

## บทที่ 5

### สรุปผล อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอการการคัดเลือกคุณสมบัติด้วยเทคนิค Gain Ratio Feature Selection ให้ทราบถึงประสิทธิภาพความถูกต้องแม่นยำในการคัดเลือกคุณสมบัติ สำหรับสนับสนุนหรือเป็นข้อมูลประกอบการตัดสินใจในการพัฒนาชุมชนท้องถิ่น ดังนั้นในบทนี้จะเป็นการสรุปผลการดำเนินการศึกษาวิจัยทั้งหมดที่ได้นำเสนอรายงานนี้ มีรายละเอียดดังต่อไปนี้

#### 5.1 สรุปผลการวิจัย

วิจัยเรื่อง การวิเคราะห์ปัจจัยสำคัญของข้อมูลเศรษฐกิจครัวเรือนด้วยการคัดเลือกคุณสมบัติ Gain Ratio Feature Selection มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยสำคัญของข้อมูลเศรษฐกิจครัวเรือนด้วยการคัดเลือกคุณสมบัติแบบ Gain Ratio โดยนำเอาเทคนิคการทำเหมืองข้อมูลมาเพื่อวิเคราะห์หาองค์ความรู้ที่อยู่ในข้อมูลจำนวนมาก เพื่อศึกษาการวิเคราะห์ปัจจัยสำคัญที่เหมาะสม ข้อมูลที่ใช้ในการศึกษาวิจัยในครั้งนี้ เป็นข้อมูลประชากรจากภาคครัวเรือน ซึ่งมี 12 หมู่บ้าน จำนวน 2,909 ครัวเรือน โดยช่วงเวลาที่ทำการเก็บรวบรวมข้อมูล คือ ปี พ.ศ. 2563 – 2564 และจากฐานข้อมูลเศรษฐกิจครัวเรือน (สำนักวิทยบริการและเทคโนโลยีสารสนเทศ, 2563: ออนไลน์) ในฐานข้อมูลนี้เป็นข้อมูลจากโครงการศาสตร์พระราชามีการเก็บข้อมูลออกเป็น 10 ส่วน รวมทั้งหมด 178 ปัจจัย จำนวน 17,933 ระเบียบ ผ่านการลดระเบียบครัวเรือนให้เหลือ 1,751 ระเบียบ (ครัวเรือน) เนื่องจากแต่ละครัวเรือนมีสมาชิกแตกต่างกันจึงใช้ขั้นตอนการแปลงรูปแบบข้อมูลให้เหลือครัวเรือนละ 1 ระเบียบ ใช้สำหรับเป็นข้อมูลในการสร้างตัวแบบวิเคราะห์ปัจจัย

เมื่อได้ข้อมูลที่พร้อมแล้วจึงนำข้อมูลไปทำการวิเคราะห์ปัจจัยด้วยโปรแกรม RapidMiner ซึ่งทำการวิเคราะห์ปัจจัยที่สำคัญด้วยเทคนิค Gain Ratio ได้ปัจจัยที่สำคัญออกมา 10 ปัจจัย นำมาสร้างตัวแบบด้วยเทคนิค Decision Tree และ ทำการสุ่มข้อมูลตามค่าสัดส่วนร้อยละ 60:40, 70:30 และ 80:20 ทำการแบ่งข้อมูลสำหรับทดสอบ การเรียนรู้ โดยใช้วิธี 5-Fold Cross Validation และวิธี 10-Fold Cross Validation นำไปเปรียบเทียบระหว่าง All Feature ซึ่งเป็นข้อมูลที่ไม่ได้ผ่านเทคนิคใด ๆ จากนั้นประเมินประสิทธิภาพด้วยค่าค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision: P) ค่าความระลึก (Recall: R) ค่าถ่วงดุล (F-Measure) และค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ย (Root Mean Squared Error: RMSE)

ผลการดำเนินงานวิจัยพบว่า การคัดเลือกคุณสมบัติด้วยเทคนิค Gain Ratio ที่มีการแบ่งข้อมูลสัดส่วนร้อยละ 70:30 10-Fold Cross Validation ให้ค่าความถูกต้องโดยรวมที่เหมาะสมอยู่ที่ 99.51% ค่าความแม่นยำที่เหมาะสมอยู่ที่ 99.61% ค่าความระลึกที่เหมาะสมอยู่ที่ 99.55% ค่าถ่วงดุล

ที่เหมาะสมอยู่ที่ 99.58% และค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยน้อยที่สุดอยู่ที่ 0.062 และผลของชุดข้อมูล All Feature ที่ให้ผลลัพธ์ที่สูงที่สุดคือ ที่มีการแบ่งข้อมูลสัดส่วนร้อยละ 80:20 การแบ่งสัดส่วนทดสอบประสิทธิภาพด้วยวิธี 5-Fold Cross Validation และ วิธี 10-Fold Cross Validation ให้ค่าความถูกต้องโดยรวมสูงสุด 99.43% ค่าความแม่นยำสูงสุด 99.28% ให้ค่าความระลึกสูงที่สุด 99.39% ให้ค่าถ่วงดุลสูงที่สุด 99.67% และค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยน้อยที่สุด 0.075

ดังนั้นวิธีที่เหมาะสมสำหรับการวิเคราะห์ปัจจัยของข้อมูลเศรษฐกิจครัวเรือนคือ การคัดเลือกคุณสมบัติด้วยเทคนิค Gain Ratio Feature Selection เนื่องจากให้ประสิทธิภาพค่าความถูกต้องในการวิเคราะห์ปัจจัยสำคัญได้ดีที่สุด

## 5.2 อภิปรายผล

งานวิจัยนี้เป็นการใช้เทคนิค Gain Ratio เพื่อวิเคราะห์ปัจจัยที่สำคัญ จากผลการวิเคราะห์ตัวแบบพบว่าได้ปัจจัยที่สำคัญทั้งหมด 10 ปัจจัยได้ แก่ รวมรายได้เฉลี่ย/ปี รายงาน รายสูงอายุ รายเรียน รวมค่าใช้จ่าย/ปี การใช้อินเทอร์เน็ตที่ก่อให้เกิดรายได้ อาชีพ รวมปริมาณหนี้สิน มูลค่าทรัพย์สิน และหนี้ในระบบ แล้วนำมาสร้างตัวแบบด้วยเทคนิค Decision Tree ซึ่งจากการทดลองทำให้เห็นว่า การวิเคราะห์ปัจจัยสำคัญด้วย Gain Ratio Feature Selection มีประสิทธิภาพในการวิเคราะห์ปัจจัยที่สำคัญได้ดีที่สุด

ข้อมูลที่ใช้สำหรับการพัฒนาตัวแบบ ผู้วิจัยได้ใช้ข้อมูลจากฐานข้อมูลเศรษฐกิจครัวเรือน โดยคัดเลือกปัจจัยแบบเจาะจงที่สอดคล้องกับเศรษฐกิจครัวเรือน เพื่อนำมาวิเคราะห์ปัจจัยที่สำคัญ จากนั้นจะได้ปัจจัยสำคัญที่มีผลต่อเศรษฐกิจครัวเรือน ซึ่งลักษณะของข้อมูลจะอยู่ในรูปแบบข้อมูล Numeric และ Nominal การเลือกข้อมูลเศรษฐกิจครัวเรือนมาใช้ในการศึกษานั้นจากการศึกษา งานวิจัยที่เกี่ยวข้องพบว่าอายุ และรายได้ มีผลค่อนข้างมากต่อประสิทธิภาพในการพยากรณ์เศรษฐกิจครัวเรือน ซึ่งสอดคล้องกับงานวิจัยของ วราวุธ พลาศศรี (2556) ได้อธิบายว่าสภาพเศรษฐกิจของครัวเรือนในชนบท สถานการณ์ความยากจน ลักษณะของครัวเรือนที่ยากจน และปัจจัยที่มีผลต่อความยากจนของครัวเรือนในชนบทจังหวัดมหาสารคาม โดยกลุ่มประชากรที่ใช้ในการศึกษาคือ ครัวเรือนที่อยู่ในเขตพื้นที่ชนบทจังหวัดมหาสารคาม จำนวน 180,328 ครัวเรือน ขนาดกลุ่มตัวอย่างเท่ากับ 400 ครัวเรือน โดยครัวเรือนที่ยากจนในเขตชนบทจะมีลักษณะรวมคือ หัวหน้าครัวเรือนมีระดับการศึกษาต่ำ มีครัวเรือนขนาดใหญ่ มีระดับรายได้ต่ำ มีขนาดพื้นที่ที่ใช้ในการประกอบอาชีพการเกษตรน้อย มีระดับความมั่งคั่งต่ำและมีหนี้สิน

การคัดเลือกวิเคราะห์ปัจจัยที่สำคัญผู้วิจัยเลือกใช้เทคนิค Gain Ratio และวิเคราะห์ปัจจัยที่สำคัญด้วยเทคนิค Decision Tree โดยเลือกจากความนิยมและผลลัพธ์ที่ได้จากงานวิจัยที่เกี่ยวข้องพบว่าใช้เทคนิค Gain Ratio และเทคนิค Decision Tree ในการวิเคราะห์ปัจจัยที่สำคัญ จะให้ผลการคัดเลือกที่ดีและมีประสิทธิภาพ เมื่อได้ทำการทดลองแล้วพบว่าปัจจัยที่ได้จากเทคนิค Gain

Ratio มีประสิทธิภาพที่ดีกว่า All Feature ซึ่งสอดคล้องกับงานวิจัยของ จิรัฏฐา ภูบุญชอบ (2554) ได้อธิบายว่า การเปรียบเทียบประสิทธิภาพการคัดเลือกคุณลักษณะ 2 แบบ ได้แก่ 1) การเลือกคุณลักษณะแบบควบรวม (Wrapper) โดยใช้วิธี Backward Elimination และ 2) การเลือกคุณลักษณะแบบกรอง (Filter) ด้วยวิธี Gain ratio จากนั้นจำแนกข้อมูลด้วยเทคนิค Neural Network เพื่อสร้างแบบจำลองในการพยากรณ์การออกกลางคันของนักศึกษาโดยการลดการนำเข้าที่ละตัวแปร ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลที่มีความสัมพันธ์ต่อการออกกลางคันของนักเรียน นักศึกษา ปีการศึกษา 2554 – 2558 ประกอบด้วย 25 ตัวแปร 2,803 ข้อมูล

### 5.3 ข้อจำกัด และข้อเสนอแนะ

วิธีการทำเหมืองข้อมูลมีหลากหลายวิธี ซึ่งแต่ละวิธีนั้นมีความเหมาะสมที่แตกต่างกันไป ดังนั้นควรจะนำเทคนิคอื่น มาวิเคราะห์เพื่อเปรียบเทียบหาวิธีที่เหมาะสม