัดสินใจ (Decision tree) หรืออัลกอริทึมการจำแนกประเภทตามโครงข่ายประสาทเทียม (Neural network-based classification algorithms) กระบวนการจำแนกข้อมูลเกี่ยวข้องกับการเรียนรู้และการจำแนก ในการเรียนรู้ข้อมูลการฝึกอบรม (Training data) จะถูกวิเคราะห์โดยอัลกอริทึมการจำแนก ส่วนการจำแนกประเภทของข้อมูลการทดสอบ (Test data) ถูกใช้เพื่อประเมินความถูกต้องของกฎการจำแนกประเภท หากความถูกต้องเป็นที่ยอมรับกฎสามารถนำไปใช้ข้อมูลใหม่ได้

ต้นไม้ตัดสินใจเป็นวิธีการในการจัดรูปแบบการจัดหมวดหมู่ในการจำแนก โดยแสดงภาพขั้นตอนที่นำไปสู่การจำแนกประเภท ต้นไม้ตัดสินใจเป็นวิธีการแสดงข้อมูลจากอัลกอริทึมการเรียนรู้ของเครื่อง (Machine learning algorithm) ที่รวดเร็วเมื่อเทียบกับเทคนิคอื่นและมีประสิทธิภาพในการแสดงโครงสร้างของข้อมูล โดยอัลกอริทึม J48 จะให้หลายตัวเลือกที่เกี่ยวข้องกับการตัดแต่งต้นไม้ซึ่งสร้างผลลัพธ์ที่กระชับและง่ายขึ้น ผลลัพธ์ที่ได้มีความยืดหยุ่นสามารถนำไปใช้โดยแปลงเป็นกฎได้ ให้ความแม่นยำสูง และนำไปประยุกต์ใช้กับการค้นหาข้อมูลด้วยภาษาสอบถามเชิงโครงสร้าง (SQL) ได้ (Witten and Frank, 2005) เทคนิคเหมืองข้อมูลได้ถูกนำไปใช้เพื่อพัฒนาคุณภาพการศึกษาของการศึกษาระดับอุดมศึกษาหลากหลายด้าน รวมถึงการศึกษาที่เกี่ยวข้องกับการแก้ปัญหาการฟื้นฟูสภาพของนักศึกษา

ชนิดาภา บุญประสม และจรัญ แสนราช (2561) ทำการวิจัยเพื่อ 1) วิเคราะห์หาปัจจัยที่เกี่ยวข้องในการลาออกกลางคันของนักศึกษาระดับปริญญาตรี 2) สังเคราะห์โมเดลสำหรับการทำนายการออกกลางคันของนักศึกษาระดับปริญญาตรี และ 3) เปรียบเทียบ ประสิทธิภาพการจำแนกข้อมูลของโมเดลด้วยเทคนิควิธี Decision tree, K-Nearest neighbors, Naive Bayes โดยใช้ข้อมูลจากฐานข้อมูลงานทะเบียนของมหาวิทยาลัยราชภัฏอุบลราชธานีของนักศึกษาระดับปริญญาตรีระหว่างปีการศึกษา 2558-2560 มีจำนวน 11 แอททริบิวต์และ 13,729 ชุดข้อมูล เมื่อนำมาวิเคราะห์ค่าน้ำหนักของแอททริบิวต์ ด้วยวิธีการ Information theory พบว่า 1) มีปัจจัยที่เกี่ยวข้องในการลาออกกลางคันของนักศึกษาจำนวน 8 ปัจจัย 2) นำปัจจัยที่ได้มาทำการสร้างเป็นโมเดลทดสอบผลลัพธ์ด้วยวิธีการ 10-fold cross validation และวัดประสิทธิภาพด้วย ค่า 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 เกรตเฉลี่ย อาชีพของมารดา และอาชีพของบิดา

จิระนันต์ เจริญรัตน์ และวีระศักดิ์ เจริญรัตน์ (2559) ได้ทำการวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาที่มีผลการเรียนปกติที่มีผลการเรียนของเกรดเฉลี่ยสะสมมากกว่า 2.00 โดยใช้ต้นไม้ตัดสินใจ โดยใช้ข้อมูลนักศึกษาระดับปริญญาตรี คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏสกลนคร จำนวน 3,385 ข้อมูล พบว่า ปัจจัยที่มีความสำคัญต่อการพัฒนาของนักศึกษาที่มีผลการเรียนปกติแบ่งเป็น 4 กลุ่ม คือ กลุ่ม Best ปัจจัยคือ วุฒิมัธยมศึกษาเดิม กลุ่ม Excellent ปัจจัยคือ อาชีพมารดา และสาขาวิชาที่เรียน กลุ่ม Good ปัจจัยคือ ทุนกู้ยืมเพื่อการศึกษา สถานภาพของครอบครัว รายได้บิดามารดา ซึ่งผลที่ได้สามารถนำไปพัฒนาระบบพยากรณ์การพัฒนาของนักศึกษา เพื่อเป็นแนวทางในการบริหารจัดการให้อาจารย์ที่ปรึกษาได้วางแผนการเรียนของนักศึกษาต่อไป

พฤทธิพงศ์ เพ็งศิริ และคณะ (2556) ได้ทำการวิเคราะห์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษานักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ซึ่งผลลัพธ์ที่ได้นั้นจะส่งผลทำให้เอื้อประโยชน์กับแนวทางในการปรับปรุงบริหารจัดการและส่งเสริมการวางแผนของหลักสูตรการศึกษาที่เกี่ยวข้อง

นอกจากนี้ยังมีงานวิจัยที่นำเทคนิคเหมืองข้อมูลมาใช้ในการจัดกลุ่มนักศึกษาชั้นปีที่ 1 ของมหาวิทยาลัยในประเทศเบลเยียม ที่ทำการรวบรวมในเดือนพฤศจิกายน ปีการศึกษา 2003 -2004 โดยแบ่งนักศึกษาออกเป็น 3 กลุ่ม คือ กลุ่มที่มีความเสี่ยงน้อย (Low-risk) กลุ่มที่มีความเสี่ยงปานกลาง (Medium-risk) และกลุ่มที่มีความเสี่ยงสูง (High-risk) ในการพัฒนา เพื่อนำผลการวิจัยที่ได้ไปสู่การวางแผนการลดจำนวนนักศึกษาที่จะพัฒนา (Superby *et al.*, 2006) รวมถึงการประยุกต์ใช้ต้นไม้ตัดสินใจในการทำนายผลการเรียนของนักศึกษาที่ลงทะเบียนเรียนรายวิชาการเขียนโปรแกรมภาษาซีของมหาวิทยาลัยYarmouk ประเทศจอร์แดน โดยใช้อัลกอริทึม ID3 C4.5 และ Naïve Bayes ซึ่งได้ผล

การทำนายที่ดีกว่าการใช้อัลกอริทึมอื่น (Al-Radaideh, 2006) จากการวิจัยดังกล่าวจะเห็นได้ว่าเทคนิคการทำเหมืองข้อมูลสามารถนำมาใช้กับฐานข้อมูลด้านการศึกษาของมหาวิทยาลัยเพื่อค้นหารูปแบบที่ซ่อนอยู่และความสัมพันธ์ระหว่างข้อมูลที่เป็นประโยชน์ช่วยในการตัดสินใจได้ดียิ่งขึ้น

วิธีการดำเนินการวิจัย

1. กรอบการศึกษา

การวิจัยนี้มีกรอบในการศึกษา 5 ขั้นตอน คือ การเตรียมข้อมูลก่อนประมวลผล อัลกอริทึมสำหรับการจำแนกประเภท การวัดประสิทธิภาพการจำแนกประเภท การสร้างแบบจำลอง และการพัฒนาและประเมินผลระบบ

- การเตรียมข้อมูลก่อนประมวลผล (Data pre-processing) เป็นขั้นตอนแรกก่อนการทำเหมืองข้อมูลเนื่องจากข้อมูลในฐานข้อมูลที่จะนำมาวิเคราะห์อาจมีข้อมูลไม่สมบูรณ์ (Incomplete data) ข้อมูลมีค่าผิดพลาด (Error) หรือมีค่าผิดปกติ (Outliers) ตลอดจนข้อมูลไม่สอดคล้อง (Inconsistent data) ที่ต้องตรวจสอบความถูกต้องในเรื่องของรูปแบบข้อมูล ค่าของข้อมูลก่อนการประมวลผล โดยงานวิจัยนี้ได้ทำการคัดข้อมูลที่เป็นส่วนรบกวน หรือข้อมูลที่ไม่เกี่ยวข้องออกไป (Data cleaning) ทำการคัดเลือกข้อมูลและแปลงข้อมูลให้เหมาะกับการทำเหมืองข้อมูล (Data selection and transformation)

- อัลกอริทึมสำหรับการจำแนกประเภท (Classifier algorithm) ทำการจำแนกประเภทด้วยอัลกอริทึม J48 เพื่อให้ได้ผลลัพธ์ที่มีความยืดหยุ่นสามารถนำไปใช้โดยแปลงเป็นกฎได้ ให้ความแม่นยำสูง และนำไปประยุกต์ใช้กับการค้นหาข้อมูลด้วยภาษาสอบถามเชิงโครงสร้าง (SQL) ได้

- การวัดประสิทธิภาพการจำแนกประเภท (Evaluation classifier) โดยใช้วิธีการตรวจสอบไขว้ (K-fold cross-validation) เพื่อประเมินผลแบบจำลองจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure)

- การสร้างแบบจำลอง (Academic DSS model) เมื่อทำการสร้างแบบจำลองทำให้ได้กฎจากอัลกอริทึมการจำแนกประเภทเพื่อนำไปสู่การพัฒนาระบบทำนายการพัฒนาระบบของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ที่นำมาวางแผนการจัดการศึกษาของคณะและมหาวิทยาลัย รวมถึงการติดตามดูแลนักศึกษาของอาจารย์ที่ปรึกษา

การพัฒนาและประเมินผลระบบ (Development and evaluation system) โดยเลือกใช้ภาษา PHP HTML JavaScript และระบบจัดการฐานข้อมูล MySQL ในรูปแบบเว็บแอปพลิเคชัน (Web application)

1. การรวบรวมข้อมูลเพื่อการประมวลผล

คลังข้อมูลของมหาวิทยาลัยราชภัฏบุรีรัมย์จะเก็บฐานข้อมูลที่เกี่ยวข้องกับการทำงานของมหาวิทยาลัย เช่น ฐานข้อมูลการศึกษา ฐานข้อมูลการเงิน ฐานข้อมูลการประกันคุณภาพ เป็นต้น เพื่อ

นำมาใช้ประโยชน์ในการบริหารงานของมหาวิทยาลัย โดยข้อมูลที่น่ามาใช้ในการวิจัยครั้งนี้คือ ข้อมูลจากฐานข้อมูลการศึกษา ที่จัดเก็บข้อมูลของนักศึกษาตั้งแต่ปีการศึกษา 2556 - 2559 โดยใช้ข้อมูลตัวอย่างจากนักศึกษาคณะวิทยาศาสตร์ ด้วยมีอัตราการพ้นสภาพมากที่สุดในมหาวิทยาลัยราชภัฏบุรีรัมย์ จากนั้นนำข้อมูลจากหลายตารางมารวมกันผ่านกระบวนการเตรียมข้อมูลให้ได้ข้อมูลจำนวน 3,604 ระเบียบ

2. การคัดเลือกและแปลงข้อมูล (Data selection and transformation)

โดยการวิจัยนี้ได้กำหนดสมมติฐานของปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาคือ ปัจจัยที่เกี่ยวข้องกับผู้เรียน คือ สถานภาพครอบครัว อาชีพ/รายได้ของบิดาและมารดา ปัจจัยการเรียนรู้ในระดับมัธยมศึกษา คือ ผลการเรียนรู้เฉลี่ย แผนการเรียนที่จบ ขนาดโรงเรียน และปัจจัยระหว่างเรียนในมหาวิทยาลัย คือ สาขาวิชา เกเรตเฉลี่ยแต่ละภาคเรียน จำนวน 6 ภาคเรียน สถานการณ์กู้ยืม รวม 16 ปัจจัย จากนั้นทำการวิเคราะห์ปัจจัยที่เกี่ยวข้องกับการพ้นสภาพของนักศึกษาระดับปริญญาตรี โดยการลดมิติข้อมูล (Attribute selection) การคำนวณค่าน้ำหนักของเอทริบิวต์ พบว่า มีจำนวน 11 ปัจจัยที่เกี่ยวข้องกับการพ้นสภาพของนักศึกษา รายละเอียดของแต่ละปัจจัยดังแสดงในตารางที่ 1

ตารางที่ 1 ตัวแปรที่เกี่ยวข้องกับนักศึกษา

ตัวแปร	คำอธิบาย	ค่าที่เป็นไปได้
Program	สาขาวิชาที่ศึกษาในคณะวิทยาศาสตร์	{210, 249, 267, 417, 416, 415, 232, 231, 230, 240, 266, 456, 268, 243, 265}
GPA1-GPA6	เกรดเฉลี่ยในภาคเรียนที่ 1 – 6 (ในปีการศึกษา 2556 - 2559)	{Weak, Medium, Good, Best} → Weak =GPA< 1.6 Medium=GPA 1.6-1.99 Good=GPA 2.0-2.5 Best=GPA>2.5
SchoolGPAX	เกรดเฉลี่ยจากโรงเรียนมัธยม	ตัวเลข
SchoolProgram	แผนการเรียนที่ศึกษาในโรงเรียนมัธยม	{1, 2, 3} → 1 = วิทยาศาสตร์/คณิตศาสตร์ 2 = ศิลปคำนวณ 3 = อื่น ๆ
SchoolSize	ขนาดโรงเรียน	{Small,Medium,Large}
Loan	สถานะการกู้ยืมเพื่อการศึกษา	{Yes, No}→ Yes = กองทุนกู้ยืม No = ไม่กู้ยืม
DropOut	สถานะการพ้นสภาพ	{Yes, No}

3. การวิเคราะห์ข้อมูล

การวิจัยนี้ใช้เทคนิคต้นไม้ตัดสินใจเพื่อสร้างแบบจำลอง (Model) ดังนี้

1.1 การสร้างและทดสอบแบบจำลอง

งานวิจัยนี้ใช้โปรแกรม WEKA 3.7.9 ในการสร้างตัวแบบการวิเคราะห์ปัจจัยและทดสอบตัวแบบด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree technique) โดยใช้วิธีการจำแนกประเภท (Classification) เลือกใช้ J48 คือ อัลกอริทึม C4.5 เพื่อใช้ในการเรียนรู้ ซึ่งสามารถสร้างผลลัพธ์ที่กระชับและง่ายขึ้น มีความยืดหยุ่นสามารถนำไปใช้โดยแปลงเป็นกฎได้ และนำไปประยุกต์ใช้กับการค้นหาข้อมูลด้วยภาษาสอบถามเชิงโครงสร้าง จากนั้นสร้างแบบจำลองต้นไม้เพื่อการวิเคราะห์ปัจจัย โดยตัวแบบที่ได้จะอยู่ในรูปของกฎการจำแนกประเภท (Classification rule) จากการเรียนรู้ด้วยชุดการเรียนรู้ (Training set) แล้วนำไปทดสอบด้วยชุดทดสอบ (Test set) โดยใช้วิธีการตรวจสอบไขว้ กำหนดค่า k เป็น 10, 100 และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งเป็นร้อยละ

1.2 การวัดค่าประสิทธิภาพของตัวแบบการพยากรณ์

การวัดค่าประสิทธิภาพของตัวแบบการวิเคราะห์ปัจจัย จากการนำตัวแบบที่ได้จากชุดการเรียนรู้มาทดสอบด้วยชุดข้อมูลทดสอบ โดยการวัดค่าที่ได้ คือ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure)

4. การพัฒนาและการประเมินผลระบบ

สถาปัตยกรรมของระบบถูกสร้างขึ้นเพื่อใช้งานผ่านระบบเครือข่ายผ่านเว็บเบราว์เซอร์ (Web browser) ซึ่งใช้ระบบจัดการฐานข้อมูล (Database management system) คือ MySQL โดยใช้ภาษาสคริปต์เป็นพีเอชพี (PHP Hypertext Preprocessor) สำหรับทำงานบนเว็บเซิร์ฟเวอร์ (Web server) ซึ่งมีการทำงานในรูปแบบ Client-side script ตามผู้ใช้งานระบบแต่ละระดับ และทำการศึกษาความพึงพอใจจากอาจารย์ที่ปรึกษา และนักศึกษาคณะวิทยาศาสตร์ จากการสุ่มแบบเจาะจง จำนวน 30 คน เพื่อเป็นกลุ่มตัวอย่าง โดยการตอบแบบสอบถามความพึงพอใจเกี่ยวกับความสามารถในการใช้งานระบบ ที่มีลักษณะเป็นแบบเลือกตอบมาตราส่วนประมาณค่า (Likert scale) 5 ระดับตามเกณฑ์ของลิเคิร์ท (ปรีชา บุญรอด, 2541)

ผลการวิจัยและอภิปรายผล

1. ผลการสร้างแบบจำลอง

แบบจำลองเพื่อทำนายการพัฒนสภาพของนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ การวิเคราะห์โดยใช้ปัจจัยทั้ง 12 ปัจจัยที่มีผลต่อการพัฒนสภาพเพื่อนำมาวิเคราะห์คุณลักษณะของนักศึกษาที่ศึกษาอยู่ในคณะวิทยาศาสตร์ พบว่า รูปแบบการทำนายการพัฒนสภาพด้วยวิธีต้นไม้ตัดสินใจ มีจำนวน 32 กฎ ประเมินโดยใช้ 10-folds cross validation มีค่าความถูกต้อง 95.57% ค่าความแม่นยำ 95.4

% ค่าความระลึก 95.6% และค่าความถ่วงดุล 95.3% (นนทวัฒน์ ประทุมชาติ และคณะ, 2562) ซึ่งสอดคล้องกับงานวิจัยของชนิดาภา บุญประสม และจรรย์ แสนราช (2561) และจีระนันต์ เจริญรัตน์ และวีระศักดิ์ เจริญรัตน์ (2559) ที่ได้ทำการวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาที่มีผลการเรียนปกติที่มีผลการเรียนของเกรดเฉลี่ยสะสมมากกว่า 2.00 โดยใช้ต้นไม้ตัดสินใจ

2. ผลการพัฒนาและประเมินผลระบบ

การพัฒนากระบวนการทำนายนการพัฒนาศาของนักศึการระดับปริญญาตรี ในรูปแบบเว็บแอปพลิเคชันที่ประกอบด้วยกระบวนการทำงานที่สำคัญ คือ 1) ส่วนการใช้งานสำหรับนักศึกษา เพื่อนำปัจจัยของนักศึกษาสู่ระบบในการทำนายนการพัฒนาศา โดยผลการทำนายนข้อมูลของระบบแสดงผลเป็น 2 ส่วน คือ ส่วนของผลการทำนายน และส่วนของคำแนะนำจากระบบ หน้าจอการทำงานดังภาพประกอบ 1-2

มหาวิทยาลัยราชภัฏบุรีรัมย์
Buriram Rajabhat University

ระบบการทำนายนการพัฒนาศาของนักศึการ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์

580112415009

เข้าสู่ระบบ

ลงทะเบียน

สวัสดี นาย นนทวัฒน์ ทริชาติ

ระบบการทำนายนการพัฒนาศาของนักศึการสำหรับนักศึการ

กรณารับนายนข้อมูลดังต่อไปนี้

สาขาที่กำลังศึการ : --โปรดเลือก--

ปี 1	ปี 2	ปี 3
เทอม 1 : 0.00	เทอม 1 : 0.00	เทอม 1 : 0.00
เทอม 2 : 0.00	เทอม 2 : 0.00	เทอม 2 : 0.00

เกรดเฉลี่ยโรงเรียนเดิม

GPA : 0.00

สาขาวิชาที่จบจากโรงเรียนเดิม : --โปรดเลือก--

จบจากโรงเรียนขนาดใด : --โปรดเลือก--

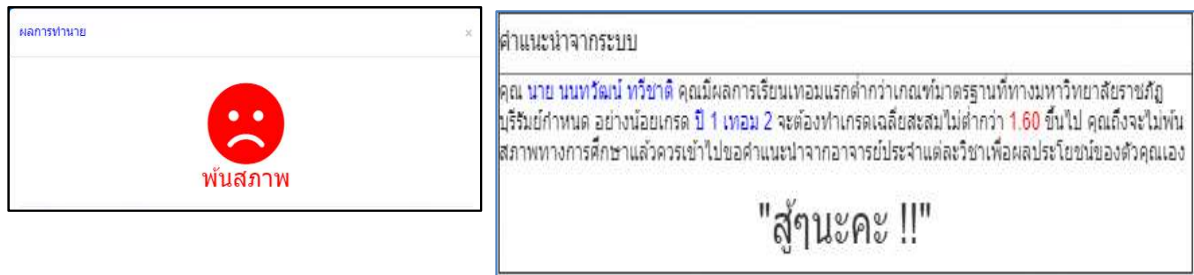
ได้ทุนกู้ยืมหรือไม่

☒ ฐ ☐ ไม่ได้ฐ

จำนวนเหตุ ฐภูมิ คมศ. หรือ ครอ.

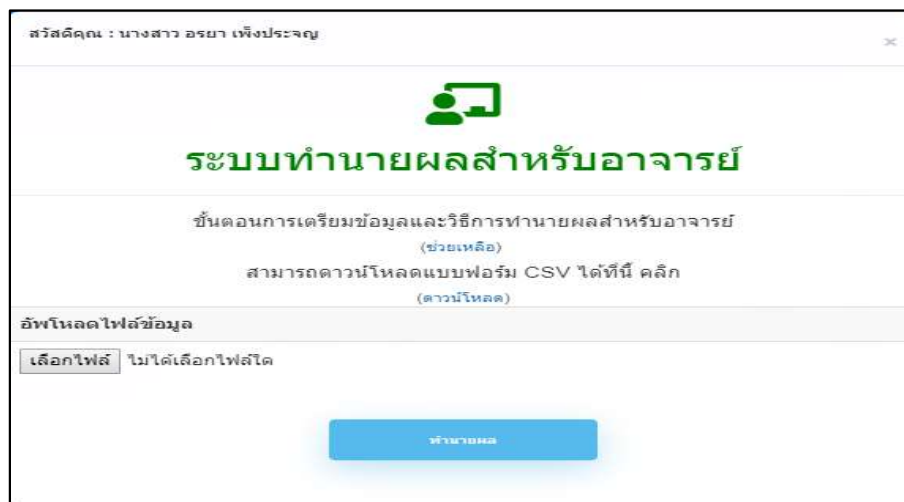
ทำนายนผล

ภาพประกอบ 1 การใช้งานระบบการทำนายนการพัฒนาศาของนักศึการสำหรับนักศึการ



ภาพประกอบ 2 ผลการทำนายและคำแนะนำจากระบบ กรณีนักศึกษาที่มีความเสี่ยงต่อการพ้นสภาพ

2) ส่วนการใช้งานสำหรับอาจารย์ที่ปรึกษา ที่สามารถดาวน์โหลดไฟล์ข้อมูลของนักศึกษาเพื่อตรวจสอบสถานะการเรียนของนักศึกษาว่ามีความเสี่ยงต่อการพ้นสภาพหรือไม่ เพื่อดำเนินการติดตามนักศึกษาต่อไป ดังภาพประกอบ 3 - 4



ภาพประกอบ 3 การใช้งานระบบการทำนายการพ้นสภาพสำหรับอาจารย์

จากภาพที่ 3 อาจารย์ที่ปรึกษาจะทำการลงทะเบียนเพื่อเข้าสู่ระบบ จากนั้นทำการดาวน์โหลดไฟล์ข้อมูลของนักศึกษาในรูปแบบ CSV เพื่อนำเข้าสู่ระบบการทำนายผล โดยระบบจะมีส่วนช่วยเหลือในการทำงานเพื่ออำนวยความสะดวกสำหรับอาจารย์ จากนั้นเลือกรูปแบบในการแสดงผลทั้งเป็นรายบุคคล และเป็นกลุ่มตามหมู่เรียน ดังภาพประกอบ 4 และ 5

เลือกรูปแบบการแสดงผล : ☒ แสดงข้อมูลเป็นจำนวนบุคคล : ☐ แสดงข้อมูลแบบแบ่งกลุ่ม

รีเซ็ต

ตารางแสดงผลทำนายตามรายบุคคล

ลำดับที่	ผลการทำนาย	ค่าความน่าจะเป็น
1	🔴 พ้นสภาพ	100%
2	😊 ไม่พ้นสภาพ	98.1%
3	😊 ไม่พ้นสภาพ	92.2%
4	😊 ไม่พ้นสภาพ	98.1%
5	😊 ไม่พ้นสภาพ	96.9%
6	😊 ไม่พ้นสภาพ	98.7%

ภาพประกอบ 4 ผลการทำนายการพ้นสภาพแบบเป็นรายห้องโดยเรียงตามเลขที่ห้อง

ในรูปแบบที่สอง คือการแสดงผลการทำนายการพ้นสภาพแบบแบ่งกลุ่มพ้นสภาพกับไม่พ้นสภาพ ดังภาพประกอบ 5

เลือกรูปแบบการแสดงผล : ☐ แสดงข้อมูลเป็นจำนวนบุคคล : ☒ แสดงข้อมูลแบบแบ่งกลุ่ม

รีเซ็ต

ตารางแสดงผลแบบแบ่งกลุ่ม

😊 ไม่พ้นสภาพ	🔴 พ้นสภาพ
ลำดับที่ ผลการทำนาย	ลำดับที่ ผลการทำนาย
2 😊 ไม่พ้นสภาพ	1 🔴 พ้นสภาพ
3 😊 ไม่พ้นสภาพ	9 🔴 พ้นสภาพ
4 😊 ไม่พ้นสภาพ	
5 😊 ไม่พ้นสภาพ	

ภาพประกอบ 5 ผลการทำนายการพ้นสภาพแบบแบ่งกลุ่มพ้นสภาพกับไม่พ้นสภาพ

ในการประเมินผลการทำงานของระบบโดยการศึกษาความพึงพอใจของการใช้งานระบบจากอาจารย์ที่ปรึกษา และนักศึกษาคณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ด้วยแบบสอบถามความพึงพอใจเกี่ยวกับความสามารถในการใช้งานระบบ (System usability) มีผลการศึกษาดังตารางที่ 2

ตารางที่ 2 ผลการศึกษาความพึงพอใจโดยรวม

รายการ	\bar{X}	S.D.	ระดับความ คิดเห็น
1. ระบบง่ายต่อการใช้งาน	3.90	0.55	มาก
2. ผลลัพธ์ที่ได้จากการทำนายมีความถูกต้อง น่าเชื่อถือ	3.55	0.51	มาก
3. ความชัดเจนของผลการทำนาย	3.80	0.77	มาก
4. ผลการทำนายที่ได้สามารถใช้ประกอบการวางแผนการศึกษาได้	3.50	0.51	มาก
5. ความเร็วในการประมวลผลของระบบ	3.70	0.47	มาก
6. ความชัดเจนของข้อความที่แสดงบนจอ	4.10	0.55	มาก
7. ความเหมาะสมของการใช้สีของตัวอักษรและรูปภาพ	3.45	0.51	มาก
8. การใช้ถ้อยคำบนจอภาพสามารถสื่อความหมายให้เข้าใจได้ง่าย	3.55	0.69	มาก
9. การกำหนดรหัสผู้ใช้งาน และรหัสผ่านในการตรวจสอบผู้เข้าใช้ระบบ	3.75	0.64	มาก
10. ผลลัพธ์ที่ได้จากการทำนายตรงกับความต้องการของผู้ใช้	3.75	0.77	มาก
โดยรวม	3.71	0.60	มาก

จากตารางที่ 2 ผลการศึกษาความพึงพอใจเกี่ยวกับความสามารถในการใช้งานระบบ พบว่า โดยภาพรวมผู้ใช้งานมีความพึงพอใจในระดับมาก (\bar{X} = 3.71, S.D. = 0.60) โดยประเด็นที่ผู้ใช้งานมีความพึงพอใจมากที่สุดคือ ความชัดเจนของข้อความที่แสดงบนจอที่มีความพึงพอใจในระดับมาก (\bar{X} = 4.10, S.D. = 0.55) รองลงมาคือ ระบบง่ายต่อการใช้งาน ความชัดเจนของผลการทำนาย การกำหนดรหัสผู้ใช้งาน และรหัสผ่านในการตรวจสอบผู้เข้าใช้ระบบ และผลลัพธ์ที่ได้จากการทำนายตรงกับความต้องการของผู้ใช้ ซึ่งแสดงให้เห็นว่าโดยภาพรวมระบบสามารถช่วยในการตัดสินใจเพื่อการวางแผนการศึกษาให้จบหลักสูตรของนักศึกษานอกจากนี้ยังช่วยให้อาจารย์ที่ปรึกษาวางแผนในการกำกับ และติดตามนักศึกษาเพื่อลดอัตราการพ้นสภาพของนักศึกษา ที่สอดคล้องกับงานวิจัยของอัล-ลาไดเดห์ (Al-Radaideh, 2006; Superby *et al.*, 2006)

ข้อเสนอแนะ

1. ในกระบวนการทำเหมืองข้อมูลขั้นตอนในการเตรียมข้อมูลถือว่ามีความสำคัญเป็นอย่างมาก ประสพการณ์ที่ได้จากการวิจัยนี้ ที่ใช้ข้อมูลที่เก็บไว้ในฐานข้อมูลเพื่อการศึกษาของมหาวิทยาลัยราชภัฏบุรีรัมย์มาทำการวิเคราะห์ที่มีตารางจำนวนมาก และข้อมูลมีหลากหลายรูปแบบ รวมถึงการมีค่าผิดพลาด (Missing value) เก็บไว้จำนวนมาก ดังนั้นในขั้นตอนการเตรียมข้อมูลจึงต้องวางแผนเป็นอย่างดี ซึ่งต้องใช้เวลานาน และต้องกระทำอย่างระมัดระวัง
2. การเลือกแอททริบิวต์ (Attributes selection) ที่จะนำมาวิเคราะห์ปัจจัยที่มีผลต่อการพัฒนาของนักศึกษาที่มีความสำคัญอีกขั้นตอนหนึ่ง โดยงานวิจัยนี้พบว่า ความเหมาะสมของแอททริบิวต์สำหรับการจัดประเภทข้อมูล (Data classification) ค่าข้อมูลจะต้องมีค่าที่วนซ้ำ และไม่หลากหลาย
3. จากส่วนหนึ่งของผลการวิจัย อัตราการพัฒนาศักยภาพของนักศึกษาชั้นปีที่ 1 มีอัตราที่สูงกว่าชั้นปีอื่น ดังนั้นมหาวิทยาลัยจะต้องให้ความสำคัญสำหรับนักศึกษาใหม่ทั้งในเรื่องของผลการเรียนและพฤติกรรมของผู้เรียนด้วย

เอกสารอ้างอิง

- จิระนันท์ เจริญรัตน์ และ วีระศักดิ์ เจริญรัตน์. (2559). การวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาที่มีผลการเรียนปกติโดยใช้ต้นไม้ตัดสินใจ. *บทความวิจัยสาขาวิชาคอมพิวเตอร์ธุรกิจ*, คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏสกลนคร. สกลนคร: มหาวิทยาลัยราชภัฏสกลนคร.
- ชนิดาภา บุญประสม และ จริญญา แสนราช. (2561). การวิเคราะห์การทำนายการลาออกกลางคันของนักศึกษา ระดับปริญญาตรีโดยใช้เทคนิควิธีการทำเหมืองข้อมูล. *วารสารวิชาการครุศาสตร์อุตสาหกรรม พระจอมเกล้าพระนครเหนือ*, 9(1), 142 – 151.
- นนทวัฒน์ ทวีชาติ, อรยา เพ็งประจัญ, วิไลรัตน์ ยาทองไชย และชูศักดิ์ ยาทองไชย. (2562). การวิเคราะห์ปัจจัยที่มีผลต่อการพัฒนาของนักศึกษาระดับปริญญาตรี ด้วยเทคนิคการทำเหมืองข้อมูล. ในงานประชุมวิชาการ การประชุมวิชาการระดับชาติการจัดการเทคโนโลยีและนวัตกรรม ครั้งที่ 5 มหาวิทยาลัยราชภัฏมหาสารคาม จังหวัดมหาสารคาม. มหาสารคาม: มหาวิทยาลัยราชภัฏมหาสารคาม
- ปรีชา บุญรอด. (2541). *มาตรวัดการวิจัยทางสังคมศาสตร์ (ตอนที่ 2)*. จาก : http://www.journal.au.edu/abac_newsletter/1998/june98/scale.html.
- พลทิพย์ เพ็งศิริ, พันธนา ก้อนเชื้อรัตน์, ชัชฎา ขวรางกูร และอัจฉราพรรณ คชเดช. (2556). การวิเคราะห์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษา นักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ. *บทความวิจัยคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ*.
- อยุธยา: พระนครศรีอยุธยา.

- Dekker, G.W., Pechenizkiy, M. and Vleeshouwers, J.M. (2009). Predicting students drop out: a Case study. In T. Barnes, M. Desmarais, C. Romero, and S. Ventura, editors, *In Proceedings of the 2nd International Conference on Educational Data Mining* (pp. 41-50).
- Superby, J.F., Vandamme, J.P. and Meskens, N. (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods, *In Proceedings of 8th International Conference on Intelligent Tutoring Systems* (pp. 37-44).
- Jadrić, M., Garača, Ž. and Ćukušić, M. (2010) Student dropout analysis with application of data mining methods, *Management*, 15(1), 31-46.
- Al-Radaideh, Q.A., Al-Shawakfa, E.M. and Al-Najjar, M.I. (2006). Mining student data using decision trees, *In Proceedings of International Arab Conference on Information Technology* (pp.1-5).
- Kotsiantis, S. (2009) Educational Data Mining: A Case Study for Predicting Dropout – Prone Students. *International Journal of Knowledge Engineering and Soft Data Paradigms*, 1(2), 101–111.
- Fayadd, U., Piatetsky-Shapiro, G. and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37-54.