

ชื่อเรื่อง	การวิเคราะห์ปัจจัยสำคัญของข้อมูลเศรษฐกิจครัวเรือนด้วยการคัดเลือกคุณสมบัติ Gain Ratio Feature Selection
ชื่อนักศึกษา	นางสาวชราภรณ์ เจริญมา รหัส 62102105141
ปริญญา	วิทยาศาสตรบัณฑิต
สาขาวิชา	วิทยาการคอมพิวเตอร์
ปีการศึกษา	2565
อาจารย์ที่ปรึกษา	ดร.นิภาพร ชนะมาร

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยสำคัญของข้อมูลเศรษฐกิจครัวเรือนด้วยการคัดเลือกคุณสมบัติแบบ Gain Ratio Feature Selection ประยุกต์ใช้เทคนิคเหมืองข้อมูล โดยใช้เทคนิคการคัดเลือกคุณสมบัติ (Gain Ratio) สร้างตัวแบบด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) นำข้อมูลปัจจัยทั้งหมด All Feature มาเปรียบเทียบกับวิธีการคัดเลือกคุณสมบัติแบบ Gain Ratio ข้อมูลประชากรจากภาคครัวเรือน ซึ่งมี 12 หมู่บ้าน จำนวน 2,909 ครัวเรือน โดยช่วงเวลาที่ทำการเก็บรวบรวมข้อมูล คือ ปี พ.ศ. 2563 – 2564 รวมทั้ง 178 ปัจจัย จำนวน 17,933 ระเบียบ ผ่านการลดระเบียนครัวเรือนให้เหลือ 1,751 ระเบียบ (ครัวเรือน) 16 ปัจจัย ได้แก่ วัยเรียน วัยทำงาน วัยสูงอายุ อาชีพ รวมรายได้เฉลี่ย/ปี มูลค่าทรัพย์สิน วัตถุประสงค์การเลี้ยงสัตว์ พื้นที่ก่อให้เกิดรายได้ ต้นทุนการผลิตทำการเกษตร ผลผลิตที่ได้จากการทำการเกษตร รวมค่าใช้จ่าย/ปี หนี้ในระบบ หนี้ในระบบ รวมปริมาณหนี้สิน การใช้อินเทอร์เน็ต ที่ก่อให้เกิดรายได้ ช่องทางการขายสินค้าที่ก่อให้เกิดรายได้ และเมื่อคัดเลือกคุณสมบัติด้วยเทคนิค Gain Ratio ได้ 10 ปัจจัยสำคัญ ได้แก่ วัยเรียน วัยทำงาน วัยสูงอายุ อาชีพ มูลค่าทรัพย์สิน หนี้ในระบบ รวมปริมาณหนี้สิน รวมค่าใช้จ่าย/ปี รวมรายได้เฉลี่ย/ปี การใช้อินเทอร์เน็ตที่ก่อให้เกิดรายได้ จากนั้นจะทำการสุ่มข้อมูลตามค่าสัดส่วนร้อยละ 60:40 70:30 และ 80:20 และเปรียบเทียบประสิทธิภาพของตัวแบบด้วยวิธี 5-Fold Cross Validation และ 10-Fold Cross Validation พัฒนาด้วยโปรแกรม RapidMiner Studio พบว่า เทคนิคที่มีประสิทธิภาพการวิเคราะห์ปัจจัยสำคัญของข้อมูลเศรษฐกิจครัวเรือนดีที่สุดคือ การคัดเลือกคุณสมบัติด้วยเทคนิค Gain Ratio ที่มีการสุ่มแบ่งข้อมูลแบบร้อยละ 70:30 ด้วยวิธี 10-Fold Cross Validation ค่า Maximal Depth มีค่าเท่ากับ 10 ให้ค่าความถูกต้องโดยรวมเหมาะสมที่สุดอยู่ที่ 99.51% ค่าความแม่นยำอยู่ที่ 99.61% ค่าความระลึกอยู่ที่ 99.55% ค่าถ่วงดุลอยู่ที่ 99.58% และค่ารากที่สองของค่าความคลาดเคลื่อนเฉลี่ยน้อยที่สุดเท่ากับ 0.062

คำสำคัญ : เหมืองข้อมูล การคัดเลือกคุณสมบัติ เทคนิคต้นไม้ตัดสินใจ เทคนิคแกนโรโซ