

## แบบเสนอเค้าโครงโครงการคอมพิวเตอร์ ปีการศึกษา2/2564

### 1. ชื่อโครงการ

การวิเคราะห์ปัจจัยสำคัญของเศรษฐกิจครัวเรือนด้วยการคัดเลือกคุณสมบัติแบบ Gain Ratio Feature Selection

### 2. ผู้เสนอโครงการ

ชื่อ-สกุล: นางสาว วชิราภรณ์ เจริญมา รหัสประจำตัว: 62102105141

หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชา วิทยาการคอมพิวเตอร์

อาจารย์ที่ปรึกษาหลัก อาจารย์ ดร.นิภาพร ชนะมาร

### 3. หลักการและเหตุผล

การฟื้นตัวในระยะข้างหน้าจะมีความแตกต่างกันในแต่ละกลุ่มธุรกิจและครัวเรือนที่สำคัญ หากการระบาดของโควิด-19 กินระยะเวลายาวนานขึ้น ความแตกต่างของการฟื้นตัว จะยิ่งทวีความรุนแรงมากขึ้น ซึ่งในท้ายที่สุดจะไปซ้ำเติมความเปราะบางของโครงสร้างเศรษฐกิจไทยที่เป็นต้นทุนอยู่เดิม (Rachot Leingchan. 2564: ออนไลน์) ซึ่งมหาวิทยาลัยราชภัฏสกลนครมีนโยบายและพันธกิจในการพัฒนาท้องถิ่น อย่างยั่งยืนบนพื้นฐานเศรษฐกิจพอเพียงเพื่อพัฒนาท้องถิ่น โดยมีโครงการน้อมนำศาสตร์พระราชาสู่การพัฒนาท้องถิ่น ยกระดับรายได้ซึ่งได้ดำเนินการมาตั้งแต่ปี 2561 จนถึงปัจจุบัน (มรสน. 2560: ออนไลน์) ในการพัฒนาสภาพทางเศรษฐกิจของคนในชุมชนซึ่งมหาวิทยาลัยราชภัฏสกลนคร ได้รับผิดชอบ 19 ตำบลในจังหวัดสกลนคร โดยมีการเก็บข้อมูลสภาพทางเศรษฐกิจไว้เป็นฐานข้อมูล ส่วนกลางของมหาวิทยาลัยราชภัฏสกลนคร ซึ่งปัจจุบันยังไม่มีวิเคราะห์ปัจจัยที่ส่งผลต่อสภาพเศรษฐกิจครัวเรือนที่จะสามารถสนับสนุนการตัดสินใจหรือการวางแผนการพัฒนาชุมชน

เทคโนโลยีในปัจจุบันมีหลากหลายเทคโนโลยีและมีหลากหลายศาสตร์ที่จะสามารถนำมาวิเคราะห์ข้อมูลขนาดใหญ่ได้โดยเฉพาะอย่างยิ่งหลักการทำให้เหมือนข้อมูลซึ่งเป็นหลักการในการวิเคราะห์เป็นหลักการของสถิติขั้นสูงที่จะสามารถวิเคราะห์ข้อมูลขนาดใหญ่ได้ โดยมีนักวิจัยหลากหลายศาสตร์ที่ได้นำเทคนิคการทำเหมือนข้อมูลมาประยุกต์ใช้งาน เช่น งานวิจัยของ นิภาพร ชนะมาร และพรณี สิทธิเดช (2557) ได้ศึกษาการประยุกต์ใช้เทคนิคการทำเหมือนข้อมูลในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิตที่ศึกษาและสำเร็จการศึกษาแล้วในหลักสูตรเดียวกัน จำนวน 180 ระเบียบ โดยใช้เทคนิคการคัดเลือกคุณสมบัติที่สำคัญ แล้วสร้างตัวแบบการพยากรณ์ด้วยเทคนิค BPNN และเทคนิค SVMs เมื่อทดลองใช้เทคนิคการรวมกลุ่ม ด้วยวิธี Bagging ร่วมกับ BPNN และ SVMs พบว่าผลการพยากรณ์ของ Baggingร่วมกับ BPNN (Bagging BPNN) มีค่าความผิดพลาดอยู่ใน ระดับต่ำสุด (RMSE=0.1051) งานของ วราวุธ พลาศรี (2556) ได้ศึกษาการศึกษาปัจจัยที่มีผลต่อความยากจนของครัวเรือนในชนบท : กรณีศึกษาจังหวัดมหาสารคาม การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาสภาพเศรษฐกิจของครัวเรือนในชนบท สถานการณ์ความยากจน ลักษณะของครัวเรือน

ที่ยากจน และปัจจัยที่มีผลต่อความยากจนของครัวเรือนในชนบทจังหวัดมหาสารคาม โดยกลุ่มประชากรที่ใช้ในการศึกษาคือ ครัวเรือนที่อยู่ในเขตพื้นที่ชนบทจังหวัดมหาสารคาม จำนวน 180,328 ครัวเรือน ขนาดกลุ่มตัวอย่างเท่ากับ 400 ครัวเรือน และงานวิจัยของ อัจจิมา มณฑาพันธุ์(2562) งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเกี่ยวกับการเปรียบเทียบวิธีการคัดเลือกคุณลักษณะที่สำคัญเพื่อนำมาใช้ในการ ปรับปรุงการพยากรณ์การเป็นมะเร็งเต้านม โดยใช้วิธีการคัดเลือกคุณลักษณะจากเทคนิคต่าง ๆ จำนวน 7 เทคนิค ได้แก่เทคนิค Correlation Based Feature Selection เทคนิค Information Gain เทคนิค Gain Ratio เทคนิค Chi-Square เทคนิค Forward Selection เทคนิค Backward Elimination และเทคนิค Evolutionary Selection หลังจากคัดเลือกคุณลักษณะ ที่สำคัญจึงนำผลที่ได้จากแต่ละเทคนิคมาคำนวณหาประสิทธิภาพในการพยากรณ์การเป็นมะเร็งเต้านมโดยใช้เทคนิคซัพพอร์ต เวกเตอร์แมชชีน ผลการทดลองพบว่าร้อยละของความถูกต้องในการพยากรณ์การเป็นมะเร็งเต้านม จากจำนวนคุณลักษณะของข้อมูลทั้งหมด 30 คุณลักษณะเท่ากับ 91.39

จากที่กล่าวมาข้างต้นผู้วิจัยจึงมีความสนใจที่จะศึกษาวิเคราะห์ปัจจัยสำคัญทางเศรษฐกิจครัวเรือน ด้วยวิธีการ Gain Ratio Feature Selection โดยใช้โปรแกรม RapidMiner เพื่อให้ได้ปัจจัยที่สำคัญสำหรับการวางแผนและการพัฒนาด้านเศรษฐกิจของชุมชนท้องถิ่นเพื่อให้สามารถมีคุณภาพชีวิตที่ดีขึ้น

#### 4.วัตถุประสงค์ของโครงการ

เพื่อวิเคราะห์ปัจจัยสำคัญของเศรษฐกิจครัวเรือนด้วยการคัดเลือกคุณสมบัติแบบ Gain Ratio Feature Selection

#### 5. ขอบเขตของโครงการ

##### 5.1 ด้านข้อมูล

ข้อมูลที่ใช้ในการศึกษาค้างนี้ คือข้อมูลประชากรจากภาคครัวเรือนเฉพาะครัวเรือนในเขตพื้นที่ชนบทของจังหวัดสกลนคร ซึ่งมี 20 หมู่บ้าน 12 ตำบล 12 อำเภอ โดยช่วงเวลาที่ทำการเก็บรวบรวมข้อมูลคือ ปี พ.ศ. 2563 – 2564 และจากฐานข้อมูลสภาพทางเศรษฐกิจครัวเรือนเป้าหมายตามโครงการจ้างงาน ประชาชนที่ได้รับผลกระทบจากสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) (สำนักวิทยบริการและเทคโนโลยีสารสนเทศ, 2563: ออนไลน์)

5.1.1 ประชากรที่ใช้ในการศึกษา ได้แก่ ครัวเรือนตำบลที่อยู่ในช่วงปี พ.ศ. 2561 - 2563 ได้มาจากข้อมูล 12 ตำบล ซึ่งมีจำนวน 17,933 ครัวเรือน

5.1.2 กลุ่มตัวอย่างที่ใช้ในการศึกษาและวิเคราะห์ข้อมูล ได้แก่ กลุ่มครัวเรือนบ้านในช่วงปี พ.ศ. 2561 - 2563 ได้มาจากการเลือกแบบเจาะจง (Purposive Sampling) จำนวน 3,233 ครัวเรือน

ข้อมูลจากฐานข้อมูล สภาพทางเศรษฐกิจครัวเรือนเป้าหมายตามโครงการจ้างงาน ประชาชนที่ได้รับผลกระทบจากสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) โดยในฐานข้อมูลมีการเก็บข้อมูลจากฐานเศรษฐกิจชุมชนซึ่งมีการเก็บข้อมูลออกเป็น 10 ส่วน ดังนี้

ส่วนที่ 1 ข้อมูลทั่วไปครัวเรือน

ส่วนที่ 2 ทรัพย์สินของครัวเรือน

ส่วนที่ 3 อาชีพและรายได้ของครัวเรือน

ส่วนที่ 4 รายจ่ายของครัวเรือน

ส่วนที่ 5 หนี้สินของครัวเรือน

ส่วนที่ 6 ผลกระทบจากสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019

(COVID - 19)

ส่วนที่ 7 การใช้เทคโนโลยีสารสนเทศ

ส่วนที่ 8 การเข้าร่วมการละเล่น การฟ้อน การรำ พิธีกรรมตามวิถีวัฒนธรรมชุมชน

ส่วนที่ 9 การเข้าร่วมโครงการที่ผ่านมาย้อนหลัง 3 ปี

ส่วนที่ 10 ข้อคิดเห็นและข้อเสนอแนะเพิ่มเติม

และได้ศึกษางานวิจัยที่เกี่ยวข้องนำมาประยุกต์ใช้ในฐานะข้อมูล Database โดยใช้กระบวนการ CRIPS-DM (Cross Reference Industry Standard for Data Mining) ในการดำเนินงาน

## 5.2 ด้านเทคนิค

-ใช้เทคนิค Gain Ratio Feature Selection เป็นเทคนิคที่ใช้ในการหาปัจจัยสำคัญ

-เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เพื่อมาหาประสิทธิภาพเปรียบเทียบระหว่างปัจจัยเริ่มต้นและการหาปัจจัยสำคัญจากเทคนิค Gain Ratio Feature Selection มาเปรียบเทียบประสิทธิภาพจากนั้นจะได้มาซึ่งปัจจัยที่เหมาะสมที่สุด

-ใช้กระบวนการ CRIPS-DM (Cross Reference Industry Standard for Data Mining) เป็นกระบวนการที่ใช้ในการดำเนินงานวิจัย

## 5.3 ด้านเครื่องมือ

### 5.3.1 ซอฟต์แวร์

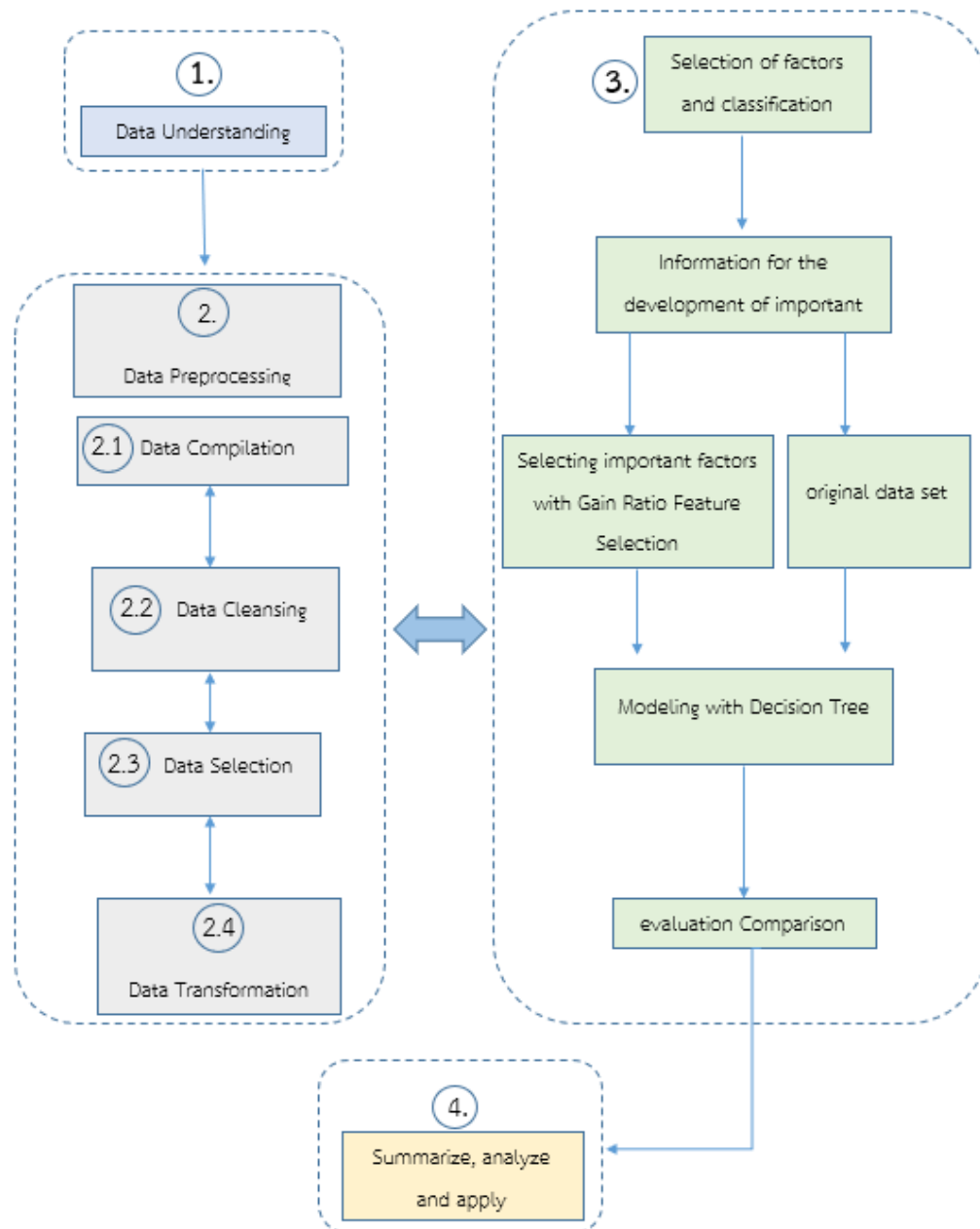
การศึกษาค้นคว้านี้ได้ทำการทดลองดำเนินการผ่านโปรแกรม RapidMiner Studio เวอร์ชัน 9.10 เป็นโปรแกรมที่ออกแบบมาสำหรับการวิเคราะห์ข้อมูล ของบริษัท RapidMiner (mypccrack 2565 : ออนไลน์)

### 5.3.2 ฮาร์ดแวร์

- เครื่องคอมพิวเตอร์ Notebook ที่ใช้ทำโครงการ
- หน่วยประมวลผล ADM Ryzen 5 2500U with Radeon Vega Mobile Gfx 2.00 GHz
- หน่วยความจำหลัก (RAM): 8.00 GB
- ระบบปฏิบัติการ (OS): Windows 10 64-bit

## 6. กรอบการดำเนินงาน

การดำเนินงานนี้ได้ใช้การประยุกต์ทำตามแนวทางในการทำเหมืองข้อมูลที่เรียกว่า กระบวนการมาตรฐานอุตสาหกรรม หรือ CRIPS-DM (Cross Reference Industry Standard for Data Mining) ที่ได้รับความนิยมมากในปัจจุบันซึ่งมีขั้นตอนการดำเนินงาน (chapman et al. 2000) ดังนี้



ภาพที่ 1 กรอบการดำเนินงานวิจัย

ที่มา : ผู้วิจัย (2565)

## 7.แผนการดำเนินงาน

1. กำหนดหัวข้อและนำเสนอหัวข้อ
2. ค้นหาปัญหา โอกาสและเป้าหมาย
3. ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง
4. เสนอเค้าโครงโครงการ
5. ศึกษาและวิเคราะห์ข้อมูล
6. ทำความเข้าใจข้อมูลและเตรียมข้อมูล
7. ดำเนินการพัฒนาโมเดล
8. ประเมินประสิทธิภาพการพัฒนาโมเดล
9. จัดทำเอกสารประกอบโครงการ
10. นำเสนอโครงการจบ

ตารางที่ 1 ระยะเวลาการดำเนินงาน

กิจกรรม	ระยะเวลาในการดำเนินงาน (พ.ศ.2565)				
	ม.ค.	ก.พ.	มี.ค.	เม.ย	พ.ค.
1.กำหนดหัวข้อและนำเสนอหัวข้อ	→				
2.ค้นหาปัญหา โอกาสและเป้าหมาย	→				
3.ศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง	→				
4.เสนอเค้าโครงโครงการ	→				
5.ศึกษาและวิเคราะห์ข้อมูล		→			
6.ทำความเข้าใจข้อมูลและเตรียมข้อมูล		→			
7.ดำเนินการพัฒนาโมเดล			→		
8.ประเมินประสิทธิภาพการพัฒนาโมเดล			→		
9.จัดทำเอกสารประกอบโครงการ				→	
10.นำเสนอโครงการจบ					→

## 8.ประโยชน์ที่คาดว่าจะได้รับ

ได้ปัจจัยที่สำคัญที่ส่งผลต่อสภาพทางเศรษฐกิจครัวเรือนเพื่อสนับสนุนหรือเป็นข้อมูลประกอบการตัดสินใจในการพัฒนาชุมชนท้องถิ่นสำหรับนักวิจัย

## 9.เอกสารและงานวิจัยที่เกี่ยวข้อง

### 9.1 ทฤษฎีที่เกี่ยวข้อง

9.1.1 การทำเหมืองข้อมูล (Data Mining)

9.1.2 การจำแนกข้อมูลด้วยการสร้างต้นไม้ตัดสินใจ (Decision Tree)

9.1.3 Feature selection การคัดเลือกคุณสมบัติ

9.1 ทฤษฎีที่เกี่ยวข้อง

9.1.4 ความแม่นยำ (Accuracy)

9.1.6 CRISP-DM (Cross-Industry Standard Process For Data Mining)

#### 9.1.1 การทำเหมืองข้อมูล (Data Mining)

เป็นเทคนิคในการวิเคราะห์ข้อมูลอย่างหนึ่ง ซึ่งมาจากคำว่า เหมืองข้อมูล นั่นคือ เป็นการค้นหาสิ่งที่มีประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่ เช่น ข้อมูลการซื้อขายสินค้าในซูเปอร์มาร์เก็ตต่าง ๆ โดยข้อมูลเหล่านี้จะเก็บจากรายการสินค้าที่ลูกค้าซื้อในแต่ละครั้ง โดยเมื่อทำการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining แล้วจะได้สิ่งที่เป็นประโยชน์ Data Mining เป็นเทคนิคในการวิเคราะห์ข้อมูลอย่างหนึ่ง ซึ่งมาจากคำว่า เหมืองข้อมูล นั่นคือ เป็นการค้นหาสิ่งที่มีประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่ เช่น ข้อมูลการซื้อขายสินค้าในซูเปอร์มาร์เก็ตต่างๆ โดยข้อมูลเหล่านี้จะเก็บจากรายการสินค้าที่ลูกค้าซื้อในแต่ละครั้ง โดยเมื่อทำการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining แล้วจะได้สิ่งที่เป็นประโยชน์เช่น ลูกค้าส่วนใหญ่ที่ซื้อเปียร์มักจะซื้อผ้าอ้อมด้วย จะเห็นว่าข้อมูลนี้เป็นข้อมูลที่ไม่เคยคิดว่าจะมีความสัมพันธ์กัน และเมื่อได้ความรู้แบบนี้ก็อาจจะนำไปเป็นออกโปรโมชั่น หรือช่วยในการจัดวางชั้นสินค้า หรือเป็นแนวทางในการสั่งซื้อสินค้าในซูเปอร์มาร์เก็ตต่อไปได้ นอกจากนี้ Data Mining ยังมีเทคนิคในการประยุกต์ใช้งานได้อย่างดี (หนึ่งพิชัย ชัยอากร.2559: ออนไลน์)

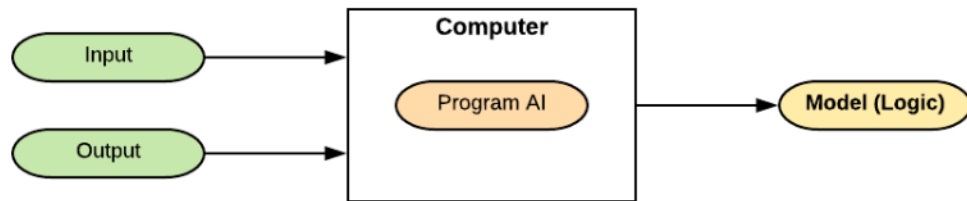
#### 9.1.2 การทำเหมืองข้อมูล จำแนกออกเป็น 2 ประเภท คือ

1) **Unsupervised Learning** การสร้างโมเดลโดยใช้ข้อมูล input เพียงอย่างเดียวไม่มี target การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง โดยการสร้างโมเดลที่เหมาะสมกับข้อมูล การเรียนรู้แบบนี้แตกต่างจากการเรียนรู้แบบมีผู้สอน คือ จะไม่มีการระบุผลที่ต้องการหรือประเภทไว้ก่อน การเรียนรู้แบบนี้จะพิจารณาวัตถุเป็นเซตของตัวแปรสุ่ม แล้วจึงสร้างโมเดลความหนาแน่นร่วมของชุดข้อมูลการเรียนรู้แบบไม่มีผู้สอนสามารถนำไปใช้ร่วมกับการอนุมานแบบเบย์ เพื่อหาความน่าจะเป็นแบบมีเงื่อนไขของตัวแปรสุ่มโดยกำหนดตัวแปรที่เกี่ยวข้องให้ นอกจากนี้ยังสามารถนำไปใช้ในการบีบอัดข้อมูล ซึ่งโดยพื้นฐานแล้ว ขั้นตอนวิธีการบีบอัดข้อมูลจะขึ้นอยู่กับ การแจกแจงความน่าจะเป็นของข้อมูลไม่อย่างชัดแจ้งก็โดยปริยาย (สุพรรณ พ้าหยง. 2562)

#### 2) Supervised Learning

เป็นการเรียนรู้ข้อมูลต่าง ๆ โดยมีผู้สอน อาศัยข้อมูลในการฝึกฝน เพื่อช่วยให้ตัวเทคโนโลยีสามารถเรียนรู้ผล และคาดคะเนผลลัพธ์ต่าง ๆ ได้อย่างแม่นยำมากยิ่งขึ้น โดยการเรียนรู้ในรูปแบบนี้

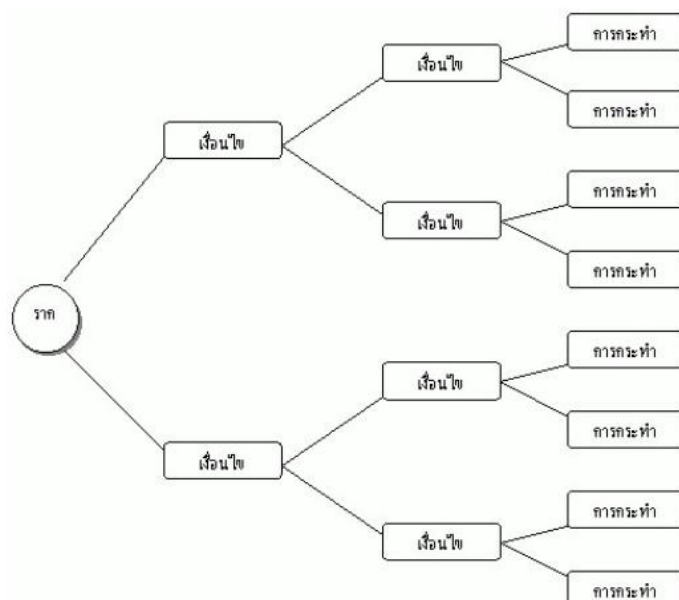
มักถูกนำมาใช้งานในเชิงธุรกิจทั้งการคำนวณราคาบ้าน การคาดคะเนค่าเงิน หรือแม้แต่การวิเคราะห์ผลการแข่งขันต่าง ๆ เป็นต้น กระบวนการสร้าง model เรียกว่าการ “เทรน” ซึ่งสามารถกินเวลาได้ตั้งแต่หลักวินาทีจนถึงหลาย ๆ วัน แล้วแต่ความซับซ้อนของโจทย์ที่เราต้องการแก้ และพลังในการประมวลผลของเครื่องคอมพิวเตอร์ที่เราใช้เทรน



ภาพที่ 2 ภาพแสดงกระบวนการเทรน เพื่อให้ได้ model ที่เราต้องการ  
ที่มา : Phuri Chalermkitsakul (2563: ออนไลน์)

### 9.1.3 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้การตัดสินใจ (decision tree) เป็นเครื่องมือที่ช่วยให้วิเคราะห์เหตุการณ์ หรือสถานการณ์เพื่อการตัดสินใจได้อย่างเป็นระบบและรวดเร็ว ต้นไม้การตัดสินใจมีลักษณะเป็นกราฟรูปต้นไม้ ซึ่งแสดงที่ตั้งต้นที่มีรากและแขนงต่างๆแตกออกมาจากต้นไม้ไปในทิศทางเดียว จนกระทั่งนำไปสู่ข้อสรุปสำหรับการตัดสินใจได้ ต้นไม้การตัดสินใจมีประโยชน์ในการสรุปการตัดสินใจที่มีความซับซ้อนในง่ายต่อความเข้าใจ ปัจจุบันต้นไม้การตัดสินใจเป็นที่นิยมใช้ในงานหลายอย่าง เช่น การแพทย์ ธุรกิจ การเขียนโปรแกรม การสร้างเครื่องที่เรียนรู้ได้เอง การสร้างระบบผู้เชี่ยวชาญ ฯลฯ (ครรชิต มาลัยวงศ์.2553: ออนไลน์)



ภาพที่ 3 เทคนิคที่แนะนำ decision tree หรือต้นไม้ตัดสินใจ  
ที่มา: Nuthdanai wangpratham. (2564: ออนไลน์)

**9.1.4 การคัดเลือกคุณสมบัติ (Feature Selection)** การคัดเลือกคุณสมบัติเป็นเทคนิคที่ช่วยลดจำนวนตัวแปรที่จะใช้ในแบบพยากรณ์ อาจกระทำเพื่อเลือกตัวแปรที่ดีที่สุดเพียงตัวเดียว หรือเลือกกลุ่มของตัวแปรที่มีความสำคัญต่อการพยากรณ์ กระบวนการคัดเลือกคุณสมบัติเป็นกระบวนการที่สำคัญในการเตรียมข้อมูลของการทำเหมืองข้อมูล เพื่อให้การสร้างแบบพยากรณ์มีประสิทธิภาพ เพราะจะช่วยลดมิติของข้อมูล และอาจช่วยให้การเรียนรู้วิธีการ พยากรณ์ดำเนินการได้เร็วขึ้นและมีประสิทธิภาพมากขึ้น ในงานวิจัยนี้ ทดลองใช้การคัดเลือกคุณสมบัติ 3 วิธี ดังต่อไปนี้

1) การคัดเลือกคุณสมบัติแบบ Correlation-based Feature Selection เป็นการคัดเลือกกลุ่ม คุณสมบัติอย่างง่าย ใช้หลักการคำนวณค่าความสัมพันธ์ระหว่างคุณสมบัติย่อยต่อค่าพยากรณ์ ซึ่งอาจใช้ คำนวณด้วยค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson's correlation) และมีการจัดอันดับตามค่าความสัมพันธ์ เพื่อประเมินค่าความสามารถในการพยากรณ์ของแต่ละคุณสมบัติ นอกจากนั้นยังพิจารณาคัดเลือกกลุ่มของ คุณสมบัติที่มีความสัมพันธ์ภายในระหว่างคุณสมบัติย่อยกันเองต่ำเพื่อลดความซ้ำซ้อนของอิทธิพลการพยากรณ์ (Hall and Smith. 1998)

2) การคัดเลือกคุณสมบัติแบบ Consistency-based Feature Selection เป็นการคัดเลือก คุณสมบัติที่ต้องกำหนดตัวชี้วัดความสอดคล้องมั่นคงไว้ก่อนเป็นอันดับแรก จากนั้นใช้ตัวชี้วัดนี้ เพื่อประเมิน ความมั่นคงของกลุ่มคุณสมบัติที่เกี่ยวข้องกับค่าพยากรณ์เดียวกัน (Liu and Setiono. 1996; Liu et al. 1998) ตัวชี้วัดความมั่นคงที่ถูกกำหนดให้เป็นตัวชี้วัดนี้ อาจกำหนดมาตรวัดเช่นเดียวกับการวัดระยะทางของ คุณสมบัติย่อยและสามารถแปลผลคุณสมบัติที่สอดคล้องกันมากด้วยค่าเข้าใกล้ศูนย์ กระบวนการคัดเลือกคุณสมบัติจะกระทำซ้ำและเลือกกลุ่มคุณสมบัติที่มีค่าตัวชี้วัดน้อยและจำนวนคุณสมบัติเท่าเดิมหรือลดลง เท่านั้น วิธีการค้นหาคุณสมบัติที่สอดคล้องมั่นคงนี้เป็นวิธีที่รวดเร็วและสามารถทราบความสัมพันธ์ ระหว่างคุณสมบัติได้ (Hall and Holmes. 2003)

3) การคัดเลือกคุณสมบัติแบบ Gain Ratio Feature Selection เป็นวิธีคัดเลือกตัวแปร โดยมี หลักการเช่นเดียวกับการเลือกตัวแปรของการสร้างต้นไม้ตัดสินใจ เพื่อให้ได้ตัวแปรที่เป็นตัวแบ่งข้อมูลออกเป็นกลุ่มย่อยที่มีสมาชิกภายในกลุ่มเป็นชนิดเดียวกันมากที่สุด (Homogeneous) ด้วยมาตรวัดการได้ประโยชน์จากการแบ่งกลุ่มย่อยเรียกว่า อัตราส่วนเกน (Gain Ratio) ซึ่งเป็นอัตราส่วนของค่าเกน (Gain หรือ Information Gain) กับ ค่าสารสนเทศการแบ่งกลุ่ม (Split Info) อันเป็นการลดอิทธิพลของตัวแปรที่มีค่าหลายค่า ผลที่ได้รับจากการใช้เทคนิคนี้จะได้ลำดับของตัวแปรซึ่งตัวแปรที่อยู่ลำดับแรกๆ จะถือว่ามีอิทธิพลในการพยากรณ์ตัวแปรเป้าหมายมากกว่าตัวแปรในลำดับถัดไป ทำให้เราสามารถพิจารณาเลือกจำนวนตัวแปรที่เหมาะสมได้อย่างมีประสิทธิภาพ (Tan, Steinbach and Kumar. 2006; Asha, Manjunath and Jayaram. 2010) เกนเรโซ (GR) เป็นการประเมินความน่าเชื่อถือของมิติข้อมูลโดยการวัด Gain Ratio ในแต่ละคลาสการคำนวณ GR โดยใช้ค่า SplitINFO ในสมการที่ 1 และการคำนวณค่าการวัด Gain Ratio ดังสมการที่ 2 (วีระยุทธ พิมพาพร และพยุง มีสัง 2557).



$$SplitINFO = \sum_{i=1}^k \frac{n_i}{n} \log_2 \frac{n_i}{n} \quad (1)$$

$$GainRatio = \frac{\Delta INFO}{SplitINFO} \quad (2)$$

### 9.1.5 ตัววัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล

ดังที่กล่าวไปแล้วว่าการนำโมเดลไปใช้งานจริงได้นั้นเราจำเป็นจะต้องทราบประสิทธิภาพของโมเดลเสียก่อน โดยทั่วไปแล้วจะมีตัววัดที่นิยมใช้กันในงานวิจัยและการทำงานต่างๆ อยู่ 5 ค่า (ธาดา จันตะคุณ, 2559) คือ

- 1) Precision เป็นการวัดค่าความแม่นยำของโมเดล โดยพิจารณาแยกทีละคลาส
  - 2) Recall เป็นการวัดค่าความถูกต้องของโมเดล โดยพิจารณาแยกทีละคลาส
  - 3) F-measure เป็นการวัดค่า Precision และ Recall พร้อมกันของโมเดล โดยพิจารณาแยกทีละคลาส
  - 4) ความแม่นยำ (Accuracy) เป็นการวัดความถูกต้องของโมเดล โดยพิจารณา
- รวมทุกคลาส

ตารางที่ 1 ตาราง Confusion Matrix ของข้อมูล Weather ซึ่งมี 2 คลาส

Predicted/Actual	Yes	No
Yes	TP	FP
No	FN	TN

จากในตารางที่ 1 ค่าที่แสดงในช่องต่าง ๆ ของตารางประกอบด้วย

- 1) True Positive (TP) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งกำลังสนใจอยู่
- 2) True Negative (TN) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งไม่ได้สนใจอยู่
- 3) False Positive (FP) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งกำลังสนใจอยู่
- 4) False Negative (FN) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งไม่ได้สนใจอยู่

หลังจากที่เราสร้างตาราง Confusion Matix ได้ดังตารางที่ 1 วิธีคำนวณค่า Precision, Recall, F-measure และ Accuracy

1) **Precision** เป็นการวัดความแม่นยำของโมเดล โดยพิจารณาแยกทีละคลาส

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3)$$

2) **Recall** เป็นการวัดความถูกต้องของโมเดล โดยพิจารณาแยกทีละคลาส

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (4)$$

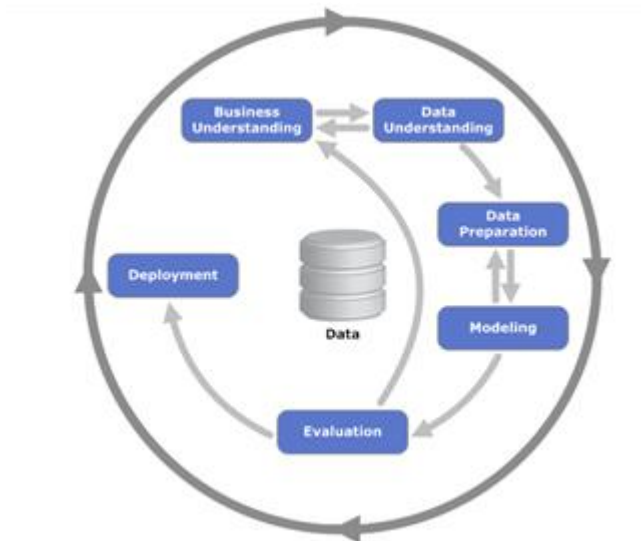
3) **F-measure** เป็นการวัดค่า Precision และ Recall พร้อมกันของโมเดล โดยพิจารณาแยกทีละคลาส

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

4) **Accuracy** เป็นการวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส คือ จำนวน True Positive ของทุกคลาสรวมกันได้เท่ากับ  $6/10 = 60\%$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (6)$$

## 9.1.6 CRISP-DM (Cross-Industry Standard Process For Data Mining)



ภาพที่ 4 แสดงกระบวนการ Cross-industry standard process for data mining

ที่มา: Thapanee Boonchob (2564: ออนไลน์)

กระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูล เพื่อทำการวิเคราะห์และนำไปใช้ประโยชน์ มีอยู่ 6 ขั้นตอน คือ

1) การทำความเข้าใจโจทย์ (Business Understanding) ขั้นตอนแรกมุ่งไปที่การทำความเข้าใจข้อมูลปัญหาและวัตถุประสงค์ของโครงการจากมุมมองข้อมูล จากนั้นแปลงปัญหาให้อยู่ในรูปของโจทย์สำหรับการวิเคราะห์ข้อมูล และวางแผนการดำเนินงานเบื้องต้น

2) การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนนี้เริ่มต้นด้วยการรวบรวมข้อมูล จากนั้นทำความเข้าใจ ตรวจสอบคุณภาพ และเลือกข้อมูลที่เกี่ยวข้องว่าจะใช้ข้อมูลใดบ้างในการวิเคราะห์ขั้นตอนที่ 1 และ 2 สามารถทำกลับไปมาได้ เนื่องจากการทำความเข้าใจข้อมูลทำให้เราเข้าใจข้อมูลมากขึ้น และการเข้าใจข้อมูลก็ทำให้เราเข้าใจข้อมูลมากขึ้นเช่นกัน

3) การเตรียมข้อมูล (Data Preparation) ขั้นตอนการเตรียมข้อมูล หมายถึง ขั้นตอนทั้งหมดที่จะทำเพื่อให้ข้อมูลดิบที่เรารวบรวมมา กลายเป็นข้อมูลสมบูรณ์ที่พร้อมจะเข้าสู่โมเดลในขั้นตอนที่ 4 เช่น การสร้างตาราง การลบข้อมูลที่ไม่ต้องการออก การแปลงข้อมูลให้อยู่ในรูปแบบที่ต้องการ เป็นต้น

4) การสร้างโมเดล (Modeling) ในขั้นตอนนี้ เราจะเลือกและทดสอบสร้างโมเดลหลายๆ แบบที่น่าจะสามารถแก้ไขปัญหที่ต้องการได้ จากนั้นค่อยๆ ปรับค่าพารามิเตอร์ในแต่ละโมเดล เพื่อให้ได้โมเดลที่เหมาะสมที่สุดมาใช้ในการแก้ไขปัญห

5) การวัดประสิทธิภาพของโมเดล (Evaluation) เราจะทำการวัดประสิทธิภาพของโมเดลที่ได้จากขั้นตอนที่ 4 เพื่อวัดว่าโมเดลมีประสิทธิภาพเพียงพอต่อการนำไปใช้งานแล้วหรือไม่ ซึ่งโมเดลแต่ละประเภทก็จะมีตัววัดประสิทธิภาพที่แตกต่างกันออกไป

6) การนำโมเดลไปใช้งานจริง (Deployment) เป็นการนำโมเดลที่เหมาะสมที่สุดไปใช้งานจริง เพื่อวิเคราะห์และแก้ปัญหที่ต้องการ (Thapanee Boonchob.2563)

## 9.3 โปรแกรมที่ใช้

### 9.3.1 โปรแกรม RapidMiner Studio Developer v9.10.0



ภาพที่ 5 โลโก้โปรแกรม RapidMiner

ที่มา: mypccrack (2565 : ออนไลน์)

RapidMiner คือซอฟต์แวร์ Data Science ใช้สำหรับการเตรียมข้อมูล การเรียนรู้เครื่อง การเรียนรู้ลึกการทำเหมืองข้อความ และการวิเคราะห์การทำนาย (Predictive analysis) เป็นซอฟต์แวร์ที่ช่วยในการจัดส่งข้อมูล และลดข้อผิดพลาดจนแทบจะไม่จำเป็นต้องเขียนโค้ดเพิ่ม แต่ที่ให้เป็นเครื่องมือที่ Data Scientist นิยมเลือกใช้เป็นเพราะว่า RapidMiner มีขั้นตอนพร้อมสำหรับการทำ Data mining (ขุดข้อมูล) และ Machine Learning ซึ่งรวมถึงการโหลดและการแปลงข้อมูล (ETL) การประมวลผลล่วงหน้าและการวาดภาพจากข้อมูล การวิเคราะห์เชิงพยากรณ์และการสร้างแบบจำลองทางสถิติ การประเมินผลและการปรับใช้ ต่างๆ ล้วนเป็นสิ่งที่ Data Scientist จำเป็นต้องทำในการเข้าใจข้อมูลมากขึ้น ที่มา : (Achieve. Plus. 2563: ออนไลน์)

## 9.4 งานวิจัยที่เกี่ยวข้อง

นิภาพร ชนะมาร และพรณี สิทธิเดช (2557) ได้ศึกษาการวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์ มีวัตถุประสงค์เพื่อการประยุกต์ใช้เทคนิคการทำเหมืองข้อมูลในการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิต โดยใช้เทคนิคการคัดเลือกคุณสมบัติที่สำคัญ แล้วสร้างตัวแบบการพยากรณ์ด้วยเทคนิค BPNN และเทคนิค SVMs จากข้อมูลที่คัดเลือกซึ่งเป็นปัจจัยการเรียนรู้ที่สำคัญ ข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลของนิสิตที่ศึกษาหลักสูตรปริญญาตรี สาขาวิชาวิทยาการคอมพิวเตอร์ ฉบับปรับปรุง พ.ศ. 2548 จำนวน 180 ระเบียน ประกอบด้วย คุณสมบัติ 23 ตัวแปร แบ่งเป็น ตัวแปรอิสระ 22 ตัวแปร ได้แก่ ข้อมูลภูมิหลังต่างๆ และข้อมูลผลการเรียนของรายวิชาที่ศึกษาในแผนการ เรียนชั้นปีที่หนึ่งและชั้นปีที่สอง ตัวแปรตามหรือตัวแปรพยากรณ์ คือ เกรดเฉลี่ยของนิสิตเมื่อสำเร็จ การศึกษา โดยข้อมูลที่ไม่เกี่ยวข้องกับการทำนาย เช่น สถานะการศึกษา ที่อยู่ หมายเลขโทรศัพท์และข้อมูลที่ ไม่มีค่า (Missing Value) ผู้วิจัยได้ทดลองสร้างตัวแบบการพยากรณ์จากข้อมูลทั้งหมดที่มีตัวแปรอิสระ 22 ตัวแปร ด้วย เทคนิคBPNN และเทคนิคSVMs ได้ผลการพยากรณ์ที่มีค่ารากที่สองของกำลังสองของข้อผิดพลาด (Root Mean Square Error: RMSE) เท่ากับ 0.2444 และ 0.1246 ตามลำดับ หลังจากนั้น จึงทำการวิเคราะห์ปัจจัยการเรียนรู้ ด้วยการคัดเลือกคุณสมบัติที่สำคัญ โดยใช้เทคนิคการคัดเลือกคุณสมบัติ 3 วิธี ได้แก่ การคัดเลือกคุณสมบัติ แบบ Correlation-based Feature Selection การ

คัดเลือกคุณสมบัติแบบ Consistency-based Feature Selection และ การคัดเลือกคุณสมบัติแบบ Gain Ratio Feature Selection ผลการทดลองทั้งสามเทคนิคสามารถลดจำนวน ของคุณสมบัติจาก 22 ตัวแปร เหลือ 9 ตัวแปร 10 ตัวแปร และ 11 ตัวแปร ตามลำดับ ผลของงานวิจัยนี้ให้ประโยชน์ในการ วิเคราะห์ปัจจัยการเรียนรู้และการพยากรณ์ผลสัมฤทธิ์ทางการเรียนของนิสิตซึ่งจะช่วยให้นิสิตสามารถ พยากรณ์ผลการเรียนของตนเองและปรับปรุงพฤติกรรมการเรียน ได้เช่น การเพิ่มถอนรายวิชาให้เหมาะสมกับ ศักยภาพตนเอง

วรายุทธ พลาศรี (2556) ได้ศึกษาปัจจัยที่มีผลต่อความยากจนของครัวเรือนในชนบท : กรณีศึกษา จังหวัดมหาสารคาม. การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาสภาพเศรษฐกิจของครัวเรือนในชนบท สถานการณ์ความยากจน ลักษณะของครัวเรือนที่ยากจน และปัจจัยที่มีผลต่อความยากจนของครัวเรือนในชนบทจังหวัดมหาสารคาม โดยกลุ่มประชากรที่ใช้ในการศึกษาคือ ครัวเรือนที่อยู่ในเขตพื้นที่ชนบทจังหวัดมหาสารคาม จำนวน 180,328 ครัวเรือน ขนาดกลุ่มตัวอย่างเท่ากับ 400 ครัวเรือน โดยได้เลือกวิธีการสุ่ม ตัวอย่างแบบหลายขั้นตอน (Multi-stage sampling method) และได้ทำการเก็บรวบรวมข้อมูลโดยใช้แบบสอบถามเป็นเครื่องมือ หลังจากนั้นจึงนำข้อมูลที่ได้มาวิเคราะห์โดยใช้สถิติพรรณนาและสถิติอนุมาน อนึ่งการศึกษาครั้งนี้ได้ใช้เส้นความยากจนของครัวเรือน ภาคตะวันออกเฉียงเหนือในเขตพื้นที่ชนบทปี 2553 ที่คำนวณโดยสำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติซึ่ง จากการคำนวณได้เส้นความยากจนเท่ากับ 1,565 บาทต่อคนต่อเดือน เป็นเกณฑ์ในการแบ่งกลุ่มครัวเรือนยากจนกับกลุ่มครัวเรือนที่ ไม่ยากจน ผลการศึกษา พบว่า ครัวเรือนในชนบทของจังหวัดมหาสารคามมีจำนวนสมาชิกในครัวเรือนเฉลี่ยครัวเรือนละ 4.16 คน มีจำนวนแรงงานในครัวเรือน เฉลี่ยครัวเรือนละ 3.15 คน และจำนวนสมาชิกที่มีรายได้ในครัวเรือนเฉลี่ยครัวเรือนละ 2.13 คน ระดับการศึกษาของหัวหน้าครัวเรือนของกลุ่มตัวอย่างส่วนใหญ่สำเร็จการศึกษาในระดับประถมศึกษาร้อยละ 49.3 อาชีพของหัวหน้าครัวเรือนส่วนใหญ่ประกอบอาชีพ เกษตรกรรมคิดเป็นร้อยละ 57.5 ครัวเรือนมีรายได้รวมเฉลี่ยครัวเรือนละ 16,036.89 บาทต่อเดือน และมีค่าใช้จ่ายสำหรับการอุปโภค บริโภค เฉลี่ยครัวเรือนละ 7,666 บาทต่อเดือน เมื่อคิดเป็นอัตราส่วนร้อยละของค่าใช้จ่ายอุปโภคบริโภคต่อรายได้จะเท่ากับ 47.80 มีหนี้สินเฉลี่ยครัวเรือนละ 187,530.38 บาท และครัวเรือนมีการเก็บออมคิดเป็นร้อยละ 76.5 ของจำนวนครัวเรือนตัวอย่างทั้งหมด สถานการณ์ความยากจนและลักษณะของครัวเรือนที่ยากจนพบว่า ครัวเรือนตัวอย่างในเขตพื้นที่ชนบทของจังหวัดมหาสารคามมี สัดส่วนของครัวเรือนที่ยากจนคิดเป็นร้อยละ 31.2 โดยครัวเรือนที่ยากจนในเขตชนบทจะมีลักษณะร่วมคือ หัวหน้าครัวเรือนมีระดับ การศึกษาต่ำ มีครัวเรือนขนาดใหญ่ มีระดับรายได้ต่ำ มีขนาดพื้นที่ที่ใช้ในการประกอบอาชีพการเกษตรน้อย มีระดับความมั่งคั่งต่ำ และมีหนี้สิน ส่วนปัจจัยที่มีผลต่อความยากจนของครัวเรือน ได้แก่ ระดับการศึกษาของหัวหน้าครัวเรือน ขนาดของครัวเรือน ขนาดพื้นที่ที่ใช้ ในการประกอบอาชีพ ความมั่งคั่งและหนี้สินของครัวเรือน

ภัทรพงศ์ พงศ์ภัทรกานต์, วิชัย พัวรุ่งโรจน์, คมยุทธ ไชยวงษ์, สุชาดา พรหมโคตร และปาริชาติ แสงระษฎ์ (2560) ได้ศึกษาการใช้เทคนิคเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยในการใช้บริการห้องสมุดของนักศึกษา งานวิจัยนี้นำเสนอการทดสอบวิเคราะห์ปัจจัยในการใช้บริการห้องสมุดของนักศึกษา มหาวิทยาลัยราชภัฏเลย โดย ใช้ข้อมูลการเข้าใช้บริการผ่านประตูอัตโนมัติในช่วงเดือนกุมภาพันธ์ถึง ตุลาคม 2559 ที่มี 9 ปัจจัยพื้นฐาน คือ วันที่เข้าใช้บริการ ช่วงเวลา เพศ คณะ ชั้นปี จังหวัดที่เกิด หมู่ เลือด จำนวนพี่น้อง และเกรดเฉลี่ยสะสม จำนวน 79,953 ชุดข้อมูล ทำการประมวลผลด้วยอัลกอริทึม C5.0, Neural Network และ CART เพื่อศึกษาและ

เปรียบเทียบประสิทธิภาพของการคัดแยกข้อมูล ผลการศึกษา พบว่า อัลกอริทึม C5.0 ให้ค่าความถูกต้อง 97.78 % และใช้ระยะเวลาในการประมวลผล น้อยกว่าอัลกอริทึมที่นำมาเปรียบเทียบ ผลจากการวิเคราะห์ด้วย อัลกอริทึม C5.0 พบว่ามี 3 ปัจจัยที่มี อิทธิพลต่อการใช้บริการห้องสมุดของนักศึกษาที่ส่งผลตามคณะ คือ เกรดเฉลี่ยสะสม มีอิทธิพลสูงสุด ร้อยละ 93.8 เพศ มีอิทธิพลร้อยละ 6.0 และช่วงเวลา มีอิทธิพลร้อยละ 0.2 ซึ่งนำมาสร้างความสัมพันธ์ ได้ 21 ระดับ ซึ่งเป็นแนวทางในการประชาสัมพันธ์และส่งเสริมนักศึกษาเข้ามาใช้บริการห้องสมุดผ่าน คณะที่สังกัดได้ โดยเฉพาะเกรดเฉลี่ยมีผลอย่างมากในการเข้ามาใช้บริการ นักศึกษาที่มีเกรดสูงมีแนวโน้มการเข้าใช้ห้องสมุดมากกว่านักศึกษาที่มีเกรดต่ำ ดังนั้น ห้องสมุดควรเน้นไปที่การเปิดบริการ หรือเชิญชวนให้นักศึกษาที่มีเกรดน้อยเข้าห้องสมุดมากขึ้น ห้องสมุดควรคิดกิจกรรมส่งเสริมใหม่เพิ่ม มากขึ้นเพื่อให้นักศึกษามีความสนใจในการเข้าใช้ห้องสมุด

รัชพลกลัดชื่น และจรัญแสนราช (2561) การเปรียบเทียบประสิทธิภาพอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา. การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทำนายและคุณลักษณะที่มีต่อผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา โดยทำการศึกษาข้อมูลนักศึกษาระดับประกาศนียบัตรวิชาชีพ จำนวน 5,100 ระเบียน ตั้งแต่ปีการศึกษา 2550 -2559 9 สาขาวิชา 27 คุณลักษณะ โดยใช้เทคนิคการจำแนกข้อมูล 3 เทคนิค ได้แก่ Decision Tree : J48graft, Naïve Bayes และ Rule Induction ทำการเปรียบเทียบประสิทธิภาพตัวแบบการทำนาย ระหว่างการใช้คุณลักษณะทั้งหมดกับการเลือกคุณลักษณะแบบ forward select ทดสอบประสิทธิภาพตัวแบบทำนายด้วยวิธีการ 10-fold cross validation โดยใช้โปรแกรม Rapid Miner Studio 8 จากนั้นนำผลการทดสอบประสิทธิภาพที่มีค่าความถูกต้องที่สูงที่สุด 2 ค่า มาทำการเปรียบเทียบด้วยวิธี T-Test ผลการศึกษาพบว่า การใช้เทคนิค Decision Tree : J48graft ด้วยการเลือกคุณลักษณะแบบ Forward Selection และ การเลือกคุณลักษณะทั้งหมด มีค่าความถูกต้องเท่ากับ 83.08% และ 81.71% ตามลำดับ และทดสอบด้วยวิธี T-Test พบว่าการทดสอบทั้งสองแบบมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จากผลการเปรียบเทียบประสิทธิภาพในครั้งนี้ สามารถนำเทคนิค Decision Tree : J48graft ไปใช้ในการทำนายผลสัมฤทธิ์ทางการเรียน และเป็นแนวทางในการสอนเสริมหรือแนะแนวให้กับนักศึกษาต่อไป

สมใจ ตามแต่รัมย์ (2560) ได้ศึกษาปัจจัยที่ส่งผลต่อระดับความสำเร็จหมู่บ้านเศรษฐกิจพอเพียงต้นแบบอำเภอพานทองจังหวัดชลบุรี. มีวัตถุประสงค์เพื่อศึกษาระดับความสำเร็จของหมู่บ้านเศรษฐกิจพอเพียงต้นแบบ อำเภอพานทอง เพื่อศึกษาปัจจัยที่ส่งผลต่อระดับความสำเร็จของหมู่บ้านเศรษฐกิจพอเพียงต้นแบบอำเภอพานทอง จังหวัดชลบุรีประชากรที่ใช้ในการศึกษาคั้งนี้คือ ประชาชน ที่อาศัยอยู่จริงในหมู่บ้านเศรษฐกิจพอเพียงต้นแบบอำเภอพานทองทั้ง 10 หมู่บ้าน จำนวนประชากร 7,342คน และกลุ่มตัวอย่างจำนวน 380คน เครื่องมือที่ใช้ในการเก็บ รวบรวมข้อมูลการวิจัยครั้งนี้คือแบบสอบถาม และสถิติที่ใช้ในการวิเคราะห์ข้อมูลได้แก่ ค่าความถี่ร้อยละค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน การวิเคราะห์ถดถอยพหุคูณ ด้วยวิธีแบบขั้นตอน ผลการศึกษาพบว่า ระดับความสำ เร็จหมู่บ้านเศรษฐกิจพอเพียงต้นแบบอำเภอพานทองจังหวัดชลบุรีอยู่บนระดับมาก

ประเสริฐ บัวทอง (2560) ได้ศึกษาปัจจัยที่มีผลต่อการตัดสินใจปลูกทุเรียนของเกษตรกรในตำบลอ่างศิระ อำเภอมะขาม จังหวัดจันทบุรี. มีวัตถุประสงค์ ประการแรกเพื่อศึกษาปัจจัยส่วนบุคคลที่มีผล ต่อการตัดสินใจปลูกทุเรียนของเกษตรกรในพื้นที่ ตำบลอ่างศิระ อำเภอมะขาม จังหวัดจันทบุรีและประการที่สองเพื่อศึกษา

ปัจจัยทางเศรษฐกิจและปัจจัยด้านกายภาพที่มีผลต่อการตัดสินใจปลูกทุเรียนของเกษตรกรในพื้นที่ ตำบลอ่างศิร อำเภอมะขาม จังหวัดจันทบุรี โดยมีกลุ่มตัวอย่าง คือเกษตรกรสวนทุเรียนที่ขึ้นทะเบียนการปลูกทุเรียนในตำบลอ่างศิร อำเภอมะขาม จังหวัดจันทบุรี จำนวน 300 ครัวเรือน และสัมภาษณ์เชิงลึก จำนวน 10 ครัวเรือน ผลการวิจัยเป็นไปตามสมมติฐานที่ตั้งไว้พบว่า เกษตรกรส่วนใหญ่เป็นเพศชายอายุ 31-40 ปี ระดับการศึกษาประถมศึกษา รายได้ต่อปีของครอบครัว 750,000-1,000,000 บาท มีจำนวนแรงงานในการปลูกทุเรียน 6-10 คน ต้นทุนในการปลูกทุเรียน 100,001-200,000 บาท สายพันธุ์ทุเรียนที่ปลูกหมอนทอง ลักษณะของดินเป็นดินร่วน ลักษณะพื้นที่เป็นที่ราบสูงขนาดของแหล่งน้ำ มีขนาดตั้งแต่ 1 ไร่ลงมาการคมนาคมสะดวกสายเป็นช่วงขนาดพื้นที่ปลูกทุเรียนมีขนาด 26-50 ไร่ ประสบการณ์ในการปลูกทุเรียน 3-6 ปีการสัมภาษณ์เกษตรกรส่วนใหญ่ให้เหตุผลว่า ทำไมถึงตัดสินใจปลูกทุเรียน เพราะทุเรียนเป็นผลไม้ที่มีความต้องการของตลาดสูง

วีระยุทธพิมพาพร และพยุหมีสัง (2557) ได้ศึกษาการวิเคราะห์องค์ประกอบของชุดข้อมูลที่ซับซ้อนด้วยวิธีการเลือกคุณลักษณะสำคัญแบบพลวัต มีวัตถุประสงค์เพื่อศึกษากระบวนการวิเคราะห์องค์ประกอบ (Factor Analysis) บนชุดข้อมูลที่ซับซ้อนด้วยวิธีการ เลือกคุณลักษณะสำคัญแบบพลวัต (Dynamic Feature Selection : DFS) โดยประยุกต์ใช้กระบวนการเลือกตัวแปร (Feature Selection) และการวิเคราะห์กลุ่ม (Clustering analysis) ข้อมูลในการประมวลผลเกิดจากกิจกรรมต่าง ๆ ในระบบการเรียนออนไลน์ (e-Learning) โดยเน้นปัจจัยที่ส่งผลโดยตรงต่อผลสัมฤทธิ์ทางการเรียน ผลการวิจัยพบว่าประสิทธิภาพโดยรวมของกระบวนการวิเคราะห์องค์ประกอบ โดยใช้อัลกอริทึมการเลือกคุณลักษณะสำคัญ แบบพลวัต ให้ค่าความถูกต้องสูงสุดที่ 45.17% โดยใช้ 3 ตัวแปร สำหรับกระบวนการวิเคราะห์องค์ประกอบโดยวิธีการคำนวณหาค่า GAIN ของข้อมูลด้วย Information Gain และ Gain ratio ให้ค่าความถูกต้องสูงสุดที่ 44.80% โดยใช้ตัวแปร 7 ตัวแปร จากผลการวิจัยสามารถสรุปได้ว่า อัลกอริทึมการเลือกคุณลักษณะสำคัญแบบพลวัตมีค่าความถูกต้องสูงกว่า และใช้จำนวนตัวแปรที่น้อยกว่า วิธีการคำนวณหาค่า GAIN ของข้อมูลด้วย Information Gain และ Gain ratio

อัจจิมา มณฑาพันธุ์(2562) งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเกี่ยวกับการเปรียบเทียบวิธีการคัดเลือกคุณลักษณะที่สำคัญเพื่อนำมาใช้ในการ ปรับปรุงการพยากรณ์การเป็นมะเร็งเต้านม โดยใช้วิธีการคัดเลือกคุณลักษณะจากเทคนิคต่าง ๆ จำนวน 7 เทคนิค ได้แก่เทคนิค Correlation Based Feature Selection เทคนิค Information Gain เทคนิค Gain Ratio เทคนิค Chi-Square เทคนิค Forward Selection เทคนิค Backward Elimination และเทคนิค Evolutionary Selection หลังจากคัดเลือกคุณลักษณะ ที่สำคัญจึงนำผลที่ได้จากแต่ละเทคนิคมาคำนวณหาประสิทธิภาพในการพยากรณ์การเป็นมะเร็งเต้านมโดยใช้เทคนิคซัพพอร์ต เวกเตอร์แมชชีน ผลการทดลองพบว่าร้อยละของความถูกต้องในการพยากรณ์การเป็นมะเร็งเต้านม จากจำนวนคุณลักษณะของ ข้อมูลทั้งหมด 30 คุณลักษณะเท่ากับ 91.39 ขณะที่เทคนิค Evolutionary Selection ให้ผลดีที่สุดโดยสามารถลดคุณลักษณะ ที่สำคัญเหลือเพียง 16 คุณลักษณะ และให้ผลการวัดค่าความถูกต้องในการพยากรณ์ได้ดีถึงร้อยละ 95.26





c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16
ลำดับที่	จังหวัด	อำเภอ	ตำบล	หมู่	บ้านเลขที่	ลำดับรายการ	รายการ	จำนวน	ราคา	สถานะ	ยอดรวม(บาท)	จำนวน(วง)	ค่าคง/เดิ	หมวดหมู่	การใช้ประโยชน์	
1							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
2							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
3							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
4							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
5							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
6							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
7							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
8							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
9							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
10							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
11							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
12							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
13							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
14							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
15							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
16							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
17							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
18							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
19							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	
20							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง					ใช้ส่วนตัว	

ภาพที่ 8 ข้อมูลจาก Database user01 ส่วนที่ 2 ทรัพย์สินของครัวเรือน

c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13
ลำดับที่	จังหวัด	อำเภอ	ตำบล	หมู่	บ้านเลขที่	ลำดับรายการ	รายการ	จำนวน	ราคา	สถานะ	ยอดรวม(บาท)	จำนวน(วง)	ค่าคง/เดิ
1							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
2							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
3							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
4							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
5							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
6							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
7							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
8							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
9							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
10							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
11							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
12							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
13							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
14							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
15							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
16							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
17							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
18							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
19							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			
20							ซื้อสด/สร้างเอง			ซื้อสด/สร้างเอง			

ภาพที่ 9 ข้อมูลจาก Database user01 ส่วนที่ 3 อาชีพและรายได้ของครัวเรือน

c0	c1	c2	c3	c4	c5	c6	c7	c8	c9
ลำต้นที่	จังหวัด	อำเภอ	ตำบล	หมู่	บ้านเลขที่	ลำต้นรายการ	รายการค่า	บาท/เดือน	รวม/ปี
1						ค่าปลูก	2000	2400	
						ทุนในการ	1500	1800	
						ค่าไฟ	60	720	
						ค่าไฟ	150	1800	
						ค่าโทรศัพท์	500	6000	
						ค่าการศึกษา	1500	18000	
						ชำระหนี้		10000	
						ค่าอื่นๆ	1000	12000	
						ค่าปลูก	4000	48000	
						ทุนในการ	2000	14000	
						ค่าเลี้ยงดู	3000	36000	
						ค่าไฟ	280	3360	
						ค่าโทรศัพท์	600	7200	
						ชำระหนี้		25800	
						ค่าจตุรกร	6200	74400	
						ค่าอื่นๆ	1500	18000	
						ค่าปลูก	3000	36000	
						ทุนในการ	1000	12000	

ภาพที่ 10 ข้อมูลจาก Database user01 ส่วนที่ 4 รายการของครัวเรือน

c0	c1	c2	c3	c4	c5	c6	c7	c8	c9
ลำต้นที่	จังหวัด	อำเภอ	ตำบล	หมู่	บ้านเลขที่	ลำต้นรายการ	แหล่งเงิน	ปริมาณ	เงิน
1						ส่งบิล	10000		
						ส่งเดือน	1287.50	ส่งงวด	
						ส่งบิล	22000		
						ส่งบิล	14088		
						ส่งทั้งต้นทั้งดอก	เบี้ย	กู้คืนได้	
						ส่งทั้งต้นทั้งดอก	เบี้ย	กู้คืนได้	
						ส่งดอกเบี้ย	เดือนละ	2700	
						ส่งบิล	13500		
						ส่งบิล	2100		
						ส่ง	14000	พร้อมดอกเบี้ย	
						ชำระบิล	1	ครั้ง	
						ชำระรายปี	อัตราดอกเบี้ย	1% ต่อปี	
						ชำระรายเดือน	ในระยะ	เวลา 1 ปี ดอกเบี้ย	0% ต่อเดือน
						ส่ง 1 ปี	(หมดแล้ว)		
						ส่งเดือน	ละ	9000	
						ส่งบิล	32500		
						ส่งบิล	ละ	2250	

ภาพที่ 11 ข้อมูลจาก Database user01 ส่วนที่ 5 หนี้สินของครัวเรือน

c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13
ลำดับที่	จังหวัด	อำเภอ	ตำบล	หมู่	บ้านเลขที่	ได้รับผลก	วิธีชีวิตประ	อาชีพ	รายได้	การศึกษา	สุขภาพ	อื่นๆ	ต้องการความช่วยเหลือ
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													

ภาพที่ 12 ข้อมูลจาก Database user01 ส่วนที่ 6 ผลกระทบจากสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID - 19)

c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18
ลำดับที่	จังหวัด	อำเภอ	ตำบล	หมู่	บ้านเลขที่	สมารถไฟ	คอมพิวเตอร์	คอมพิวเตอร์	แท็บเล็ต	ไอแพด	ช่องทางรับ	ใช้คอมพิวเตอร์	ครัวเรือนมี	ใช้เทคโนโลยี	ใช้เทคโนโลยี	ใช้เทคโนโลยี	ใช้เทคโนโลยี	ใช้เทคโนโลยี
1																		
2																		
3																		
4																		
5																		
6																		
7																		
8																		
9																		
10																		
11																		
12																		
13																		
14																		
15																		
16																		
17																		
18																		

ภาพที่ 13 ข้อมูลจาก Database user01 ส่วนที่ 7 การใช้เทคโนโลยีสารสนเทศ



c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10
ลำดับที่	จังหวัด	อำเภอ	ตำบล	หมู่	บ้านเลขที่	ลำดับรายชื่อ	ชื่อ(การละ	จัดขึ้นใน	วัตถุประสงค์	ประเภท/หน้าที่
1						1	วันขึ้นปีใหม่	เดือน1	รวมญาติ	เป็น
2						2	สงกรานต์	เดือน4	ทำบุญ	เป็น
3						3	ทำบุญ	วันพระ	ร่วม	เป็น
4						4	งานบุญ	ทุกปี	รวม	เป็น
5						5	บุญสงกรานต์	ขึ้น15	เดือน6	เป็น
6						6	ลอยกระทง	เดือน11	ขอพร	เป็น
7						7	สงกรานต์	เดือน12	เป็น	เป็น
8						8	งานบุญ	ทุกปี	รวม	เป็น
9						9	บุญสงกรานต์	ขึ้น15	เดือน6	เป็น
10						10	สงกรานต์	เดือน12	เป็น	เป็น
11						11	ปีใหม่	เดือน1	รวมญาติ	เป็น
12						12	สงกรานต์	เดือน4	ทำบุญ	เป็น
13						13	งานบุญ	ทุกปี	รวม	เป็น
14						14	บุญสงกรานต์	ขึ้น15	เดือน6	เป็น
15						15	ลอยกระทง	เดือน11	ขอพร	เป็น
16						16	สงกรานต์	เดือน12	เป็น	เป็น
17						17	ปีใหม่	เดือน1	รวมญาติ	เป็น
18						18	สงกรานต์	เดือน4	ทำบุญ	เป็น
19						19	งานบุญ	ทุกปี	รวม	เป็น
20						20	บุญสงกรานต์	ขึ้น15	เดือน6	เป็น

ภาพที่ 14 ข้อมูลจาก Database user01 ส่วนที่ 8 การเข้าร่วมการละเล่น การพ้อง การรำ พิธีกรรมตามวิถีวัฒนธรรมชุมชน

c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11
ลำดับที่	จังหวัด	อำเภอ	ตำบล	หมู่	บ้านเลขที่	ลำดับรายชื่อ	ชื่อ	หน่วยงาน	สถานที่	ประเภท/หน้าที่	วัตถุประสงค์
1						1	หมู่บ้าน	พัฒนา	มีการ	วัด/สิ่งของ(กลาง)	ความรู้ ทักษะ ประสิทธิภาพ
2						2	หมู่บ้าน	วัด	ท่องเที่ยว	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
3						3	ชุมชน	วัด	พัฒนา	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
4						4	ศาสนา	พระ	มหาวิทยาลัย	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
5						5	หมู่บ้าน	พัฒนา	มีการ	วัด/สิ่งของ(กลาง)	ความรู้ ทักษะ ประสิทธิภาพ
6						6	หมู่บ้าน	วัด	ท่องเที่ยว	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
7						7	ชุมชน	วัด	พัฒนา	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
8						8	ศาสนา	พระ	มหาวิทยาลัย	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
9						9	หมู่บ้าน	พัฒนา	มีการ	วัด/สิ่งของ(กลาง)	ความรู้ ทักษะ ประสิทธิภาพ
10						10	หมู่บ้าน	วัด	ท่องเที่ยว	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
11						11	ชุมชน	วัด	พัฒนา	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
12						12	ศาสนา	พระ	มหาวิทยาลัย	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
13						13	หมู่บ้าน	พัฒนา	มีการ	วัด/สิ่งของ(กลาง)	ความรู้ ทักษะ ประสิทธิภาพ
14						14	หมู่บ้าน	วัด	ท่องเที่ยว	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
15						15	ชุมชน	วัด	พัฒนา	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
16						16	ศาสนา	พระ	มหาวิทยาลัย	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
17						17	หมู่บ้าน	พัฒนา	มีการ	วัด/สิ่งของ(กลาง)	ความรู้ ทักษะ ประสิทธิภาพ
18						18	หมู่บ้าน	วัด	ท่องเที่ยว	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
19						19	ชุมชน	วัด	พัฒนา	มีการ	ความรู้ ทักษะ ประสิทธิภาพ
20						20	ศาสนา	พระ	มหาวิทยาลัย	มีการ	ความรู้ ทักษะ ประสิทธิภาพ

ภาพที่ 15 ข้อมูลจาก Database user01 ส่วนที่ 9 การเข้าร่วมโครงการที่ผ่านมาอย่างน้อย 3 ปี

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	c0	c1	c2	c3	c4	c5	c6												
2	1	จักรวรรดิ	กำแพง	ดาบแดง	นป	ปราบเลขที่	ข้อคิดเห็นและข้อเสนอแนะเพิ่มเติม												
3							ไม่มี												
4																			
5																			
6																			
7																			
8																			
9																			
10																			
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			

ภาพที่ 16 ข้อมูลจาก Database user01 ส่วนที่ 10 ข้อคิดเห็นและข้อเสนอแนะเพิ่มเติม

## 11. บรรณานุกรม

- กองนโยบายและแผน. (2560). แผนยุทธศาสตร์มหาวิทยาลัยราชภัฏสกลนคร ระยะ 20 ปี (พ.ศ. 2560 – 2579). สืบค้นเมื่อ 10 มกราคม 2565 จากเว็บไซต์  
<https://www.snru.ac.th/wpcontent/uploads/2018/02/-60-79.pdf?fbclid=IwAR3ac-E3KYNQbiMW8dXJ6Yc3vi3-MZ3JuV517z71ucqScne6TYUmEr6EqjE>
- กิตติคุณ แสงนิล และประสพชัย พสุนนท. (2561). "ความน่าเชื่อถือ ความถูกต้อง ความแม่นยำ และ ความเที่ยงตรง" ความสอดคล้องในวิธีการและความคลาดเคลื่อนจากการวัด ของการวิจัยทางด้านสรีรวิทยาการออกกำลังกาย. *Veridian E-Journal, Science and Technology Silpakorn University*. 5(6) หน้า 1-19.
- ครรชิต มัลลียงศ์. (2553). **ต้นไม้การตัดสินใจ**. สืบค้นเมื่อ 14 มกราคม 2565 จากเว็บไซต์  
<https://kb.hsri.or.th/dspace/handle/11228/2964?locale-attribute=th>
- ณัฐพร เห็นเจริญเลิศ. (2558). การวิเคราะห์ ข้อมูลด้วยเทคนิค Data Mining โดยซอฟต์แวร์ RapidMiner Studio 6 (ขั้นพื้นฐานและปานกลาง). สืบค้นเมื่อ 9 มกราคม 2564 จากเว็บไซต์  
[https://www.stou.ac.th/Schools/sst/main/KM/KM%20Post/58/RapidMiner.pdf?fbclid=IwAR1PBZwc7y0HGPhBqxlvy\\_YkzOuHXG3H0ijFMwbgGFHxU9mC5WxoFcg3lOQ](https://www.stou.ac.th/Schools/sst/main/KM/KM%20Post/58/RapidMiner.pdf?fbclid=IwAR1PBZwc7y0HGPhBqxlvy_YkzOuHXG3H0ijFMwbgGFHxU9mC5WxoFcg3lOQ).
- นิภาพร ชนะมาร และพรณิ สิทธิเดช. (2557). การวิเคราะห์ปัจจัยการเรียนรู้ด้วยการคัดเลือกคุณสมบัติและการพยากรณ์. *วารสารมหาวิทยาลัยราชภัฏสกลนคร*. 6(12), 46 หน้า.
- นิเวศ จิระวิชิตชัย. (2553). การค้นหาเทคนิคเหมืองข้อมูลเพื่อสร้างโมเดลการวิเคราะห์โรคอัตโนมัติ. สืบค้นเมื่อ 11 มกราคม 2565 จากเว็บไซต์ <http://ssruir.ssru.ac.th/handle/ssruir/377>.
- ประเสริฐ บัวทอง. (2560). **ปัจจัยที่มีผลต่อการตัดสินใจปลูกทุเรียนของเกษตรกรในตำบลอ่างศิระ อำเภอมะขาม จังหวัดจันทบุรี**. วิทยานิพนธ์หลักสูตรบริหารธุรกิจมหาบัณฑิต, มหาวิทยาลัยบูรพา, 78 หน้า.
- พิมเพ็ญ พรเฉลิมพงศ์. **accuracy-ความถูกต้อง-ความแม่นยำ**. สืบค้นเมื่อ 20 มกราคม 2565 จากเว็บไซต์  
<http://www.foodnetworksolution.com/wiki/word/4289/accuracy>
- ภัทรพงศ์ พงศ์ภัทรกานต์, วิชัย พัวรุ่งโรจน์, คมยุทธ ไชยวงษ์, สุชาดา พรหมโคตร และ ปาริชาติ แสงระชัย. (2560). การใช้เทคนิคเหมืองข้อมูลเพื่อวิเคราะห์ปัจจัยในการใช้บริการห้องสมุดของนักศึกษา. *PULINET Journal*;4(2), 11-18.
- ภาณุพงศ์ สุขสุวรรณ. (2562). **Model Evaluation, Model Optimization and Deployment**. สืบค้น 9 มกราคม 2565, จาก <https://medium.com/tni-university/model-evaluation-and-deployment-848f33e9b395>
- ภูริพัทธ์ ทองคำ. (2559). **อัลกอริทึมแบบรวมสำหรับการเลือกคุณสมบัติของข้อมูล**. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยธรรมศาสตร์, 119 หน้า.

- มหาวิทยาลัยราชภัฏสกลนคร. (2563). ระบบบันทึกแบบสอบถามสภาพทางเศรษฐกิจครัวเรือนเป้าหมายตามโครงการจ้างงาน ประชาชนที่ได้รับผลกระทบจากสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19). สืบค้น 9 มกราคม 2565, จาก [http://qcovid.snru.ac.th/Default.aspx?fbclid=IwAR3CISMWrlnGeVJl4Ou9UVQP3LXA\\_PkgSrlnLffWEwkHf-bAjUyM\\_mCVoQo](http://qcovid.snru.ac.th/Default.aspx?fbclid=IwAR3CISMWrlnGeVJl4Ou9UVQP3LXA_PkgSrlnLffWEwkHf-bAjUyM_mCVoQo)
- รัชพล กลัดชื่น และจรัญ แสนราช. (2561). การเปรียบเทียบประสิทธิภาพอัลกอริทึมและการคัดเลือกคุณลักษณะที่เหมาะสมเพื่อการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษา. *Research Journal Rajamangala University of Technology Thanyaburi*. 17(1). หน้า 1-10.
- วรายุทธ พลาศร. (2555). การศึกษาปัจจัยที่มีผลต่อความยากจนของครัวเรือนในชนบท:กรณีศึกษาจังหวัดมหาสารคาม. *วารสารมหาวิทยาลัยราชภัฏมหาสารคาม*. 7(1). หน้า 29-38.
- วีระยุทธ พิมพาพร และพยุ่ง มีสัจ. (2557). การวิเคราะห์องค์ประกอบของชุดข้อมูลที่ซับซ้อนด้วยวิธีการเลือกคุณลักษณะสำคัญแบบพลวัต. *วารสารศรีปทุมปริทัศน์*. 6(6). หน้า 90-99.
- สมใจ ตามแต่รัมย์. (2560). ปัจจัยที่ส่งผลต่อระดับความสำเร็จหมู่บ้านเศรษฐกิจพอเพียงต้นแบบอำเภอพานทองจังหวัดชลบุรี. *วิทยานิพนธ์หลักสูตรรัฐประศาสนศาสตรมหาบัณฑิต, มหาวิทยาลัยบูรพา*, 95 หน้า.
- สุพรรณ ฟ้าหยง. (2562). *Machine Learning 4 ประเภท*. สืบค้นเมื่อ 13 มกราคม 2565 จากเว็บไซต์: <http://codeonthehill.com/machine-learning-types>.
- หนึ่งหทัย ชัยอากร. (2559). การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไมน์นิ่ง. สืบค้นเมื่อ 14 มกราคม 2565 จากเว็บไซต์ <https://erp.mju.ac.th/articleDetail.aspx?qid=551>.
- อัจจิมา มณฑาพันธุ์. (2562). การเปรียบเทียบวิธีการคัดเลือกคุณลักษณะที่สำคัญในการปรับปรุงการพยากรณ์มะเร็งเต้านม. *Royal Thai Air Force Medical Gazette*. 65(2). หน้า 49-56
- Achieve.Plus.(2563). *Rapidminer* เสกคนไม่มีพื้นฐานให้เป็นเซียน. สืบค้นเมื่อ 9 มกราคม 2565 จากเว็บไซต์ <https://medium.com/achieve-space/rapidminer-99-9bf6ab20d1aa>
- Mpcrkadmn. (2565). *MY PC Crack Full Version Free Download*. สืบค้นเมื่อ 9 มกราคม 2565 จากเว็บไซต์ [https://mypccrack.com/rapidminer-studio-crack/?fbclid=IwAR20TxwTMGCHbTOMzT5hDSEoEZcupYN-hwXK6Pur3\\_jQQ2AnnKpMqrf\\_nDo](https://mypccrack.com/rapidminer-studio-crack/?fbclid=IwAR20TxwTMGCHbTOMzT5hDSEoEZcupYN-hwXK6Pur3_jQQ2AnnKpMqrf_nDo)
- Nuthdanai wangpratham. (2564). *Predictive Modeling*. สืบค้นเมื่อ 15 มกราคม 2565 จากเว็บไซต์ <https://nutdnuy.medium.com/predictive-modeling-f9881b3e4c02>
- Panupong Sukswan. (2016). *Model Evaluation ,Model Optimization and Deployment*. สืบค้นเมื่อ 14 มกราคม 2565 จากเว็บไซต์
- Phuri Chalermkiatsakul. (2563). *Supervised Learning คืออะไร? ทำงานยังไง?*. สืบค้นเมื่อ 16 มกราคม 2565 จากเว็บไซต์ <https://phuri.medium.com/supervised-learning>  
<https://medium.com/tni-university/model-evaluation-and-deployment-848f33e9b395>.
- Rachot Leingchan. (2564). *เศรษฐกิจไทยจะเป็นอย่างไร หากเราต้องอยู่กับโควิด-19 ไปตลอดกาล*. สืบค้นเมื่อ 17 มกราคม 2565 จากเว็บไซต์ <https://www.krungsri.com/th/research/research-intelligence/ri-covid-recovery-2021?fbclid=>

Thapanee Boonchob. (2563). **เข้าใจ CRISP-DM ฉบับเร่งรัด**. สืบค้นเมื่อ 11 มกราคม 2565 จากเว็บไซต์  
<https://kamboonchob.medium.com/94-b0913050198f>