

# การประยุกต์ใช้แบบจำลองการพยากรณ์ปัจจัยที่มีผลต่อ การสำเร็จการศึกษาของนิสิตคณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม โดยการทำเหมืองข้อมูล Utilization of Rule-based Predicting Models Faculty of Architecture, Urban Design and Creative Arts Students at Mahasarakham University by Data Mining

วารินทร์ ปัญญาวงษ์

นักวิชาการคอมพิวเตอร์ คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์  
มหาวิทยาลัยมหาสารคาม อำเภอกันทรวิชัย จังหวัดมหาสารคาม 44150

Wararin Panyawong

Computer Technical Officer, Affiliation, Address, City, Province, Postal Code  
Faculty of Architecture, Urban Design and Creative Arts,  
Maha Sarakham University, Kantarawichai District, Maha Sarakham, Thailand, 44150

Email: Wararin.p@msu.ac.th

## บทคัดย่อ

งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อการเลือกเรียนต่อระดับปริญญาตรี และสร้างแบบจำลองในการพยากรณ์การสำเร็จการศึกษาระดับปริญญาตรี โดยข้อมูล ถูกรวบรวมจากคณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม 6 หลักสูตร ได้แก่สถาปัตยกรรม สถาปัตยกรรมผังเมือง ภูมิสถาปัตยกรรม การจัดการงานก่อสร้าง สถาปัตยกรรมภายใน นฤมิตศิลป์ ในปีการศึกษา 2557-2559 จำนวน 929 ระเบียบ ในการนี้เทคนิค Decision Table, Decision Rule, Repeated Incremental Pruning to Produce Error Reduction (RIPPER) และ OneR ได้ถูกนำมาใช้ในการสร้างแบบจำลองเพื่อการพยากรณ์ ในการวัดประสิทธิภาพแบบจำลอง 10-fold cross validation ได้ถูกนำมาใช้ในการแบ่งข้อมูลออกเป็น ชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ โดยประสิทธิภาพของแบบจำลองที่สร้างจากแต่ละเทคนิคใช้ ค่าความเที่ยงตรง (Precision) ค่าความระลึก (Recall) และ ค่า F-measure จากการศึกษาพบว่าปัจจัยที่มีผลต่อการเลือกเรียนต่อระดับปริญญาตรี 5 ลำดับแรก คือ อำเภอ ประเภทวิชา เพศ เกรดเฉลี่ย กลุ่มสุขศึกษาพลศึกษา และอาชีพมารดา ส่วนการสร้างแบบจำลองเทคนิค RIPPER สามารถสร้างแบบจำลองการพยากรณ์ที่มีประสิทธิภาพมากกว่า Decision Table, Decision Rule, and OneR ซึ่งมีความถูกต้อง (Precision) เท่ากับ ร้อยละ 91 ค่าความระลึก (Recall) เท่ากับ ร้อยละ 91.10 และค่า F-measure เท่ากับ ร้อยละ 90.90 ต้นแบบกฎการจำลองนี้สามารถ

นำไปใช้ในการพัฒนาระบบหรือโปรแกรมการพยากรณ์การสำเร็จการศึกษาระดับปริญญาตรี คณะสถาปัตยกรรมศาสตร์  
ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคามได้

**คำสำคัญ:** แบบจำลองกฎการพยากรณ์, เหมืองข้อมูล, แบบจำลองการพยากรณ์ตามกฎ

## Abstract

This research are to study factors influencing selecting graduation bachelor and to create rule-based prediction models of graduation bachelor. The data were collected from 6-course education institutions in Faculty of Architecture, Urban Design and Creative Arts Students at Mahasarakham University from year 2014-2016. The data consist of 929 records. In this paper, Decision Table, Decision Rule, Repeated Incremental Pruning to Produce Error Reduction (RIPPER) and OneR are used for building the rule-base prediction models. In order to measure the performance of the rule-base prediction models, 10-fold cross validation is applied to divide the data set into learning and testing sets. The performance of the prediction models created from each technique is measured with precision, recall and F-measure. The experimental results revealed that Factors influencing selecting graduation bachelor of the top five as follows: 1) district 2) Type of course 3) gender 4) grade point average of health and physical education and 5) mother's career. For building prediction model, RIPPER is superior to Decision Table, Decision Rule, and OneR. It can create effective rule-based prediction models with 91% of precision, 91.10% of recall and 90.90% of F-measure.

**Keywords:** Rule-based Prediction Models, Data Mining, Rule-based Prediction Models.

**Received:** March 4, 2020; **Revised:** March 27, 2020; **Accepted:** August 17, 2020

## 1. บทนำ

คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม เป็นหน่วยงานหนึ่งที่มีการผลิตบัณฑิต  
ด้านงานสถาปัตยกรรมและสาขาที่เกี่ยวข้องกับการออกแบบต่าง ๆ ซึ่งได้จัดการเรียนการสอนใน 2 ระดับการศึกษา  
คือ ระดับปริญญาตรี และปริญญาโท โดยการเรียนการสอนในระดับปริญญาบัณฑิต (ปริญญาตรี) ใช้ระบบทวิภาค คือ แบ่ง  
ปีการศึกษาหนึ่งออกเป็น 2 ภาคการศึกษา ได้แก่ ภาคการศึกษาต้นและภาคการศึกษาปลาย (หรืออาจมีภาคฤดูร้อนก็ได้)  
หลักสูตรระดับปริญญาบัณฑิตเป็น 6 หลักสูตร ได้แก่ สถาปัตยกรรม ภูมิสถาปัตยกรรม สถาปัตยกรรมผังเมือง สถาปัตยกรรม  
ภายใน นฤมิตศิลป์ สถาปัตยกรรมการจัดการงานก่อสร้าง ทั้ง 6 หลักสูตรได้รับการรับรองจากสภามหาวิทยาลัยมหาสารคาม  
สถาบันอุดมศึกษามีข้อมูลที่ถูกรวบรวมอยู่เป็นจำนวนมากข้อมูลต่าง ๆ เพิ่มมากขึ้นตาม การเติบโตควบคู่กับความทันสมัย  
ของเทคโนโลยีประกอบกับมีฐานข้อมูลขนาดใหญ่และมีข้อมูล สารสนเทศภายในมีแฝงองค์ความรู้สำคัญซ่อนอยู่ จากความ  
ก้าวหน้าของเทคโนโลยีจึงเกิดศาสตร์ด้าน ปัญญาประดิษฐ์ (Artificial Intelligence) ร่วมกับวิวัฒนาการข้อมูลโดย Group  
(2010) อ้างถึงใน สายชล สินสมบูรณ์ทอง (2558) กล่าวว่า การทำเหมืองข้อมูลจะเป็นกระบวนการค้นหาความสัมพันธ์ รูปแบบ  
หรือกฎเกณฑ์ โดยนำข้อมูลปริมาณมากจากแหล่งจัดเก็บข้อมูลขนาดใหญ่ที่เก็บไว้นานมาเข้าสู่ กระบวนการวิธีทาง  
คณิตศาสตร์ และสถิติมาช่วยในการวิเคราะห์ข้อมูลและจัดการข้อมูล การทำ เหมืองข้อมูลนั้นมีเทคนิคที่สามารถประยุกต์  
ใช้ได้มากมายโดยได้มีงานวิจัยด้านการศึกษามาจากข้อมูลที่มี อยู่ในสถาบันอุดมศึกษา เช่น งานวิจัยของ สมฤทัย กลัดแก้ว  
อรไท ชั่วเจริญ และชานัญ เจริญรุ่งเรือง (2557) ได้ศึกษาปัจจัยที่มีผลต่อการตัดสินใจเลือกตำแหน่งงานของนักศึกษา  
มหาวิทยาลัยราชภัฏหมู่บ้านจอมบึงจากข้อมูลภาวะมีงานทำของบัณฑิตที่สำเร็จการศึกษาปีการศึกษา 5 พ.ศ. 2555 – 2557

จำนวน 1,933 คน โดยจำแนกประเภทข้อมูล (Classification) ด้วยเทคนิค Decision Tree เปรียบเทียบกับเทคนิค Logistic Regression พบว่าแบบจำลองที่ใช้เทคนิค Decision Tree ให้ ความถูกต้องสูงสุดที่ร้อยละ 57.37 และงานวิจัยของ เสกสรรค์ วิสัยลักษณ์ วิชา เจริญภัณฑารักษ์ และ ดวงดาว วิชาตากุล (2558) ได้พัฒนาแบบจำลองทำนายผลการเรียนนักเรียน โรงเรียนสาธิตแห่ง มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา จากข้อมูลนักเรียน ระดับมัธยมศึกษาปีที่ 4 ระหว่างปีการศึกษา 2553 – 2556 จำนวน 525 ระเบียบ และ 16 คุณลักษณะ พบว่า การคัดเลือกคุณลักษณะด้วยชุดข้อมูลแบบไม่จัดกลุ่มวิธี Correlation – based Feature Selection ร่วมกับเทคนิค Neural Network ให้ค่าความถูกต้องสูงสุดที่ร้อยละ 94.48 และมีค่าความ คลาดเคลื่อนสมบูรณ์เฉลี่ยที่ 0.1880 ซึ่งมีความเหมาะสมที่สุด ส่วนงานวิจัยของ ชุตินา อุตมะมุณี และประสงค์ ปราณีตพลกรัง, (2553) ได้หาปัจจัยที่มีผลต่อการตัดสินใจเลือกสาขาวิชาเรียนของ นักศึกษาระดับอุดมศึกษา จากข้อมูลนักศึกษาระดับอุดมศึกษามหาวิทยาลัยภาครัฐ และเอกชน จำนวน 9 แห่ง โดยใช้เทคนิค Decision Tree และเปรียบเทียบเทคนิค Bayesian Networks เพื่อ พัฒนาแบบจำลองทำนายระบบตัดสินใจเลือกสาขาวิชาเรียน พบว่าเทคนิค Bayesian Networks มี ค่าความแม่นยำสูงสุดร้อยละ 91.35

การทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการค้นหาความรู้ที่มีอยู่ในฐานข้อมูลขนาดใหญ่ ซึ่งเทคโนโลยี ที่เข้ามาช่วยในการเปลี่ยนแปลงข้อมูลเหล่านั้นให้เป็นข้อมูลที่สามารถนำไปประยุกต์ใช้งานด้านต่าง ๆ ได้ โดยที่องค์กรต่าง ๆ มักจะมองข้ามความสำคัญของข้อมูลที่เก็บไว้ นักวิจัยหลายท่าน ได้นำเอากระบวนการของเหมืองข้อมูลมาใช้ในการสร้างแบบจำลองเพื่อการพยากรณ์ (เสกสรรค์ วิสัยลักษณ์ วิชา เจริญภัณฑารักษ์ และดวงดาว วิชาตากุล, 2558) มีการเสนอการวิเคราะห์ข้อมูลจากแบบสำรวจสุขภาพประชาชนโดยใช้ข้อมูลของ ประชาชนในเขตเทศบาลเมืองจังหวัดบุรีรัมย์ ปี พ.ศ. 2555 โดยในการศึกษาการคัดกรองสุขภาพเบื้องต้นของประชาชน พบว่าอัลกอริทึม Partial Rules ให้ความถูกต้องมากกว่าอัลกอริทึมต้นไม้ ตัดสินใจ C4.5 โดยมีค่าความถูกต้อง (precision) เท่ากับ ร้อยละ 88.60 ค่าระลึก (recall) เท่ากับ ร้อยละ 89.20 และค่าความเหวี่ยง (F-measure) เท่ากับ ร้อยละ 88.80

ดังนั้นผู้วิจัยจึงมีแนวคิดศึกษาปัจจัยที่มีผลต่อการสำเร็จการศึกษาในระดับปริญญาตรีคณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม และสร้างแบบจำลองการพยากรณ์การสำเร็จการศึกษาระดับปริญญาตรี ด้วยเทคนิค Decision Table, Decision Rule, RIPPER และ OneR เพื่อให้คณะอาจารย์และผู้บริหารได้นำไปประยุกต์ใช้ในการวิเคราะห์ความสำเร็จการศึกษานิสิต แนะนำแนวการเข้าศึกษาต่อระดับปริญญาตรี ในสถานศึกษาคณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคามได้อย่างแม่นยำที่สุด

## 2. แนวคิดที่เกี่ยวข้อง

สมฤทัย กลัดแก้ว (2557) ได้วิจัย “ระบบสนับสนุนการตัดสินใจการเลือก ตำแหน่งงานให้สอดคล้องกับความสามารถของบัณฑิต” เพื่อศึกษาปัจจัยที่มีผลต่อการเลือกตำแหน่ง งานที่สอดคล้องกับความสามารถบัณฑิตด้วยเทคนิค Logistic Regression Analysis โดยการจำแนก ประเภทข้อมูลด้วยเทคนิค Decision Tree และเปรียบเทียบความถูกต้องการทำนายระหว่างเทคนิค Logistic Regression Analysis และการจำแนกประเภทข้อมูลด้วยเทคนิค Decision Tree จากนั้นได้ทำการพัฒนาแบบจำลองช่วยในการตัดสินใจเลือกตำแหน่งงานให้สอดคล้องกับความสามารถของ บัณฑิต จากข้อมูลภาวะมีงานทำบัณฑิตที่สำเร็จการศึกษา 2555 – 2557 จำนวน 1,933 ระเบียบ จำนวน 2,825 คน เป็นกลุ่มตัวอย่าง และได้รวบรวมแบบสำรวจข้อมูลภาวะมีงานทำของศูนย์ 31 คอมพิวเตอร์ มหาวิทยาลัยราชภัฏหมู่บ้านจอมบึงเป็นเครื่องมือที่ใช้ศึกษา และได้แบ่งตำแหน่งงาน หรือกลุ่มอาชีพเป็น 8 กลุ่ม ผลการวิจัยพบว่าปัจจัยจาก 12 คุณลักษณะ มีผลต่อการเลือกตำแหน่ง งานหรือกลุ่มอาชีพมี 4 คุณลักษณะ คือ เพศ คุณวุฒิการศึกษาที่สำเร็จการศึกษา สาขาวิชาที่เรียน และ ความสามารถพิเศษ ซึ่งมีระดับนัยสำคัญทางสถิติอยู่ที่ 0.05 และจากการวัดค่าประสิทธิภาพตัว แบบด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ซึ่ง ได้วิเคราะห์ผลเปรียบเทียบค่าความถูกต้องของการพยากรณ์เทคนิคการจำแนกประเภทข้อมูลพบว่า เทคนิค Decision Tree อยู่ที่ร้อยละ 57.37 โดยมีค่ามากกว่าเทคนิค Logistic Regression Analysis เล็กน้อยที่ร้อยละ 56.30 จึงได้พัฒนาแบบจำลองระบบสนับสนุนการตัดสินใจ การเลือกตำแหน่งงานให้ สอดคล้องกับความสามารถของบัณฑิตโดยใช้เงื่อนไข 4 คุณลักษณะดังกล่าวมาสร้างเป็นแบบจำลองระบบโดยออกแบบส่วนติดต่อประสานงานผู้ใช้ด้วย Microsoft Visual Basic

เสกสรรค์ วิสัยลักษณ์ วิชา เจริญภัณฑารักษ์ และดวงดาว วิชาตากุล, (2558) ได้วิจัย “การใช้เทคนิคการทำเหมืองข้อมูล เพื่อพยากรณ์ผลการเรียนของนักเรียนโรงเรียนสาธิตแห่ง มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสนศูนย์วิจัย และพัฒนาการศึกษา” เพื่อพัฒนาคัดเลือกข้อมูลและสร้างแบบจำลองทำนายผลการเรียนของนักเรียนโรงเรียนสาธิตแห่ง มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสนศูนย์วิจัยและพัฒนาการศึกษา โดยใช้ข้อมูลนักเรียน ระดับมัธยมศึกษา ตอนต้น ระหว่างปีการศึกษา 2548 – 2556 นำมาพัฒนาคัดเลือกข้อมูลด้วยโครงสร้าง แบบ Snowflake Schema จากนั้นใช้ ข้อมูลนักเรียนระดับมัธยมศึกษาตอนปลาย ระหว่างปี การศึกษา 2556 – 2556 จำนวน 525 ระเบียบ ประกอบด้วย 16 คุณลักษณะ นำมาพัฒนา แบบจำลองทำนาย โดยใช้ชุดข้อมูล 2 แบบ ได้แก่แบบไม่จัดกลุ่ม (Original Data) และแบบ จัดกลุ่ม (Cluster Data) จากนั้นทำการคัดเลือกคุณลักษณะด้วยวิธี Correlation-based Feature Selection และวิธี Information Gain แล้วใช้เทคนิค Neural Network แบบ Multi-Layer Perceptron เทคนิคSupport Vector Machine และเทคนิค Decision Tree มาพัฒนาแบบจำลอง 32 ทำนายพร้อมเปรียบเทียบและวัดประสิทธิภาพด้วยวิธี 10-fold Cross Validation พบว่าคัดเลือกข้อมูลที่พัฒนาขึ้นผู้ใช้งานมีความพึงพอใจอยู่ในระดับดีและจากการเปรียบเทียบแบบจำลอง ทำนายโดยชุดข้อมูลแบบไม่จัดกลุ่ม (Original Data) ที่คัดเลือกด้วยวิธี Correlation-based Feature Selection ร่วมกับ เทคนิค Neural Network แบบ Multi-Layer Perceptron จำนวน 5 คุณลักษณะ คือ แผนการเรียน ผลการเรียนเฉลี่ย วิทยาศาสตร์ ผลการเรียนเฉลี่ยสังคมศึกษา ผลการเรียนเฉลี่ย ภาษาอังกฤษ และผลการเรียนเฉลี่ยระดับชั้นมัธยมศึกษาปีที่ 3 มีความเหมาะสมมากที่สุด โดยมีค่าความ ถูกต้องสูงสุดที่ร้อยละ 94.48 และมีค่ารากที่สองของความคลาดเคลื่อน (RMSE) น้อยที่สุดที่ 0.1880 จากนั้นนำมาพัฒนาระบบพยากรณ์ผลการเรียนด้วยภาษา PHP

นิภาพร ชนะมาร และพรณี สิทธิเดช (2557) ได้วิจัย “การวิเคราะห์ปัจจัย การเรียนรู้ด้วยการคัดเลือกคุณสมบัติ และการพยากรณ์” เพื่อประยุกต์ใช้เทคนิคเหมืองข้อมูลทำนาย ผลสัมฤทธิ์ทางการเรียนของนิสิตระดับปริญญาตรี สาขา วิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัย นครสวรรค์ จำนวน 180 ระเบียบ ประกอบด้วยคุณสมบัติ 23 คุณลักษณะ แบ่งเป็น ตัวแปรอิสระ 22 คุณลักษณะ ได้แก่ ข้อมูลภูมิหลังและข้อมูลผลการเรียนรายวิชาในชั้นปีที่ 1 และ 2 ตัวแปรตามหรือตัว แปรทำนายคือเกรดเฉลี่ยเมื่อสำเร็จการศึกษา จากนั้นวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือก คุณลักษณะที่สำคัญ 3 วิธี ได้แก่ (1) วิธีCorrelation-based (2) Consistency-based และ (3) วิธี Gain Ratio จากนั้นนำมาพัฒนาแบบจำลอง การทำนาย ด้วยเทคนิค Neural Network แบบ Backpropagation และเทคนิคSupport Vector Machines ร่วมกับการ วัดประสิทธิภาพด้วยวิธี 10-fold Cross Validation และวัดค่ารากที่สองของความคลาดเคลื่อน (RMSE) พบว่าข้อมูลภูมิ หลังไม่ใช่ ข้อมูลสำคัญในการทำนายผลสัมฤทธิ์ทางการเรียน และผลการเรียนรายวิชาจำนวน 10 คุณลักษณะ นั้นเป็นตัวแปร สำคัญ สำหรับแบบจำลองทำนายด้วยเทคนิค Neural Network แบบ Backpropagation และเทคนิคSupport Vector Machines จากคุณลักษณะสำคัญดังกล่าวมีค่าความ ผิดพลาดอยู่ในระดับต่ำกว่าแบบจำลองทำนายที่ใช้ตัวแปรตั้งต้นจำนวน 22 คุณลักษณะ ในขณะที่ เทคนิคการรวมกลุ่มประจำแนกประเภทด้วยวิธี Bagging ร่วมกับเทคนิค Neural Network แบบ Back-propagation และเทคนิค Support Vector Machines พบว่าผลการพยากรณ์ของ Bagging ร่วมกับเทคนิค Neural Network แบบ Back-propagation มีค่ารากที่สองของความคลาดเคลื่อน (RMSE) อยู่ระดับต่ำสุดที่ 0.1051 มีประสิทธิภาพดีที่สุดจึงนำไปใช้พยากรณ์

Osiris Villacampa (2015) ได้วิจัยเรื่อง “Feature Selection and Classification Methods for Decision Making: A Comparative Analysis” เพื่อศึกษาการประมวลผลการทำเหมืองข้อมูลเพื่อลดมิติข้อมูลด้วยการคัดเลือกคุณลักษณะ หรือการเลือก คุณลักษณะซึ่งจะช่วยให้การทำเหมืองข้อมูลมีความถูกต้องและความมีประสิทธิภาพ โดยใช้ข้อมูล ประวัติ การบริการและการขายรถจากตัวแทนจำหน่ายรถ จำนวน 15,417 ระเบียบ จำนวน 40 คุณลักษณะ และข้อมูลประวัติ การเงินของลูกค้าจากธนาคาร จำนวน 10,578 ระเบียบ จำนวน 17 คุณลักษณะ เพื่อหารูปแบบจำแนกประเภทลูกค้าที่มีฐานะ อยู่ในกลุ่มผู้ซื้อรถใหม่ โดยนำเทคนิคการคัดเลือกคุณลักษณะแบบ Filters, Wrappers และ Hybrid ได้แก่วิธี Information Gain วิธี Correlation Based Feature Selection วิธี Relief-F และวิธี Wrappers นำมาใช้เพื่อลดจำนวน คุณลักษณะ ในชุดข้อมูลจากนั้นทำการจำแนกประเภทและพัฒนาแบบจำลองด้วยเทคนิค Decision Tree เทคนิคk-Nearest Neighbor และเทคนิค Support Vector Machines ทดสอบ ประสิทธิภาพแบบจำลองด้วยวิธี 5-fold Cross Validation และวิเคราะห์ เปรียบเทียบการคัดเลือก คุณลักษณะและความ แตกต่างด้วยค่า Accuracy, ค่า Area Under Receiver Operating Characteristic Curve (AUC), ค่า F-Measure, ค่า TP Rate และค่า FP Rate พบว่าค่าความถูกต้องของเทคนิค Decision

Tree ร่วมกับการคัดเลือกคุณลักษณะแบบ Wrapper ที่ค่าความเชื่อมั่น 0.10% มีความถูกต้องแม่นยำถึง 86.4% ในการจำแนกประเภท/การคัดเลือกคุณลักษณะแบบเดียวกันนี้ ยังให้ค่า AUC สูงสุดที่ 90.4% ส่วนเทคนิค k-Nearest Neighbor และเทคนิค Support Vector Machines มีค่าความถูกต้องเพียง 84.9% และ 79.8% ตามลำดับ

### 3. วิธีการศึกษา

ดำเนินการวิจัยในงานวิจัยได้นำเอาขั้นตอนการทำเหมืองข้อมูลมาใช้ เพื่อสร้างระบบ พยากรณ์การสำเร็จการศึกษาระดับปริญญาตรี คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม ซึ่งมีขั้นตอนการดำเนินการ ดังนี้ การเตรียมข้อมูล การสร้างแบบจำลอง และการวัดประสิทธิภาพแบบจำลอง

#### 3.1 การเตรียมข้อมูล

การเตรียมข้อมูลของการศึกษาค้นคว้าอิสระครั้งนี้ ได้ข้อมูลมาจากการสำรวจข้อมูลจำนวน นิสิตระดับปริญญาตรี จากระดับปริญญาตรีในสถานศึกษา คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม ประจำปีการศึกษา 2557-2559 จำนวน 5 สาขา ได้แก่ สาขาภูมิสถาปัตยกรรม สาขาสถาปัตยกรรม สาขาสถาปัตยกรรมผังเมือง

สาขาสถาปัตยกรรมภายใน และนฤมิตศิลป์

ตารางที่ 1 จำนวนนิสิตคณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ ประจำปีการศึกษา 2557-2559

| สาขาวิชา               | จำนวนนิสิต ปีการศึกษา<br>2557 (คน) | จำนวนนิสิต ปีการศึกษา<br>2558 (คน) | จำนวนนิสิต ปีการศึกษา<br>2559 (คน) |
|------------------------|------------------------------------|------------------------------------|------------------------------------|
| 1. ภูมิสถาปัตยกรรม     | 26                                 | 46                                 | 35                                 |
| 2. สถาปัตยกรรม         | 68                                 | 84                                 | 55                                 |
| 3. สถาปัตยกรรมผังเมือง | 64                                 | 77                                 | 69                                 |
| 4. สถาปัตยกรรมภายใน    | 58                                 | 71                                 | 61                                 |
| 5. นฤมิตศิลป์          | 48                                 | 63                                 | 46                                 |
| รวม                    | 264                                | 341                                | 266                                |
| <b>รวมทั้งสิ้น</b>     | <b>871 คน</b>                      |                                    |                                    |

## ตารางที่ 2 การจำแนกชนิดของตัวแปร

| ชื่อตัวแปร                | คำอธิบายตัวแปร                      | ระดับการวัดผล |
|---------------------------|-------------------------------------|---------------|
| 1. Sex                    | เพศ                                 | Nominal       |
| 2. Age                    | อายุ                                | Ratio         |
| 3. Number of siblings     | จำนวนพี่น้อง                        | Ratio         |
| 4. Father Status          | สถานภาพบิดา                         | Nominal       |
| 5. Father Career          | อาชีพบิดา                           | No            |
| 6. Mother Status          | สถานภาพมารดา                        | Nominal       |
| 7. Mother Career          | อาชีพมารดา                          | No            |
| 8. Amphoe                 | อำเภอ                               | Nominal       |
| 9. Thai                   | เกรดกลุ่มภาษาไทย                    | Ratio         |
| 10. Math                  | เกรดกลุ่มคณิตศาสตร์                 | Ratio         |
| 11. Science               | เกรดกลุ่มวิทยาศาสตร์                | Ratio         |
| 12. Social                | เกรดกลุ่มสังคมศึกษาศาสนาและวัฒนธรรม | Ratio         |
| 13. Health                | เกรดกลุ่มสุขศึกษา พลศึกษา           | Ratio         |
| 14. Art                   | เกรดกลุ่มศิลปะ                      | Ratio         |
| 15. Career and Technology | เกรดกลุ่มการงานอาชีพและเทคโนโลยี    | Ratio         |
| 16. Foreign language      | เกรดกลุ่มภาษาต่างประเทศ             | Ratio         |
| 17. Type of course        | ประเภทวิชาที่จบ                     | Nominal       |
| 18. Graduate              | จบการศึกษา                          | Nominal       |

### 3.2 การสร้างแบบจำลอง

การสร้างแบบจำลองพยากรณ์การสำเร็จการศึกษา เพื่อช่วยในการตัดสินใจว่าเมื่อเข้าศึกษาต่อแล้วจะสามารถสำเร็จการศึกษาได้หรือไม่ด้วยตัวแปรทั้ง 18 ตัวแปรที่ได้จากการสำรวจข้อมูลจำนวนนักเรียนในปีการศึกษา 2557-2559 นั้น ด้วยการทำให้เหมือนข้อมูลด้วยโปรแกรม Weka 3.8.1 และเทคนิคที่นำมาใช้ในการสร้าง แบบจำลองมีจำนวน 4 เทคนิค คือ 1) เทคนิค Decision Table 2) เทคนิค Decision Rule 3) เทคนิค RIPPER 4) เทคนิค OneR

### 3.3 การวัดประสิทธิภาพของแบบจำลอง

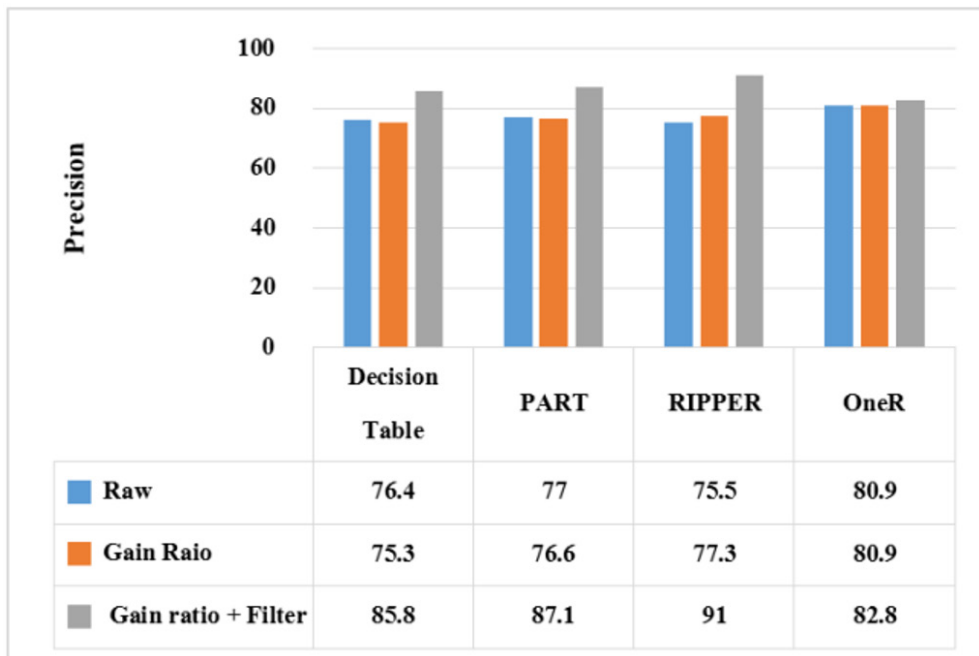
ในการวัดประสิทธิภาพของแบบจำลองนั้น โดยการนำข้อมูลที่เตรียมไว้ ทดสอบทั้งหมด 929 ระเบียบ ผู้วิจัยได้ทำการทดลองแบ่งชุดข้อมูลเป็น 2 ชุด คือ ข้อมูลชุดสอน และข้อมูลชุดทดสอบ ด้วยหลักการ 10-fold cross validation เพื่อให้ข้อมูลทุกตัวมีโอกาส เป็นชุดทดสอบและชุดสอน โดยแบ่งข้อมูลออกเป็น ชุดข้อมูลสอน (Training Data) และข้อมูลทดสอบ (TestingData) โดยแบ่งข้อมูลออกเป็น 10 ส่วนเท่า ๆ กัน ใช้ 9 ส่วนเป็นชุด ข้อมูลสอน และอีก 1 ส่วนเป็นชุดข้อมูลทดสอบ ซึ่งจะทำการสลับกันจนครบทั้งหมด 10 รอบ ในส่วนการวัดประสิทธิภาพของตัวแบบในงานวิจัยนี้ได้วัดจาก ค่าความแม่นยำ ค่าความไว และค่าความจำเพาะ

## 4. ผลการวิจัย

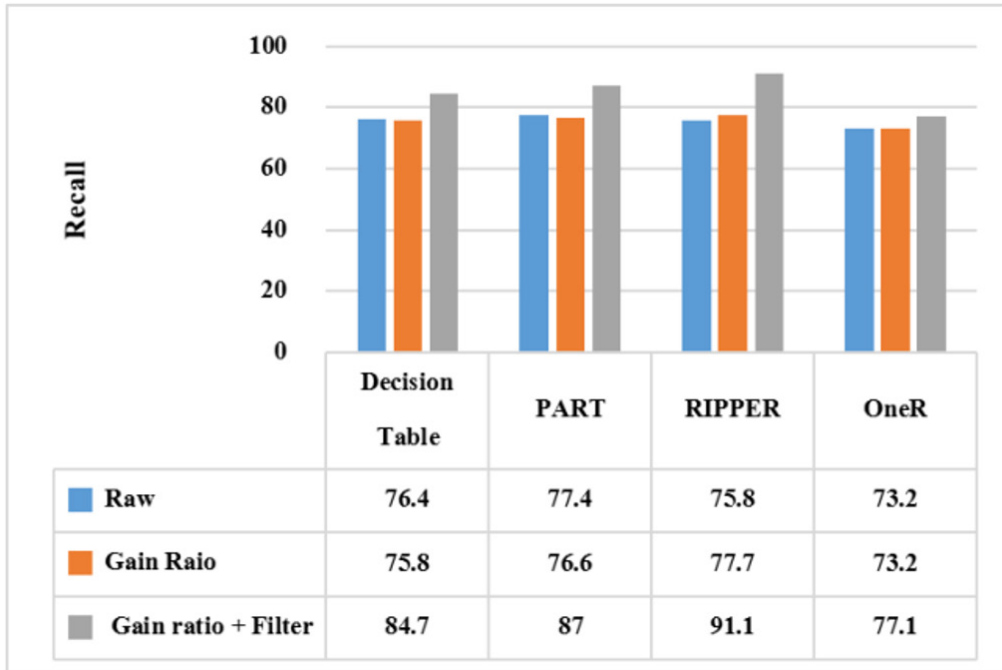
ในการดำเนินการศึกษาค้นคว้าครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาปัจจัยที่มีผลต่อการเลือกเรียนต่อ สำเร็จการศึกษาระดับปริญญาตรี ในสถานศึกษา คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม และเพื่อสร้าง แบบจำลองในการพยากรณ์การสำเร็จการศึกษาระดับปริญญาตรี ในสถานศึกษา คณะสถาปัตยกรรมศาสตร์ ผังเมือง และนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม พยากรณ์ (rule based) ที่มีประสิทธิภาพในการพยากรณ์ โดยมีผลการวิเคราะห์ปัจจัยที่มีผลต่อการ เลือกเรียนต่อระดับประกาศนียบัตรวิชาชีพชั้นสูง และผลการวัดประสิทธิภาพแบบจำลองทั้ง 4 เทคนิค

### 4.1 ผลการวัดประสิทธิภาพแบบจำลองทั้ง 4 เทคนิค

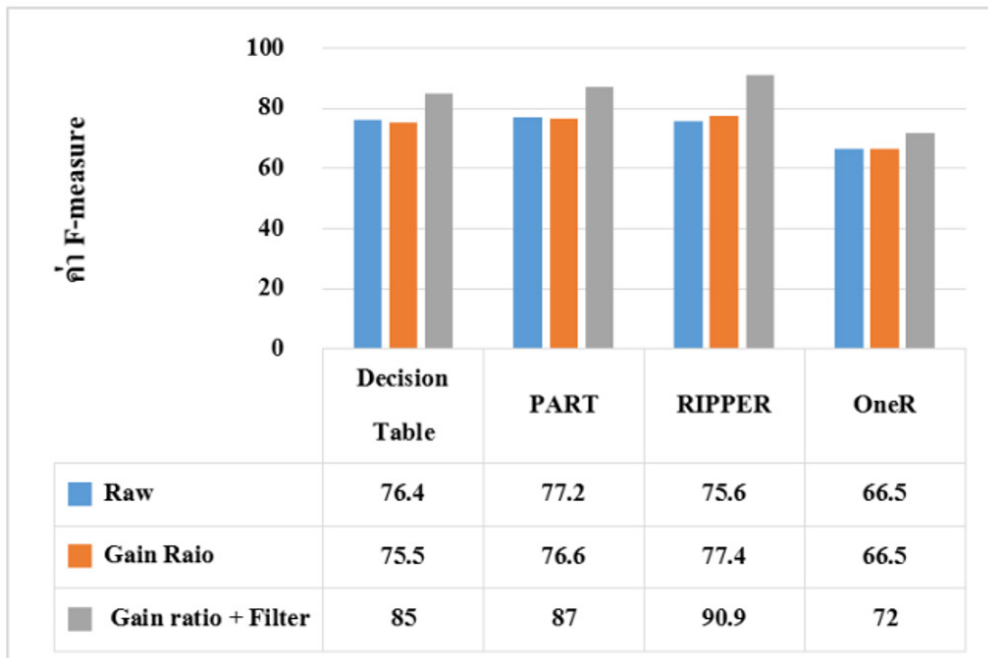
การวัดประสิทธิภาพของแบบจำลองนี้คณะผู้วิจัยได้นำเอาโปรแกรม WEKA เวอร์ชัน 3.8.1 มาเป็นเครื่องมือมาใช้ในการสร้างแบบจำลองด้วยเทคนิค ใช้เทคนิค Decision Table, Decision Rule, RIPPER และ OneR โดยใช้ Gain Ratio และ Filter ในการคัดเลือกแอตทริบิวต์ การวัดประสิทธิภาพ ของแบบจำลองด้วย ค่าความถูกต้อง (Precision) ค่าความระลึก (Recall) และ F-measure ได้ผลการทดลอง



ภาพที่ 1 ค่า Precision ของกฎการพยากรณ์



ภาพที่ 2 ค่า Recall ของกฎการพยากรณ์



ภาพที่ 3 ค่า F-measure ของกฎการพยากรณ์



=== Classifier model (full training set) ===

JRIP rules:

=====

```
(amper = 5) => education=2 (40.0/0.0)
(amper = 1) and (type = 1) and (health = 2) => education=2 (49.0/0.0)
(amper = 1) and (type = 1) and (art = 1.5) and (sci = 2.5) => education=2 (22.0/3.0)
(amper = 8) and (health = 2.5) and (thai = 2.5) => education=2 (19.0/1.0)
(amper = 9) => education=2 (23.0/0.0)
(amper = 1) and (type = 1) and (sci = 2) and (health = 1.5) => education=2 (15.0/0.0)
(amper = 8) and (health = 2.5) and (career and tech = 2.5) => education=2 (16.0/5.0)
(amper = 8) and (health = 2) and (thai = 2) => education=2 (7.0/0.0)
(amper = 1) and (type = 1) and (art = 1.5) and (health = 2.5) => education=2 (8.0/0.0)
(amper = 8) and (career and tech = 2) and (health = 2) => education=2 (15.0/5.0)
(amper = 1) and (type = 1) and (MotherCareer = 4) and (art = 2.5) => education=2 (19.0/5.0)
(amper = 1) and (type = 1) and (sex = 1) and (MotherCareer = 5) => education=2 (6.0/1.0)
(amper = 1) and (type = 1) and (FatherCareer = 5) and (art = 2) and (health = 2.5) => education=2 (6.0/1.0)
(amper = 8) and (health = 2.5) and (social = 2.5) and (sci = 1.5) => education=2 (4.0/0.0)
=> education=1 (556.0/19.0)
```

Number of Rules : 15

#### ภาพที่ 4 แสดงให้เห็นว่า แบบจำลองที่สร้างจากเทคนิค RIPPER

ภาพที่ 4 แสดงให้เห็นว่า แบบจำลองที่สร้างจากเทคนิค RIPPER ได้ถูกสร้างกฎ จำนวน 15 กฎ ซึ่งมีการแปลงกฎของ การพยากรณ์ดังนี้ ตัวอย่างการแปลงกฎการพยากรณ์

(amper=5) => education=2 (40.0/0.0) พยากรณ์ว่านักเรียนคนนี้จะไม่จบการศึกษา ด้วยค่า ความถูกต้องที่ ร้อยละ 100

(amper=1) and (type=1) and (health=2) => education = 2 (49.0/00) ถ้านักเรียน เรียนในสาขาพาณิชยกรรม และเรียนกลุ่มสุขศึกษา พลศึกษา ได้เกรดระดับ 2 พยากรณ์ว่านักเรียนคนนี้จะไม่จบการศึกษา ด้วยค่าความถูกต้องที่ ร้อยละ 100

(amper=1) and (type=1) and (art=1.5) and (sci=2.5) => education = 2 (22.0/3.0) ถ้านักเรียน เรียนในสาขา พาณิชยกรรม และเรียนกลุ่มวิชาศิลปะ ได้เกรดระดับ 1.5 กลุ่มวิชาวิทยาศาสตร์ได้เกรดระดับ 2.5 พยากรณ์ว่านักเรียนคน นี้จะไม่จบการศึกษา ด้วยค่าความถูกต้องที่ร้อยละ 88

(amper=8) and (health=2.5) and (thai=2.5) => education = 2 (19.0/1.0) ถ้านักเรียนเรียนกลุ่มสุขศึกษา พลศึกษา ได้เกรดระดับ 2.5 กลุ่มวิชาภาษาไทยได้เกรดระดับ 2.5 พยากรณ์ว่านักเรียนคนนี้จะไม่จบการศึกษาด้วย ค่าความถูกต้องที่ ร้อยละ 95

(amper=9) => education=2 (23.0/0.0) พยากรณ์ว่านักเรียนคนนี้จะไม่จบการศึกษาด้วย ค่าความถูกต้องที่ ร้อยละ 100

## 5. อภิปรายผล

ผลการวัดประสิทธิภาพแบบจำลองด้วยค่าความเที่ยงตรง (Precision) ทั้ง 4 เทคนิคจากการวิเคราะห์ปัจจัยด้วยเทคนิค Gain Ratio Attribute Evaluation ผลปรากฏว่า ก่อนและหลังการ คัดเลือกตัวแปรในเทคนิค OneR ให้ค่าความเที่ยงตรง ร้อยละ 80.90 เท่าเดิม เทคนิค RIPPER ให้ค่าความ เที่ยงตรง ร้อยละ 75.50 หลังจากคัดเลือกตัวแปร ให้ค่าความเที่ยงตรง ร้อยละ 77.30 ทำให้ค่าความเที่ยงตรง เพิ่มขึ้น ร้อยละ 1.8 เทคนิค Decision Rule ให้ค่าความเที่ยงตรง ร้อยละ 77 หลังจากคัดเลือกตัวแปร ให้ค่าความ เที่ยงตรง ร้อยละ 76.60 ทำให้ค่าความเที่ยงตรงลดลง ร้อยละ 0.4 เทคนิค Decision Table ให้ค่าความเที่ยงตรง ร้อยละ 76.4 หลังจากคัดเลือกตัวแปร ให้ค่าความเที่ยงตรง ร้อยละ 75.30 ทำให้ค่าความเที่ยงตรงลดลง ร้อยละ 1.1 ด้วย 10-fold cross validation

ผลการวัดประสิทธิภาพแบบจำลองด้วยค่าความเที่ยงตรง (Precision) ทั้ง 4 เทคนิคจากการ วิเคราะห์ปัจจัยด้วยเทคนิค Gain Ratio Attribute Evaluation และใช้การคัดกรองค่าผิดพลาด ผลปรากฏว่า ก่อนและหลังการคัดเลือกตัวแปร และใช้การคัดกรองค่าผิดพลาด ในเทคนิค RIPPER ให้ค่าความเที่ยงตรง มากที่สุด ร้อยละ 91 ประสิทธิภาพสูงขึ้น ร้อยละ 15.5 ต่อมา เทคนิค Decision Rule ให้ค่าความเที่ยงตรง ร้อยละ 87.10 ประสิทธิภาพสูงขึ้น ร้อยละ 10.1 เทคนิค Decision Table ให้ค่าความ เที่ยงตรง ร้อยละ 85.80 ประสิทธิภาพสูงขึ้น ร้อยละ 9.4 เทคนิค OneR ให้ค่าความเที่ยงตรง ร้อยละ 82.80 ประสิทธิภาพสูงขึ้น ร้อยละ 1.9 ด้วย 10-fold cross validation ตามลำดับ

ผลการวัดประสิทธิภาพแบบจำลองด้วยค่าความระลึก (Recall) ทั้ง 4 เทคนิคจากการ วิเคราะห์ปัจจัยด้วยเทคนิค Gain Ratio Attribute Evaluation ผลปรากฏว่า ก่อนและหลังการ คัดเลือกตัวแปรในเทคนิค OneR ให้ค่าความระลึก ร้อยละ 73.20 เท่าเดิม เทคนิค RIPPER ให้ค่าความระลึก ร้อยละ 75.80 หลังจากคัดเลือกตัวแปร ให้ค่าความระลึก ร้อยละ 77.70 ทำให้ค่าความระลึกเพิ่มขึ้น ร้อยละ 1.9 เทคนิค Decision Rule ให้ค่าความระลึก ร้อยละ 77.40 หลังจาก คัดเลือกตัวแปร ให้ค่าความระลึก ร้อยละ 76.60 ทำให้ค่า ความระลึกลดลง ร้อยละ 0.8 เทคนิค Decision Table ให้ค่าความระลึก ร้อยละ 76.40 หลังจากคัดเลือกตัวแปร ให้ค่าความระลึก ร้อยละ 75.80 ทำให้ค่าความระลึกลดลง ร้อยละ 0.6 ด้วย 10-fold cross validation

ผลการวัดประสิทธิภาพแบบจำลองด้วยค่าความระลึก (Recall) ทั้ง 4 เทคนิคจากการวิเคราะห์ปัจจัยด้วยเทคนิค Gain Ratio Attribute Evaluation และใช้การคัดกรองค่าผิดพลาด ผลปรากฏว่า ก่อนและหลังการคัดเลือกตัวแปร และใช้การคัดกรองค่าผิดพลาด ในเทคนิค RIPPER ให้ค่า

ความระลึกมากที่สุด ร้อยละ 91.10 ประสิทธิภาพสูงขึ้น ร้อยละ 15.3 ต่อมา เทคนิค Decision Rule ให้ค่าความระลึก ร้อยละ 87 ประสิทธิภาพสูงขึ้น ร้อยละ 9.6 เทคนิค Decision Table ให้ค่าความระลึก ร้อยละ 84.70 ประสิทธิภาพสูงขึ้น ร้อยละ 8.3 เทคนิค OneR ให้ค่าความระลึก ร้อยละ 77.10 ประสิทธิภาพสูงขึ้น ร้อยละ 3.9 ด้วย 10-fold cross validation ตามลำดับ ผลการวัดประสิทธิภาพแบบจำลองด้วยค่า F-measure ทั้ง 4 เทคนิคจากการวิเคราะห์ปัจจัย ด้วยเทคนิค Gain Ratio Attribute Evaluation ผลปรากฏว่า ก่อนและหลังการคัดเลือกตัวแปรใน เทคนิค OneR ให้ ค่า F-measure ร้อยละ 66.50 เท่าเดิม เทคนิค RIPPER ให้ค่า F-measure ร้อยละ 75.60 หลังจากคัดเลือกตัวแปร ให้ค่า F-measure ร้อยละ 77.40 ทำให้ค่า F-measure เพิ่มขึ้น ร้อยละ 1.8 เทคนิค Decision Rule ให้ค่า F-measure ร้อยละ 77.20 หลัง จากคัดเลือกตัวแปร ให้ค่า F-measure ร้อยละ 76.60 ทำให้ค่า F-measure ลดลง ร้อยละ 0.6 เทคนิค Decision Table ให้ค่า F-measure ร้อยละ 76.40 หลังจาก คัดเลือกตัวแปร ให้ค่า F-measure ร้อยละ 75.50 ทำให้ค่า F-measure ลดลง ร้อยละ 0.9 ด้วย 10-fold cross validation

ผลการวัดประสิทธิภาพแบบจำลองด้วยค่า F-measure ทั้ง 4 เทคนิคจากการวิเคราะห์ปัจจัย ด้วยเทคนิค Gain Ratio Attribute Evaluation และใช้การคัดกรองค่าผิดพลาด ผลปรากฏว่า ก่อน และหลังการคัดเลือกตัวแปรและใช้การคัดกรองค่าผิดพลาด ในเทคนิค RIPPER ให้ค่า F-measure มากที่สุด ร้อยละ 90.90 ประสิทธิภาพสูงขึ้น ร้อยละ 15.3 ต่อมา เทคนิค Decision Rule ให้ค่า F-measure ร้อยละ 87 ประสิทธิภาพสูงขึ้น ร้อยละ 9.8 เทคนิค Decision Table ให้ค่า F-measure ร้อยละ 85 ประสิทธิภาพสูงขึ้น ร้อยละ 8.6 เทคนิค OneR ให้ค่า F-measure 72 ประสิทธิภาพสูงขึ้น ร้อยละ 5.5 ด้วย 10-fold cross validation ตามลำดับ

## 6. สรุป

สรุปได้ว่า เทคนิค RIPPER มีความเหมาะสมในการสร้างแบบจำลองการพยากรณ์การสำเร็จการศึกษาระดับปริญญาตรี ในสถานศึกษา คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม โดยใช้เทคนิค ในการเปรียบเทียบได้แก่ Decision Table, Decision Rule, RIPPER และ OneR โดยใช้ Gain Ratio และ Filter ในการคัดเลือก แอตทริบิวต์ ผลการทดลองพบว่าเทคนิค RIPPER มีความเหมาะสมมากที่สุด ในการสร้างแบบจำลองการพยากรณ์การสำเร็จ การศึกษา โดยการวิเคราะห์ประสิทธิภาพด้วยค่าความถูกต้อง (Precision), ค่าความระลึก (Recall) และ F-measure

## 7. ข้อเสนอแนะ

7.1 ควรเพิ่มการวิเคราะห์ปัจจัยให้หลายหลายมากขึ้น

7.2 แบบจำลองต้นแบบนี้ สามารถนำไปใช้ในการพัฒนาระบบการพยากรณ์การสำเร็จการศึกษาระดับปริญญาตรี ในสถานศึกษา คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ มหาวิทยาลัยมหาสารคาม

## 8. เอกสารอ้างอิง

- สายชล สีนสมบูรณ์ทอง. (2558). *การทำเหมืองข้อมูล (Data Mining)*. กรุงเทพฯ: จามจุรีโปรดักส์ สมฤทัย กลัดแก้ว. (2557). *ระบบสนับสนุนการตัดสินใจการเลือกตำแหน่งงานให้สอดคล้องกับ ความสามารถของบัณฑิต*. ปริญญาวิทยาศาสตรมหาบัณฑิต สาขากระบวนสนับสนุน การตัดสินใจ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง.
- เสกสรรค์ วิสัยลักษณ์ วิชา เจริญภัณฑารักษ์ และดวงดาว วิชาดากุล. (2558). การใช้เทคนิคการทำ เหมืองข้อมูลเพื่อพยากรณ์ผลการเรียนของนักเรียน โรงเรียนสาธิตแห่ง มหาวิทยาลัยเกษตรศาสตร์. *ชุดิมา อุตมะมุณี และประสงค์ ปราณีตพลกรัง*. (2553). การพัฒนาตัวแบบระบบสนับสนุนการ ตัดสินใจ แบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา. *Journal of Information Science and Technology*, 1(2), 39 - 48.
- นิภาพร ชนะมาร และพรณี สิทธิเดช. (2557). การวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือก คุณสมบัติ และการพยากรณ์. *วารสารมหาวิทยาลัยราชภัฏสกลนคร*, 6(12), 34 - 35.
- Osiris, V. (2015). *Feature Selection and Classification Methods for Decision Making: A Comparative Analysis*. Ph.D. dissertation, Information Systems, College of Engineering and Computing, Nova Southeastern University,
- Jiawei, H., & Micheline, K. (2006). *Data Mining: Concepts and Techniques Second Edition*. San Francisco: Morgan Kaufmann.
- Group, T. G. (2010). What is data mining? [Online]. Retrieved from: <http://www.gartner.com> [accessed 4 March 2014].
- PangNing, T. Michael, S., & Vipin, K. (2006). *Introduction to Data Mining*. Boston: Pearson Addison Wesley.
- David, H., Heikki, M., and Padhraic S. (2001). *Principles of Data Mining*. [Online]. Retrieved from: [ftp://gamma.sbin.org.pub/doc/books/Principles\\_of\\_Data\\_Mining.pdf](ftp://gamma.sbin.org.pub/doc/books/Principles_of_Data_Mining.pdf) [accessed 4 March 2014].