

การวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาระดับปริญญาตรี ด้วยเทคนิคการทำเหมืองข้อมูล

Factor Analysis with Data Mining Technique in Higher Educational Student Drop Out

นนทวัฒน์ ทวีชาติ¹ อรยา เพ็งประจัญ² วิไลรัตน์ ยาทองไชย^{3*} และชูศักดิ์ ยาทองไชย⁴

สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์^{1,2,3,4}

tongnam33@gmail.com, Wilairat.bru@gmail.com*, chusak.bru@gmail.com

บทคัดย่อ

การเพิ่มขึ้นของอัตราการพ้นสภาพของนักศึกษาระดับอุดมศึกษาเป็นปัญหาที่สำคัญประการหนึ่งในสถาบันการศึกษา การค้นหาความรู้จากฐานข้อมูลด้านการศึกษาด้านการทำเหมืองข้อมูลที่มีประสิทธิภาพในการวิเคราะห์หาปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาสามารถนำไปสู่การวางแผนการศึกษาและการจัดการเพื่อลดอัตราการพ้นสภาพของนักศึกษารวมถึงการสร้างระบบสารสนเทศสำหรับสนับสนุนการตัดสินใจของผู้บริหารสถานศึกษาเพื่อปรับปรุงคุณภาพระบบการศึกษาระดับอุดมศึกษา ทำการวิเคราะห์ข้อมูลเพื่อสร้างแบบจำลองโดยใช้เทคนิคการทำเหมืองข้อมูลแบบการจำแนกข้อมูลด้วยผังต้นไม้ตัดสินใจ และใช้อัลกอริทึม J48 ในการทดสอบแบบจำลองด้วยวิธีการตรวจสอบไขว้ และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งเป็นร้อยละ โดยใช้โปรแกรมWEKA ใช้ชุดข้อมูลของนักศึกษาคณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ ที่เข้าศึกษาระหว่างปี พ.ศ. 2556 – 2559 มีจำนวน 3,604 ชุดข้อมูล ได้ปัจจัยที่มีผลต่ออัตราการพ้นสภาพ จำนวน 11 ปัจจัย คือ ปัจจัยที่เกี่ยวข้องกับการเรียนในระดับมัธยมศึกษา และปัจจัยระหว่างเรียนในมหาวิทยาลัย โดยวัดค่าความถูกต้องได้ร้อยละ 96.73 ความแม่นยำได้ร้อยละ 96.6 ค่าความระลึกได้ร้อยละ 96.7 และค่าความถ่วงดุลได้ร้อยละ 96.5 จากผลการวิจัยนี้สามารถนำผลการจำแนกข้อมูลมาใช้ในการพัฒนาระบบทำนายการพ้นสภาพนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ต่อไป

คำสำคัญ: ระดับอุดมศึกษา, การตกรอก, เทคนิคเหมืองข้อมูล, การจำแนกข้อมูล

ABSTRACT

The increase of students' drop out rate in higher education is one of the important problems in most institutions. The discovery of hidden knowledge from the educational data system by the effective process of data mining technology to analyze factors affecting student drop out can lead to a better academic planning and management to reduce students drop out rate, as well as can inform valuable information for decision making of stake holder to improve the quality of higher educational system. In this paper, we consider three issues of factors affecting students' drop out rate. These factors are conditions related to the students before admission, and factors related to the students during the study periods in the university including the target value to be predict for factors analysis. We use tree-based classification algorithm, J48 or C4.5 to analyze the data. To evaluated the model, we use both 10-fold

cross validation and percentage split set methods. Accuracy rate was satisfactory and the induced models are actionable and potentially applicable to higher education planning.

Keyword: Higher education, Student drop out, Data mining technique, Classification

บทนำ

เทคโนโลยีสารสนเทศได้เข้ามามีบทบาทสำคัญในการจัดการและเก็บรวบรวมข้อมูลขนาดใหญ่ขององค์กรโดยข้อมูลที่เก็บไว้สามารถนำมาใช้ในการสร้างสารสนเทศเพื่อใช้ในการประกอบการตัดสินใจสำหรับองค์กร เหมือนข้อมูลคือ กระบวนการในการวิเคราะห์ข้อมูลโดยอัตโนมัติเพื่อช่วยให้ค้นพบและสร้างรูปแบบของข้อมูลจากข้อมูลที่ถูเก็บไว้ (Fayadd, Piatessky-Shapiro, and Smyth, 1996) การใช้เทคนิคเหมืองข้อมูลเพื่อวิเคราะห์ฐานข้อมูลทางการศึกษาถือว่าเป็นประโยชน์อย่างยิ่งต่อสถาบันอุดมศึกษา

ปัจจุบันข้อมูลทางการศึกษา เช่น ข้อมูลของผู้เรียน รายละเอียดหลักสูตร การวัดและการประเมินผล และอื่น ๆ เพิ่มขึ้นอย่างมากเป็นผลให้หลายปัจจัยมีส่วนเกี่ยวข้องกับคุณภาพของระบบการศึกษาระดับอุดมศึกษา คุณภาพเป็นปัจจัยหลักที่สำคัญในการศึกษาระดับอุดมศึกษา การได้มาซึ่งคุณภาพนั้นต้องได้รับการวางแผน ติดตาม และควบคุมในแต่ละกระบวนการ การศึกษาโดยมีวัตถุประสงค์หลักเพื่อการปรับปรุงประสิทธิภาพของผู้เรียน ตัวบ่งชี้จุดอ่อนของคุณภาพการศึกษาคือ จำนวนการตกออกของผู้เรียนที่เพิ่มสูงขึ้น การทำนายจำนวนผู้เรียนที่ตกออก และปัจจัยที่มีผลต่อสถานการณ์การตกออกต้องใช้กระบวนการที่มีประสิทธิภาพ เดกเกอร์และคณะ (Dekker, Pechenizkiy, and Vleeshouwers, 2009) ใช้เทคนิคการทำเหมืองข้อมูลเพื่อทำนายการฟื้นสภาพและระบุปัจจัยความสำเร็จของนักศึกษาวิศวกรรมไฟฟ้า กอตเซนทิส (Kotsiantis, 2009) ยังใช้เทคนิคการทำเหมืองข้อมูลเพื่อการศึกษาและทำนายการออกกลางคันซึ่งถือเป็นความล้มเหลวของโรงเรียนที่เป็นสิ่งสำคัญในการแก้ไขปัญหา เช่นเดียวกับ จาดริค และคณะ (Jadrić, Garača, and Ćukušić, 2010) ได้ทำการวิเคราะห์ปัญหาของนักศึกษาที่เลิกเรียนระดับอุดมศึกษาโดยใช้วิธีการเหมืองข้อมูล ซึ่งจะเห็นได้ว่าได้มีงานวิจัยที่เกี่ยวข้องกับการใช้เหมืองข้อมูลเพื่อการพัฒนาการศึกษาและค้นหาความรู้จากสภาพแวดล้อมทางการศึกษามากขึ้นทั้งนี้เพื่อเพิ่มประสิทธิภาพในการบริหารจัดการด้านการศึกษา

บทความนี้นำเสนอการใช้สารสนเทศทางการศึกษาจากฐานข้อมูลทางการศึกษาของมหาวิทยาลัยบนแนวทางการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลที่มีประสิทธิภาพในการวิเคราะห์ปัจจัยหลักที่มีผลต่อการฟื้นสภาพของนักศึกษาในระดับปริญญาตรี โดยเป้าหมายหลักของการศึกษาคือการนำแบบจำลองที่ได้จากการวิเคราะห์มาพัฒนาระบบทำนายการฟื้นสภาพของนักศึกษาเพื่อวางแผนในการจัดการศึกษา รวมถึงการติดตามดูแลเพื่อลดจำนวนการฟื้นสภาพของนักศึกษาต่อไป

1. วัตถุประสงค์การวิจัย

เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการฟื้นสภาพของนักศึกษาในระดับปริญญาตรี โดยใช้เทคนิคเหมืองข้อมูล

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

เทคนิคการจำแนกประเภท (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 ให้

ความแม่นยำสูง และนำไปประยุกต์ใช้กับการค้นหาข้อมูลด้วยภาษาสอบถามเชิงโครงสร้าง (SQL) ได้ (Witten and E. Frank, 2005) เทคนิคเหมืองข้อมูลได้ถูกนำไปใช้เพื่อพัฒนาคุณภาพการศึกษาของการศึกษาระดับอุดมศึกษาหลายด้าน

จอร์จันต์ เจริญรัตน์ และวีระศักดิ์ เจริญรัตน์ (2559) ได้ทำการวิเคราะห์ปัจจัยที่ส่งผลต่อการพัฒนาของนักศึกษาที่มีผลการเรียนปกติที่มีผลการเรียนของเกรดเฉลี่ยสะสมมากกว่า 2.00 โดยใช้ต้นไม้ตัดสินใจ โดยใช้ข้อมูลนักศึกษาระดับปริญญาตรี คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏสกลนคร จำนวน 3,385 ข้อมูล พบว่า ปัจจัยที่มีความสำคัญต่อการพัฒนาของนักศึกษาที่มีผลการเรียนปกติ แบ่งเป็น 4 กลุ่ม คือ กลุ่ม Best ปัจจัยคือ วุฒิก่อนศึกษาต่อ กลุ่ม Excellent ปัจจัยคือ อาชีพมารดา และสาขาวิชาที่เรียน กลุ่ม Good ปัจจัยคือ ทุนกู้ยืมเพื่อการศึกษา สถานภาพของครอบครัว รายได้บิดามารดา ซึ่งผลที่ได้สามารถนำไปพัฒนาระบบพยากรณ์การพัฒนาศักยภาพของนักศึกษา เพื่อเป็นแนวทางในการบริหารจัดการให้อาจารย์ที่ปรึกษาได้วางแผนการเรียนของนักศึกษาต่อไป

พุดพิงศ์ 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) ในการพัฒนาศักยภาพ เพื่อนำผลการวิจัยที่ได้ไปสู่การวางแผนการลดจำนวนนักศึกษาที่จะพัฒนาศักยภาพ (Superby et al., 2006) รวมถึงการประยุกต์ใช้ต้นไม้ตัดสินใจในการทำนายผลการเรียนของนักศึกษาที่ลงทะเบียนเรียนรายวิชาการเขียนโปรแกรมภาษาซี ของมหาวิทยาลัยYarmouk ประเทศจอร์แดน โดยใช้อัลกอริทึม ID3 C4.5 และ Naive Bayes ซึ่งได้ผลการทำนายที่ดีกว่าการใช้ อัลกอริทึมอื่น (Al-Radaideh, 2006) จากการวิจัยดังกล่าวจะเห็นได้ว่าเทคนิคการทำเหมืองข้อมูลสามารถนำมาใช้กับฐานข้อมูลด้านการศึกษาของมหาวิทยาลัยเพื่อค้นหารูปแบบที่ซ่อนอยู่และความสัมพันธ์ระหว่างข้อมูลที่เป็นประโยชน์ช่วยในการตัดสินใจได้ดียิ่งขึ้น

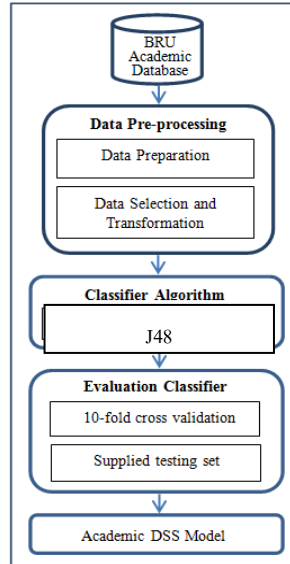
วิธีดำเนินการวิจัย

1. กรอบการศึกษา

การวิจัยนี้มีกรอบในการศึกษา 4 ขั้นตอน คือ การเตรียมข้อมูลก่อนประมวลผล อัลกอริทึมสำหรับการจำแนกประเภท การวัดประสิทธิภาพการจำแนกประเภท และการสร้างแบบจำลอง

- **การเตรียมข้อมูลก่อนประมวลผล (Data Pre-processing)** เป็นขั้นตอนแรกก่อนการทำเหมืองข้อมูลเนื่องจากข้อมูลในฐานข้อมูลที่จะนำมาวิเคราะห์อาจมีข้อมูลไม่สมบูรณ์ (Incomplete data) ข้อมูลมีค่าผิดพลาด (Error) หรือมีค่าผิดปกติ (Outliers) ตลอดจนข้อมูลไม่สอดคล้อง (Inconsistent data) ที่ต้องตรวจสอบความถูกต้องในเรื่องของรูปแบบข้อมูล ค่าของข้อมูลก่อนการประมวลผล โดยงานวิจัยนี้ได้ทำการคัดข้อมูลที่เป็นส่วนรบกวน หรือข้อมูลที่ไม่เกี่ยวข้องออกไป (Data cleaning) ทำการคัดเลือกข้อมูลและแปลงข้อมูลให้เหมาะสมกับการทำเหมืองข้อมูล (Data selection and transformation)
- **อัลกอริทึมสำหรับการจำแนกประเภท (Classifier Algorithm)** ได้ทำการจำแนกประเภทด้วยอัลกอริทึม J48
- **การวัดประสิทธิภาพการจำแนกประเภท (Evaluation Classifier)** โดยใช้วิธีการตรวจสอบไขว้ (K-fold cross-validation) และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งเป็นร้อยละ (Percentage Split) เพื่อประเมินผลแบบจำลองจากค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าความถ่วงดุล (F-measure)

- การสร้างแบบจำลอง (Academic DSS Model) เมื่อทำการสร้างแบบจำลองจะทำให้ได้ถูกจากอัลกอริทึมการจำแนกประเภทเพื่อนำไปสู่การพัฒนาระบบทำนายการผันสภาพนักศึกษาในระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์ต่อไป เพื่อนำมาใช้ในการวางแผนติดตามและปรับปรุงประสิทธิภาพการเรียนของผู้เรียนเพื่อลดอัตราการผันสภาพ ดังแสดงในภาพที่ 1



ภาพที่ 1 กรอบการศึกษา

2. การรวบรวมข้อมูลเพื่อการประมวลผล

คลังข้อมูลของมหาวิทยาลัยราชภัฏบุรีรัมย์จะเก็บฐานข้อมูลที่เกี่ยวข้องกับการทำงานของมหาวิทยาลัย เช่น ฐานข้อมูลการศึกษา ฐานข้อมูลการเงิน ฐานข้อมูลการประกันคุณภาพ เป็นต้น เพื่อนำมาใช้ประโยชน์ในการบริหารงานของมหาวิทยาลัย โดยข้อมูลที่น่ามาใช้ในการวิจัยครั้งนี้คือ ข้อมูลจากฐานข้อมูลการศึกษา ที่จัดเก็บข้อมูลของนักศึกษาตั้งแต่ปีการศึกษา 2556 - 2559 โดยใช้ข้อมูลตัวอย่างจากนักศึกษาคณะวิทยาศาสตร์ ด้วยมีอัตราการผันสภาพมากที่สุดในมหาวิทยาลัย จากนั้นนำข้อมูลจากหลายตารางมารวมกันผ่านกระบวนการเตรียมข้อมูลทำให้ได้ข้อมูลจำนวน 3,604 ระเบียบ

3. การคัดเลือกและแปลงข้อมูล (Data selection and transformation)

โดยการวิจัยนี้ได้กำหนดสมมติฐานของปัจจัยที่มีผลต่อการผันสภาพของนักศึกษาคือ ปัจจัยที่เกี่ยวข้องกับการเรียนในระดับมัธยมศึกษา คือ ผลการเรียนเฉลี่ย แผนการเรียนที่จบ ขนาดโรงเรียน และปัจจัยระหว่างเรียนในมหาวิทยาลัย คือ สาขาวิชา เกรดเฉลี่ยแต่ละภาคเรียน จำนวน 6 ภาคเรียน สถานการณ์กู้ยืม 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 = ศิลป์คำนวณ 3 = อื่น ๆ
SchoolSize	ขนาดโรงเรียน	{Small,Medium,Large}
Loan	สถานะการกู้ยืมเพื่อการศึกษา	{Yes, No} → Yes = กองทุนกู้ยืม No = ไม่กู้ยืม
DropTerm	ภาคเรียนที่พ้นสภาพ	{1, 2, 3, 4, 5, 6, No}
DropOut	สถานะการพ้นสภาพ	{Yes, No}

4. การวิเคราะห์ข้อมูล

การวิจัยนี้ใช้เทคนิคต้นไม้ตัดสินใจในการจำแนกประเภทของข้อมูลเพื่อสร้างแบบจำลอง (Model) ดังนี้

4.1 การสร้างและทดสอบแบบจำลอง

งานวิจัยนี้ใช้โปรแกรม WEKA ในการสร้างตัวแบบการวิเคราะห์ที่ปัจจัยและทดสอบตัวแบบด้วยเทคนิคต้นไม้ตัดสินใจ (Decision tree technique) โดยใช้วิธีการจำแนกประเภท (Classification) เลือกใช้ J48 คือ อัลกอริทึม C4.5 เพื่อใช้ในการเรียนรู้ (Learning) และสร้างแบบจำลองต้นไม้เพื่อการวิเคราะห์ปัจจัย โดยตัวแบบที่ได้จะอยู่ในรูปของกฎการจำแนกประเภทจากการเรียนรู้ด้วยชุดการเรียนรู้ (Training set) แล้วนำไปทดสอบด้วยชุดทดสอบ (Test data) โดยใช้วิธีการตรวจสอบไขว้ กำหนดค่า k เป็น 10, 100 และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งเป็นร้อยละ

วิธีการตรวจสอบไขว้เป็นการแบ่งข้อมูลแบบสุ่มออกเป็นจำนวนกลุ่มตามตัวเลข คือ ถ้าใส่ตัวเลข ค่า N เข้าไปจะมีการแบ่งข้อมูลออกเป็นส่วนย่อยจำนวน N และจะเก็บข้อมูล 1 ส่วนย่อยเพื่อไว้ทดสอบ ส่วนข้อมูลที่เหลือนำมาสร้างแบบจำลอง และจะทำการไปจนกว่าข้อมูลถูกใช้ทดสอบจนหมดทุกส่วนการวิจัยครั้งนี้ได้กำหนดค่า k ในวิธีการตรวจสอบไขว้คือ 10, 100

4.2 การวัดค่าประสิทธิภาพของตัวแบบการพยากรณ์

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

ผลการวิจัยและอภิปรายผล

ในการวิจัยครั้งนี้ได้สรุปผลข้อมูลและนำเสนอผลการวิเคราะห์ที่ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาได้ดังนี้

1) การวิเคราะห์ที่ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา โดยการวิเคราะห์การจำแนกประเภทด้วยเทคนิคต้นไม้ตัดสินใจ ได้ผลในรูปแบบของต้นไม้ตัดสินใจที่สามารถแปลในรูปของกฎโดยตัวแบบที่ได้จะอยู่ในรูปของกฎการจำแนกประเภทข้อมูลจากการเรียนรู้ด้วยชุดการเรียนรู้ ซึ่งได้กฎการจำแนกประเภทข้อมูลจากการเรียนรู้มากมายแต่กฎที่มีความสัมพันธ์กัน มีอยู่ 9 กฎ ดังนี้

— นักศึกษาที่มีการกู้ยืมเพื่อการศึกษาจะไม่พ้นสภาพ ในขณะที่นักศึกษาที่ไม่กู้ยืมเพื่อการศึกษาที่มีเกรดเฉลี่ยจากโรงเรียนมัธยมต่ำกว่า 2.42 จะพ้นสภาพ

— นักศึกษาที่ไม่กู้ยืมเพื่อการศึกษาและมีเกรดเฉลี่ยจากโรงเรียนมัธยมมากกว่า 2.42 และศึกษาในสาขาวิชาคณิตศาสตร์จะไม่พ้นสภาพในขณะที่นักศึกษาที่ศึกษาในสาขาวิชาอื่นจะมีการพ้นสภาพ

— นักศึกษาที่พ้นสภาพหลังจากเรียนภาคเรียนแรกผ่านไปแล้วคือ ผู้ที่มีเกรดเฉลี่ยน้อยกว่า 1.6 และไม่กู้ยืมกองทุนเพื่อการศึกษา ส่วนผู้ที่มีการกู้ยืมเพื่อการศึกษาจะพ้นสภาพหลังจากเรียนภาคเรียนที่สอง

- นักศึกษาที่มีเกรดเฉลี่ยในภาคเรียนแรกอยู่ระหว่าง 1.6-1.99 และภาคเรียนที่สองน้อยกว่า 1.6 จะพ้นสภาพหลังจากเรียนภาคเรียนที่สองแล้ว
- นักศึกษาที่มีเกรดเฉลี่ยภาคเรียนที่สี่มากกว่า 1.6 จะไม่พ้นสภาพ
- นักศึกษาที่เรียนสาขาวิชาวิทยาศาสตร์การกีฬาจะไม่พ้นสภาพ
- นักศึกษาที่เรียนสาขาวิชาสาธารณสุขชุมชนและวิทยาการคอมพิวเตอร์ จะมีอัตราการพ้นสภาพสูงในภาคเรียนที่หนึ่ง
- นักศึกษาที่เรียนสาขาวิชาเทคโนโลยีสารสนเทศ (กลุ่มวิชาเทคโนโลยีสารสนเทศและกลุ่มวิชาเทคโนโลยีคอมพิวเตอร์) ที่มีเกรดเฉลี่ยในภาคเรียนแรกมากกว่า 2.5 แต่ภาคเรียนที่ 2 และ 3 น้อยกว่า 1.6 จะพ้นสภาพหลังจากเรียนภาคเรียนที่ 4
- นักศึกษาส่วนใหญ่จะพ้นสภาพในภาคเรียนแรก ด้วยเหตุผลหลักที่ว่าต้องการเปลี่ยนสาขาวิชา และเข้ามาเรียนอีกครั้งในปีต่อไปในขณะที่ส่วนหนึ่งมีปัญหาเรื่องการเงิน การย้ายที่อยู่ เป็นต้น

2) วิธีการที่ใช้ในการทวนสอบแบบจำลองคือ วิธีการตรวจสอบไขว้ กำหนดค่า k เป็น 10, 100 และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งเป็นร้อยละ โดยกำหนดร้อยละ 66 การวิเคราะห์โดยใช้ปัจจัยทั้ง 11 ปัจจัยที่มีผลต่อการพ้นสภาพ เพื่อนำมาวิเคราะห์คุณลักษณะของนักศึกษาที่ศึกษาอยู่ในคณะวิทยาศาสตร์ มีค่าผลการประเมินแบบจำลอง แสดงได้ดังตารางที่ 2

ตารางที่ 2 ผลการประเมินปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา

วิธีทวนสอบ	ค่าความถูกต้อง (%)	ค่าความแม่นยำ (%)	ค่าความระลึก (%)	ค่าความถ่วงดุล (%)
การตรวจสอบไขว้ (10 fold)	98.14	98.14	98.14	98.14
การตรวจสอบไขว้ (100 fold)	98.05	98.5	98.5	98.5
การแบ่งข้อมูลแบบสุ่มด้วยแบ่งร้อยละ 66	96.73	96.6	96.7	96.5

จากตารางที่ 2 แสดงผลการประเมินปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา โดยใช้วิธีการทวนสอบ คือ การตรวจสอบไขว้ (10 fold) การตรวจสอบไขว้ (100 fold) และการแบ่งข้อมูลแบบสุ่มด้วยแบ่งร้อยละ 66 ที่มีค่าสูงด้วยมาจากการนำข้อมูลที่มีมาหาค่าถ่วงน้ำหนักเพื่อให้ข้อมูลของแต่ละสาขาที่มีจำนวนนักศึกษาไม่เท่ากันอยู่ในสัดส่วนที่ใกล้เคียงกันในการวิเคราะห์ข้อมูล

ข้อเสนอแนะ

1. ในกระบวนการทำเหมืองข้อมูลขั้นตอนในการเตรียมข้อมูลถือว่ามีความสำคัญเป็นอย่างมาก ประสพการณ์ที่ได้จากการวิจัยนี้ ที่ใช้ข้อมูลที่เก็บไว้ในฐานข้อมูลเพื่อการศึกษาของมหาวิทยาลัยราชภัฏบุรีรัมย์มาทำการวิเคราะห์ที่มีจำนวนตารางจำนวนมาก และข้อมูลมีหลากหลายรูปแบบ รวมถึงการมีค่าผิดพลาด (Missing value) เก็บไว้จำนวนมาก ดังนั้นในขั้นตอนการเตรียมข้อมูลจึงต้องวางแผนเป็นอย่างดี ซึ่งต้องใช้เวลาและต้องกระทำอย่างระมัดระวัง

2. การเลือกแอททริบิวต์ (Attributes selection) ที่จะนำมาวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาก็มีความสำคัญอีกขั้นตอนหนึ่ง โดยงานวิจัยนี้พบว่า ความเหมาะสมของแอททริบิวต์สำหรับการจัดประเภทข้อมูล (Data classification) ค่าข้อมูลจะต้องมีค่าที่วนซ้ำ และไม่หลากหลาย

3. จากส่วนหนึ่งของผลการวิจัย อัตราการพ้นสภาพของนักศึกษาชั้นปีที่ 1 มีอัตราที่สูงกว่าชั้นปีอื่น ดังนั้นมหาวิทยาลัยจะต้องให้ความสำคัญสำหรับนักศึกษาใหม่ทั้งในเรื่องของการเรียนและพฤติกรรมของผู้เรียนด้วย

จากแบบจำลองที่ได้จากการวิเคราะห์จะนำไปสู่การพัฒนากระบวนการทำนายการพ้นสภาพของนักศึกษาในขั้นต่อไป เพื่อวางแผนในการจัดการศึกษา รวมถึงการติดตามดูแลเพื่อลดจำนวนการพ้นสภาพของนักศึกษาต่อไป

เอกสารอ้างอิง

- จิระนันท์ เจริญรัตน์ และวีระศักดิ์ เจริญรัตน์. (2559). การวิเคราะห์ปัจจัยที่ส่งผลต่อการพ้นสภาพของนักศึกษาที่มีผลการเรียนปกติโดยใช้ต้นไม้ตัดสินใจ. **บทความวิจัยสาขาวิชาคอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏสกลนคร**.
- พุดผิงค์ เพ็งศิริ, พันธนา ก้อนเชื้อรัตน์, ชัชฎา ขวรางกูร และอัจฉราพรรณ คชเดช. (2556). การวิเคราะห์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษา นักศึกษามหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ. **บทความวิจัยคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ**. พระนครศรีอยุธยา.
- G.W. Dekker, M. Pechenizkiy, and J.M. Vleeshouwers, Predicting students drop out: a Case study. In T. Barnes, M. Desmarais, C. Romero, and S. Ventura, editors, **Proceedings of the 2nd International Conference on Educational Data Mining**, 2009, pp.41-50.
- J. F. Superby, J. P. Vandamme, and N. Meskens. Determination of factors influencing the achievement of the first-year university students using data mining methods, **Proceedings of 8th International Conference on Intelligent Tutoring Systems**, 2006, pp. 37-44.
- M. Jadrić, Ž. Garača, and M. Čukušić, Student dropout analysis with application of data mining methods, **Management**, Vol.15, No.1, 2010, pp. 31-46
- Q.A. Al-Radaideh, E.M. Al-Shawakfa, and M.I. Al-Najjar, Mining student data using decision trees, **Proceedings of International Arab Conference on Information Technology**, 2006, pp.1-5.
- S. Kotsiantis, Educational Data Mining: A Case Study for Predicting Dropout – Prone Students. **International Journal of Knowledge Engineering and Soft Data Paradigms**, Vol.1, No.2, 2009, pp.101–111.
- U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, From data mining to knowledge discovery in databases, **AI Magazine**, Vol.17, No.3, 1996, pp.37-54.