

วิจัยเชิงสำรวจการใช้เทคนิคเหมืองข้อมูลเพื่อค้นหา

ปัจจัยที่ส่งผลต่อการผันสภาพของนักศึกษา

Survey Paper : Using Data Mining To Find Factors Influence Dropout Students

กฤตเมธ เพียรวิบูลย์ (Kittamet Peanviboon),

ณรงค์ ดอกคำ (Narong Dokkham),

มลฤดี จันทิมา (Monrudee Chantima),

กิตติชัย ศรีธรรมพงษ์ (Kittichai Srithummapong)

ธีรวัฒน์ พูลสวัสดิ์ (Teerawat Polsawat)

อดิพัฒน์ ประวันณา (Adipat Pawanna)

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอบทบาทและความสำคัญของการทำเหมืองข้อมูลทางการศึกษา เพื่อค้นหาปัจจัยที่ส่งผลต่อการผันสภาพของนักศึกษา ซึ่งอาศัยเทคนิคต่างๆ ของเหมืองข้อมูลเช่นเทคนิคต้นไม้ตัดสินใจเทคนิคกฎความสัมพันธ์ โดยเครื่องมือที่ผู้วิจัยนิยมใช้คือโปรแกรม WEKA จากผลการวิจัยพบว่า ปัจจัยที่สำคัญที่จะส่งผลกระทบต่อการศึกษา ได้แก่ ปัจจัย ทุนกู้ยืมเพื่อการศึกษา สถานะภาพครอบครัว อาชีพบิดามารดา รายได้บิดามารดา

คำสำคัญ: ผันสภาพนักศึกษา, เหมืองข้อมูล

Abstract

The main idea of this paper is present the techniques of creating data mining and applied work in education. For example decision tree association rules. WEKA is the most popular tool that researcher use to study the problems. The result of the research found Factors Influence Dropout Students was student loan, family status, parent's job, and father's salary.

Keyword: Dropout Students, Data mining

1. บทนำ

เทคโนโลยีสารสนเทศในปัจจุบันมีบทบาทในชีวิตประจำวันเป็นอย่างมาก ในสถานศึกษาก็มีการนำเอาเทคโนโลยีสารสนเทศมาประยุกต์ใช้ในรูปแบบต่างๆ อย่างกว้างขวาง ในหลายมิติ เช่น เทคโนโลยีสารสนเทศด้านบุคลากร ด้านการพัฒนาหลักสูตร การลงทะเบียนเรียน และการติดตาม

ประเมินผลการเรียนของนักศึกษา เป็นต้น การเก็บข้อมูลเหล่านี้มีการเก็บอย่างต่อเนื่องและมีปริมาณมากขึ้น แต่การนำข้อมูลเหล่านี้มาใช้งานยังไม่สามารถตอบสนองต่อความต้องการของผู้บริหารได้ การนำเอาเทคนิคเหมืองข้อมูลมาวิเคราะห์ข้อมูลในอดีตก็เพื่อหาความสัมพันธ์และสร้างต้นแบบในการพยากรณ์หรือทำนายแนวโน้มในอนาคต

จากปัญหาของสถานศึกษาในปัจจุบันพบว่า มีนักศึกษากว่าจำนวนไม่น้อยที่ตกออกระหว่างภาคการศึกษาซึ่งส่งผลกระทบต่อทางสถาบันการศึกษาและตัวนักศึกษาเอง ส่งผลกระทบต่อประสิทธิภาพในการบริหารจัดการ การจัดหลักสูตรการศึกษา รวมไปถึงงบประมาณของสถานศึกษา

ด้วยเหตุผลดังกล่าวงานวิจัยนี้จึงมีความสนใจที่จะศึกษาถึงปัจจัยที่มีผลต่อการตกออกของนักศึกษา เพื่อค้นหาสาเหตุที่มีผลกระทบต่อการศึกษา เพื่อให้เกิดประโยชน์ต่อการวางแผนติดตามนักศึกษากลุ่มที่มีความเสี่ยงและเพื่อช่วยลดจำนวนนักศึกษาที่จะตกออกในปีต่อๆ ไป

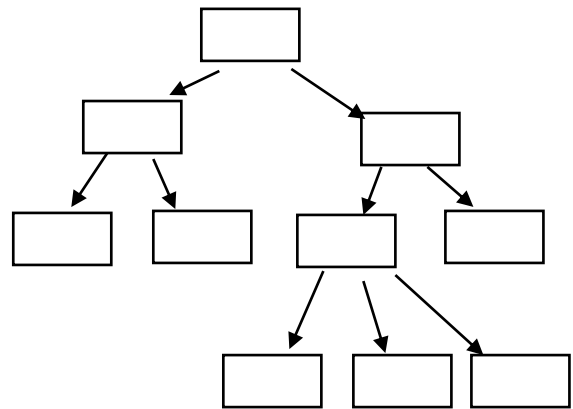
2. ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

2.1 เหมืองข้อมูล (Data Mining)

การค้นหาความสัมพันธ์และรูปแบบ (Pattern) ทั้งหมด ซึ่งมีอยู่จริงในฐานข้อมูล แต่ได้ถูกซ่อนไว้ภายในข้อมูลจำนวนมาก เพื่อค้นหาในรูปแบบ แนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง และหลักคณิตศาสตร์ความรู้ที่ได้จากการทำเหมืองข้อมูลมีหลายรูปแบบ ได้แก่ กฎความสัมพันธ์ (Association rule) แสดงความสัมพันธ์ของเหตุการณ์หรือวัตถุที่เกิดขึ้นพร้อมกัน ตัวอย่างของการประยุกต์ใช้กฎเชื่อมโยง เช่น การวิเคราะห์

ข้อมูลการขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย (POS) หรือร้านค้าออนไลน์ แล้วพิจารณาสินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่าคนที่ซื้อเทปวิดีโอมักจะซื้อเทปกาวยด้วย การจำแนกประเภทข้อมูล (Data classification) หากกฎเพื่อระบุประเภทของวัตถุจากคุณสมบัติของวัตถุ เช่น หากความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย หรือการวิจัยทางการแพทย์ ในทางธุรกิจ จะใช้เพื่อคุณลักษณะของผู้ที่จะก่อหนี้ดีหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้ การแบ่งกลุ่มข้อมูล (Data clustering) แบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์สาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน การสร้างมโนภาพ (Visualization) สร้างภาพคอมพิวเตอร์กราฟิกที่สามารถนำเสนอข้อมูลมากมายอย่างครบถ้วนแทนการใช้ข้อความนำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ่อนเร้นเมื่อดูข้อมูลชุดนั้นด้วยจินตทัศน์

2.2 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) คือ แบบจำลองทางคณิตศาสตร์ เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่ง มีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) สามารถสร้างแบบจำลองการจัดหมวดหมู่ (Clustering) ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้า (Training set) ได้โดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไขแล้วผลลัพธ์” เช่น “If Income = High and Married = No THEN Risk = Poor” “If Income = High and Married = Yes THEN Risk = Good” ส่วนประกอบของต้นไม้ตัดสินใจ ประกอบด้วย โหนด (Node) คือคุณสมบัติต่างๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใด ซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก (Root Node) กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวนของกิ่งจะเท่ากับว่าคุณสมบัติของโหนด ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดง ส่วนประกอบของต้นไม้ตัดสินใจ



ภาพที่ 1 รูปแบบต้นไม้ตัดสินใจ

2.3 กฎความสัมพันธ์ (Association Rules) เป็นกระบวนการหนึ่งในการทำ Data Mining ที่ได้รับความนิยมมาก โดยจะใช้ Association Rules ในการหาความสัมพันธ์ของข้อมูลสองชุดหรือมากกว่าสองชุดขึ้นไปภายในกลุ่มข้อมูลที่มีขนาดใหญ่ ในการหาความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน แต่ขั้นตอนวิธีที่เป็นที่รู้จักและใช้อย่างแพร่หลายคือ ขั้นตอนวิธี Apriori

การหา frequent itemset เป็นการหารูปแบบของข้อมูลที่เกิดขึ้นร่วมกันบ่อยๆ ในฐานข้อมูล หรือ มากกว่าค่า minimum support ที่ผู้ใช้กำหนด กฎความสัมพันธ์เขียนอยู่ในรูป $x \rightarrow y$ โดยที่ $x, y \subseteq I$ และ $x \cap y = \emptyset, x \rightarrow y | Support = S, Confidence$ หมายถึงกฎ $x \rightarrow y$ ที่มีค่าสนับสนุนเท่ากับ S ในเซตของรายการ (Transaction) จากฐานข้อมูล และ D หมายถึง S% ของจำนวนรายการทั้งหมดประกอบด้วย itemsets x และ y (เขียนแทนด้วย $x \cup y$) และมีค่าความเชื่อมั่น เท่ากับ C ถ้า C% ของรายการที่ประกอบด้วย itemsets x แล้วรายการเหล่านั้นมี itemsets y ด้วย \rightarrow Itemset คือ ความสัมพันธ์ของข้อมูลที่หาได้ โดย Itemset ประกอบด้วย item ที่ k-itemsets เช่น 3-itemsets {Milk, Bread, Diaper}

\rightarrow Support count คือ ความถี่ของการปรากฏ itemset เช่น {Milk, Bread, Diaper}

→ Support คือ ค่าสนับสนุน หรือ ค่าความจริงของจำนวน

รายการที่มี itemset อยู่ ทั้งนี้เป็นการบ่งบอกถึงคุณภาพของกฎอาทิส ({Milk, Bread, Diaper}) ดังแสดงใน สมการ 1

$$\text{Support}(X) = \frac{\text{จำนวน Transaction ที่ประกอบด้วย Itemset } X}{\text{จำนวน Transaction ทั้งหมดในฐานข้อมูล}}$$

→ Frequent Itemset คือ itemset ที่มีค่าสนับสนุนสูงกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำ (Minimum support) เช่น กำหนดค่า minimum support = 0.5 ดังนั้น itemset ที่ สามารถเป็น frequent itemset ได้ต้องมีค่าสนับสนุน สูง กว่าหรือเท่ากับ 0.5

→ ค่าความเชื่อมั่น (Confidence value) เป็น ค่าแสดงความเป็นจริงของรายการเปรียบเทียบกับจำนวนรายการที่มีทั้งหมด สามารถคำนวณได้ดังสมการ 2

$$\text{Confidence}(x \rightarrow y) = \frac{\text{Support}(x \cup y)}{\text{Support}(x)}$$

มาติดา สองลี และคณะ(2553) ใช้เทคนิคการทำเหมืองข้อมูลแบบ Classification โดยเลือกการทำนายข้อมูลด้วยวิธี Decision Tree และใช้อัลกอริทึมชนิด J48 เพื่อให้ได้ตัวแบบที่มีค่าความถูกต้อง ในการ ทำ น า ย ที่ น ำ เ ชื่ อ ถี อ แล้วนำผลที่ได้มาเขียนโปรแกรมพัฒนา ผลการวิจัยพบว่า จากการแบ่งนักศึกษาออกเป็น 2 กลุ่มคือกลุ่มหลักสูตร 4 ปี และกลุ่มหลักสูตรต่อเนื่อง เมื่อเปรียบเทียบความถูกต้องและความผิดพลาดของการทำนายข้อมูล แสดงให้เห็นว่า ความถูกต้องของการทำนายข้อมูลของนักศึกษากลุ่มหลักสูตรต่อเนื่อง มีความถูกต้องของข้อมูลสอนและข้อมูลทดสอบเพียงร้อยละ 60.66 และ 49.00 [3]

ภัทรพงศ์ พงศ์ภัทรกานต์(2553) เป็นคอมมิตตีแมชชีน คือ SVM ร่วมกับC5.0 มาทดลองกับชุดข้อมูลของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏเลย แล้วนำมาเปรียบเทียบกับ C5.0 และนิรอลเน็ตเวิร์ก ผลการวิจัยพบว่า ประสิทธิภาพในการจำแนกประเภทข้อมูล โดยใช้ SVM ร่วมกับ C5.0 มีค่าสูงที่สุด เท่ากับ 75.32 ส่วน C5.0 มีค่า 70.93 และนิรอลเน็ตเวิร์ก มีค่า 69.92 ตามลำดับ และสามารถวิเคราะห์ ปัจจัยที่มีความสำคัญที่ส่งผลต่อการฟื้นฟูสภาพของนักศึกษาระดับปริญญา

ตรี มหาวิทยาลัยราชภัฏเลยได้ 3 ปัจจัยหลัก คือ ขนาดโรงเรียนเดิม(old_school) ร้อยละ 64.00, อาชีพของบิดา (fat_occup) ร้อยละ 33.00 และ อาชีพของมารดา (mat_occup) ร้อยละ 3.00 [4]

จิระนันต์ เจริญรัตน์(2559) วิเคราะห์ปัจจัยที่ส่งผลต่อการฟื้นฟูสภาพของนักศึกษาระดับปริญญาตรี ที่มีผลการเรียนของเกรดเฉลี่ยสะสมตั้งแต่ 2.00 ขึ้นไป ในการวิเคราะห์ข้อมูลจะใช้ข้อมูลนักศึกษาระดับปริญญาตรี คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏ เลือกการทำนายข้อมูลด้วยวิธี Decision Tree และใช้อัลกอริทึมชนิด J48 การทดสอบโมเดลที่ได้จะทำการทดสอบแบบ 10-fold cross validation โดยใช้โปรแกรม WEKA ผลการวิเคราะห์พบว่าปัจจัยที่มีความสำคัญที่จะส่งผลต่อการฟื้นฟูสภาพของนักศึกษาที่มีผลการเรียนปกติ แบ่งเป็น 4 กลุ่ม ตามกลุ่มเกรดเฉลี่ยสะสม ดังนี้ กลุ่ม Best (GPA>3.50 ขึ้นไป) ปัจจัยคือ วุฒิสถาบันศึกษาเดิม กลุ่ม Excellent (GPA = 3.00-3.50) ปัจจัยคือ อาชีพมารดา และสาขาวิชาที่เรียน กลุ่ม Good (GPA = 2.50-2.99) ปัจจัยคือ ทุนกู้ยืมเพื่อ การศึกษา สถานภาพของครอบครัว รายได้ของบิดา รายได้ของมารดา และจังหวัด กลุ่ม Medium (GPA = 2.00-2.49) ปัจจัยคือ ทุนกู้ยืมเพื่อการศึกษา สถานภาพครอบครัว และรายได้ของมารดา [5]

3. วิธีดำเนินงาน

งานวิจัยนี้เป็นวิจัย Survey Paper โดยค้นหางานวิจัยที่เกี่ยวข้องกับการค้นหาปัจจัยที่มีผลการตกออกและนำมาวิเคราะห์จำนวน 8 ชิ้นงาน

4. ผลการดำเนินงาน

ตารางที่ 1 เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่อเรื่อง	ข้อดี	ข้อเสีย
-------	------------	-------	---------

1	การประยุกต์ใช้ เหมืองข้อมูลเพื่อการ ทำนายสถานภาพของ นักศึกษา วิทยาลัย เทคโนโลยีภาคใต้ [3]	มีการนำ ผลการวิจัยมา พัฒนาระบบ สารสนเทศ เพื่อการทำนาย สถานภาพ นักศึกษา	มีค่าความ แม่นยำที่ต่ำ
2	การวิเคราะห์ปัจจัยที่ ส่งผลต่อการฟื้น สภาพของนักศึกษา ระดับปริญญาตรี โดยใช้ คอมมิตตีแมชชีน [4]	มีการใช้คอม มิตตีแมชชีน (SVM) ร่วมกับ C5.0 ทำให้ มีค่า ความแม่นยำ เพิ่มขึ้น	ไม่ได้นำ ผลงานการ วิจัยมา เปรียบเทียบ กับ แบบสอบถาม ความพึง พอใจหรือ เหตุผลที่ต้อง ฟื้นสภาพ นักศึกษา
3	การวิเคราะห์ปัจจัยที่ ส่งผลต่อการฟื้น สภาพของนักศึกษา ที่มีผลการเรียนปกติ โดยใช้ต้นไม้ ตัดสินใจ [10]	ใช้เทคนิค เหมืองข้อมูล หลายเทคนิค ทำให้มี ประสิทธิภาพ ด้านค่าความ ถูกต้อง	Data Cleaning โดย ใช้วิธีการตัด Missing Data ซึ่งบางที่ ข้อมูลที่ตัด ออกไปอาจมี ผลต่อปัจจัยก็ ได้ การทำ DataCleaning จึงต้อง พิจารณาให้ดี และรอบคอบ

4	การหาปัจจัยที่ส่งผล ต่อความเสี่ยงของ นักศึกษาเรียนอ่อน ด้วยเทคนิคกฎ ความสัมพันธ์ กรณีศึกษา : มหาวิทยาลัยสงขล นครินทร์ [6]	วิจัยขั้นนี้ไปใช้เพื่อ ป้องกัน ไม่ให้เกิด การเรียนอ่อนด้วย การหาความเสี่ยง ของ แล้วให้ คำแนะนำและดูแล	มีชุด ข้อมูลที่ นำมา วิเคราะห์ ค่อนข้าง น้อย
5	การเปรียบเทียบ อัลกอริทึมเหมือง ข้อมูลเพื่อวิเคราะห์ ปัจจัยที่ส่งผลต่อ ระดับผลการ เรียน ของนักศึกษา [3]	สามารถนำมาใช้ เป็นข้อมูลในการ วางแผนดูแล ให้ คำแนะนำสำหรับ นักศึกษาที่คาดว่า จะอยู่ในกลุ่มเสี่ยง ผลการเรียนต่ำ	มีค่า ความ ถูกต้องที่ น้อย
6	การเลือก คุณลักษณะข้อมูลที่ มีผลกระทบต่อการ สำเร็จการศึกษาของ นักศึกษาโดยใช้ วิธีการจำแนกข้อมูล ด้วยโครงข่าย ประสาทเทียม [8]	มีข้อมูลมากที่ นำมาทำการ วิเคราะห์มาก ได้ค่าความถูกต้อง มาก	ไม่ได้ นำไปใช้ งานจริง

ลำดับ	ชื่อเรื่อง	ข้อดี	ข้อเสีย
-------	------------	-------	---------

ลำดับ	ชื่อเรื่อง	ข้อดี	ข้อเสีย
7	การวิเคราะห์ ปัจจัยที่ส่งผลต่อ	ทำให้ได้กฎ ความสัมพันธ์	ผลจากการ วิเคราะห์ปัจจัย

	การฟื้นฟูสภาพของ นักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล [1]	ที่เกี่ยวข้องกับการฟื้นฟูสภาพของนักศึกษา	ที่เกี่ยวข้องกับการฟื้นฟูสภาพมีเพียง 2 ปัจจัยซึ่งส่งผลกระทบต่อวิเคราะห์ได้ไม่ชัดเจน ไม่สามารถนำไปประยุกต์ใช้ได้จริง
8	การวิเคราะห์ความสัมพันธ์ของปัจจัยข้อมูล นักศึกษาที่มีผลต่อระดับผลการเรียนด้วยเทคนิคต้นไม้ตัดสินใจกรณีศึกษา มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ [2]	มีการตั้งสมมติฐานขึ้นมาเพื่อทดสอบกับชุดข้อมูล	ปัจจัยที่ได้ไม่มีผลต่อระดับผลการเรียน

5. สรุป

จากการสำรวจงานวิจัยที่เกี่ยวข้องกับการค้นหาปัจจัยที่มีผลต่อการตกออกของนักศึกษา พบว่า ปัจจัยที่มีผลต่อการตกออกได้แก่ปัจจัย ทนสู้ยิมเพื่อการศึกษา สถานะภาพครอบครัว อาชีพบิดามารดา รายได้บิดามารดา ซึ่งปัจจัยเหล่านี้เป็นปัจจัยส่วนบุคคลที่เกี่ยวข้องกับสถานะทางเศรษฐกิจเป็นหลัก แต่ งานวิจัยชิ้นนี้ก็มีข้อจำกัดอยู่ที่เพียงการศึกษาจากงานวิจัยชิ้นอื่นเท่านั้น งานวิจัยครั้งต่อไปผู้วิจัยจะต้องมีการรวบรวมข้อมูลจริงมาวิเคราะห์ปัจจัยการตกออกเพื่อนำมาเปรียบเทียบกับ การสำรวจงานวิจัยในครั้งต่อไป

เอกสารอ้างอิง

- [1] นางขวัญฤทัย นกแก้ว. การวิเคราะห์ปัจจัยที่ส่งผลต่อการฟื้นฟูสภาพของนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูล; 2555.
- [2] พฤทธิพงษ์ เฟ็งศิริ, พันธนา ก้อนเชื้อรัตน์, ชัชฎา ชวรางกูร, อัจฉราพรรณ คชเดช. การวิเคราะห์ความสัมพันธ์ของปัจจัยข้อมูลนักศึกษาที่มีผลต่อระดับผลการเรียน ด้วยเทคนิคต้นไม้ตัดสินใจ กรณีศึกษานักศึกษามหาวิทยาลัยเทคโนโลยี ราชมงคลสุวรรณภูมิ; 2556.
- [3] มานิตา สองสี, ไชยชนัดต์ ปาละมาณ, วีระพงศ์ วุฒิสักดิ์. การประยุกต์ใช้เหมืองข้อมูลเพื่อการทำนายสถานภาพของนักศึกษา วิทยาลัยเทคโนโลยีภาคใต้; 2553.
- [4] ภัทร์พงศ์ พงศ์ภัทรกานต์. การวิเคราะห์ปัจจัยที่ส่งผลต่อการฟื้นฟูสภาพของนักศึกษา ระดับปริญญาตรี โดยใช้คอมพิวเตอร์แมชชีน; 2553.
- [5] จิระนันต์ เจริญรัตน์. การวิเคราะห์ปัจจัยที่ส่งผลต่อการฟื้นฟูสภาพของนักศึกษาที่มีผลการเรียนปกติโดยใช้ต้นไม้ตัดสินใจ; 2559.
- [6] ณัฐธิดา สุวรรณ โฉ, อ้นธิกา สิงห์เอี่ยม. การหาปัจจัยที่ส่งผลต่อความเสี่ยงของนักศึกษาเรียนอ่อนด้วยเทคนิคกฎความสัมพันธ์กรณีศึกษา : มหาวิทยาลัยสงขลานครินทร์; 2554.
- [7] เขียวภา ภากรสำเร็จ, จิรัฐฐา อนุญอบ, วิรัตน์ พงษ์ศิริ. การเปรียบเทียบอัลกอริทึมเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยที่ส่งผลต่อระดับผลการเรียนของนักศึกษา; 2555.
- [8] ธนาวุฒิ นิลมณี, สมคิด แซ่ หลี, สรเดช ครูทจอน. การเลือกคุณลักษณะข้อมูลที่มีผลกระทบต่อความสำเร็จการศึกษาของนักศึกษาโดยใช้วิธีการ จำแนกข้อมูลด้วยโครงข่ายประสาทเทียม; 2555.
- [9] อคูลย์ ยี่มงาม (2552). การทำเหมืองข้อมูล (Data Mining). ค้นเมื่อ 20 มีนาคม 2552. จาก

http://compcenter.bu.ac.th/index.php?option=com_content&task=view&id=75&Itemid=172

- [10] ประกาศิต ช่างสุพรรณ, "การจำแนกระดับความมั่นใจของผู้เรียน โดยใช้เทคนิคต้นไม้ตัดสินใจ," มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ, 2553