

# การจัดกลุ่มข้อมูลเศรษฐกิจครัวเรือนด้วยการประยุกต์ใช้เทคนิคการจัดกลุ่ม แบบ K-Means Clustering

## Clustering of Household Economic Data Using K-Means Clustering Techniques

เจนจิรา สุกใส<sup>1</sup> และ นิภาพร ชนะมาร<sup>2\*</sup>

Janejira Suksai<sup>1</sup> and Nipaporn Chanamarn<sup>2\*</sup>

<sup>1, 2</sup>สาขาวิชาคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร 47000

<sup>1, 2</sup>Department of Computer, Faculty of Science and technology, Sakon Nakhon Rajabhat University,  
Sakon Nakhon Province, 47000

\*Corresponding author E-mail: nipaporn@snru.ac.th

Received 8 September 2022, Accepted 27 December 2022, Published 1 March 2023

### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อจัดกลุ่มข้อมูลเศรษฐกิจครัวเรือนด้วยการประยุกต์ใช้ K-Means Clustering โดยใช้ข้อมูลเศรษฐกิจครัวเรือนจากฐานข้อมูลตำบลและฐานข้อมูลขนาดใหญ่สำหรับการพัฒนาท้องถิ่นอย่างยั่งยืน ทำการเก็บรวบรวมข้อมูล ตั้งแต่ปี พ.ศ. 2563 – 2564 จำนวน 12 หมู่บ้าน 2,909 ครัวเรือน ซึ่งมีการเก็บข้อมูลทั้งหมด 10 ส่วน มี 17,933 ระเบียบ และ 178 ปัจจัย จากนั้นผู้วิจัยได้ทำความสะอาดข้อมูล คัดเลือกข้อมูล ลดมิติข้อมูล และแปลงรูปแบบข้อมูล ตามกระบวนการของ CRISP-DM ได้ข้อมูลทดลอง จำนวน 1,751 ระเบียบ 16 ปัจจัย โดยใช้เทคนิค K-Means Clustering และพัฒนาด้วยโปรแกรม RapidMiner เวอร์ชัน 9.10 ด้วยโอเพอเรเตอร์ Clustering ผลการวิจัยพบว่า การจัดกลุ่มเศรษฐกิจครัวเรือน สามารถจัดกลุ่มได้ 3 กลุ่ม (K=3) โดยพิจารณาจากค่า Centroid ที่เหมาะสม ดังนี้ ครัวเรือนที่อยู่ในระดับเศรษฐกิจน้อย มีจำนวน 304 ครัวเรือน (-3.375) ครัวเรือนที่อยู่ในระดับเศรษฐกิจปานกลาง มีจำนวน 544 ครัวเรือน (-2.597) และ ครัวเรือนที่อยู่ในระดับเศรษฐกิจสูง มีจำนวน 903 ครัวเรือน (-2.355) จากผลที่ได้สามารถนำไปประกอบการพิจารณาการตัดสินใจในการพัฒนาและส่งเสริมอาชีพเพื่อยกระดับรายได้ของครัวเรือนในแต่ละกลุ่มได้

**คำสำคัญ:** การทำเหมืองข้อมูล การจัดกลุ่ม ข้อมูลเศรษฐกิจครัวเรือน

### ABSTRACT

The goal of this study is to cluster household economic data using K-Means clustering. For sustainable local development, it makes use of household economic data from parish records and big databases. Data was gathered between 2020 and 2021 in 2,909 households across 12 villages. There are 178 factors, 17,933 records,



and 10 parts of data. After that, the researcher used data cleansing, data selection, dimension reduction, and data transformation to implement the CRISP-DM procedure. For experimental data, there are 16 factors and 1,751 records. Rapid Miner 9.10 with K-mean clustering operators is used during development. The findings demonstrated that family economic grouping can be divided into three groups based on the correct centroid values: The low economy had 304 households (Centroid = -3.375), the middle economy had 544 households (Centroid = -2.597) and the high economy had 903 households (Centroid = -2.355). The data can be used to determine decisions for career development and promotion in order to raise household income in each group.

**Keywords:** Data Mining, Clustering Analysis, Household Economic Data

## บทนำ

มหาวิทยาลัยราชภัฏสกลนคร ได้มีการเก็บรวบรวมข้อมูลเศรษฐกิจครัวเรือน ในฐานะข้อมูลตำบลและฐานข้อมูลขนาดใหญ่ สำหรับการพัฒนาท้องถิ่นอย่างยั่งยืน ตั้งแต่ปี พ.ศ. 2563 – 2564 จากข้อมูลดังกล่าว โดยมีการเก็บข้อมูลออกเป็น 10 ส่วน คือ 1) ข้อมูลทั่วไปครัวเรือน 2) ทรัพย์สินของครัวเรือน 3) อาชีพและรายได้ของครัวเรือน 4) รายจ่ายของครัวเรือน 5) หนี้สินของครัวเรือน 6) ผลกระทบจากสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID - 19) 7) การใช้เทคโนโลยีสารสนเทศ 8) การเข้าร่วมการละเล่น การฟ้อน การรำ พิธีกรรมตามวิถีวัฒนธรรมชุมชน 9) การเข้าร่วมโครงการที่ผ่านมา ย้อนหลัง 3 ปี และ 10) ข้อคิดเห็นและข้อเสนอแนะเพิ่มเติม (มหาวิทยาลัยราชภัฏสกลนคร, 2563) ซึ่งผู้วิจัยมีความสนใจที่จะนำข้อมูลมาทำการแบ่งระดับเศรษฐกิจครัวเรือน เพื่อส่งเสริมงานวิจัยและใช้ข้อมูลประกอบการตัดสินใจ ในการวางแผนพัฒนาเศรษฐกิจครัวเรือนในชุมชนเป้าหมาย

ในปัจจุบันมีเทคโนโลยีมากมายที่ใช้ในการวิเคราะห์ข้อมูล โดยเฉพาะอย่างยิ่งเทคนิคการทำเหมืองข้อมูลที่อยู่ในศาสตร์วิทยาการข้อมูล ที่กำลังเป็นที่นิยม ซึ่งมีงานวิจัยหลากหลายงานวิจัยที่ได้นำไปประยุกต์ใช้ เช่น วนิชา แผลงรักษา และนิเวศ จิระวิชิตชัย (วนิชา และนิเวศ, 2562) ได้ศึกษาการแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบ K-means สำหรับการบริหารลูกค้าสัมพันธ์ โดยทดสอบกับระบบฐานข้อมูลเกี่ยวกับลักษณะลูกค้าของผู้จัดจำหน่ายสินค้าออนไลน์แห่งหนึ่ง ทั้งหมดจำนวน 1,000 คน เพื่อนำมาวิเคราะห์การแบ่งกลุ่มคลัสเตอร์ด้วย K-means ผลการทดลองโดยใช้เทคนิคการทำคลัสเตอร์แบบ K-means สามารถแบ่งกลุ่มลูกค้าที่มีคุณลักษณะเหมือนกันไว้อยู่ในกลุ่มเดียวกันได้อย่างมีประสิทธิภาพและสามารถนำไปใช้เพื่อการส่งเสริมการขายได้ตรงกลุ่มเป้าหมาย งานวิจัยของ ชัดชัย แก้วตา และชนัญญาญจน์ แสงประสาน (ชัดชัย และชนัญญาญจน์, 2564) ได้ศึกษาการเลือกพืชสวนเศรษฐกิจเพื่อการเพาะปลูกที่เหมาะสม ในจังหวัดอุบลราชธานี โดยใช้เทคนิคการทำเหมืองข้อมูล 2 ขั้นตอน คือ การจัดกลุ่มด้วยเทคนิค K-means และการจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจ ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูล ปี พ.ศ. 2558-2560 พืชสวนเศรษฐกิจ จำนวน 14 ชนิด ผลการจัดกลุ่มพื้นที่ด้วยเทคนิค K-means พบว่า พื้นที่จังหวัดอุบลราชธานี ถูกจัดอยู่ในกลุ่มที่ 3 มีสมาชิก 15 จังหวัด รวมทั้งหมด 207 อำเภอ เมื่อนำรูปแบบการเลือกพืชสวนเศรษฐกิจมาพัฒนาโปรแกรมในรูปแบบเว็บแอปพลิเคชันกลุ่มตัวอย่างเกษตรกรมีความพึงพอใจต่อโปรแกรมการเลือกพืชสวนเศรษฐกิจที่เหมาะสมกับพื้นที่เพาะปลูกของจังหวัดอุบลราชธานี อยู่ในระดับมาก และวิจัยของ ดารณี พิมพ์ช่างทอง (ดารณี, 2561) ได้ศึกษาการวิเคราะห์จัดกลุ่มเพื่อการรณรงค์ทางการตลาดด้วยการใช้เครือข่ายสังคมออนไลน์ กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ บุคคลทั่วไป ประเภท Facebook, Line และ Instagram โดยทำการเก็บตัวอย่างทั้งหมด 400 ตัวอย่าง ใช้วิธีการสุ่มตัวอย่าง สติติที่ใช้ในการวิเคราะห์



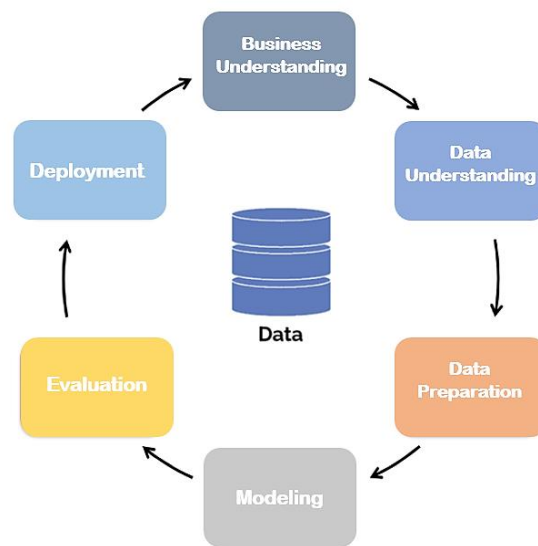
คือ ค่าร้อยละ ค่าความถี่ โดยใช้ Operator K-Mean ผลการวิเคราะห์กลุ่มข้อมูลสามารถจำแนกเป็น 2 กลุ่ม คือ กลุ่มที่ตระหนักถึงสินค้าและบริการ และกลุ่มที่ตระหนักถึงราคา เพื่อการณรงค์ทางการตลาดผ่านสื่อสังคมออนไลน์ ด้วยการผลิตสื่อโฆษณาที่ดึงดูดความสนใจ และสร้างการมีส่วนร่วมของผู้ชมซึ่งจะนำไปสู่การซื้อที่เพิ่มขึ้น

จากที่กล่าวมาข้างต้นผู้วิจัยจึงมีความสนใจที่จะนำ เทคนิค K-Means Clustering มาทำการจัดกลุ่มข้อมูล เพื่อให้สามารถนำข้อมูลมาสนับสนุนในการตัดสินใจในการพัฒนาชุมชนท้องถิ่นได้

## เอกสารและงานวิจัยที่เกี่ยวข้อง

### 1. การทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data mining) เป็นกระบวนการในการค้นหารูปแบบในชุดข้อมูลขนาดใหญ่ โดยใช้วิธีการของการเรียนรู้ของเครื่อง สถิติ และระบบฐานข้อมูล เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์หรือสร้างตัวแบบสำหรับการจำแนกหน่วยหรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยต่างๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล การทำเหมืองข้อมูล ประกอบขึ้นด้วยการนำกระบวนการทางสถิติและการเรียนรู้ผ่านระบบคอมพิวเตอร์เพื่อสร้างตัวแบบ กฎเกณฑ์ รูปแบบ การพยากรณ์และข้อความรู้จากฐานข้อมูลขนาดใหญ่ (Thaidatascience, 2017) โดยการทำให้เหมืองข้อมูล ที่นิยมใช้ในปัจจุบันมีหลายแบบแต่ที่จะนำมากล่าวคือวิธี Cross-Industry Standard Process for Data Mining (CRISP-DM) (Thapanee, 2020) การทำเหมืองข้อมูลมีหลายรูปแบบ ได้แก่ กฎความสัมพันธ์ (Association Rule) การจำแนกประเภทข้อมูล (Data classification) การแบ่งกลุ่มข้อมูล (Data Clustering) (Big Data, 2020)



ภาพที่ 1 กระบวนการ CRIPS-DM

ที่มา: (Thapanee, 2020)

จากภาพที่ 1 กระบวนการ CRIPS-DM มีทั้งหมด 6 ขั้นตอน ดังนี้

- 1) Business Understanding เป็นการทำความเข้าใจกับปัญหาหรือประเด็นที่ต้องการวิเคราะห์
- 2) Data Understanding เป็นการทำความเข้าใจกับข้อมูล การรวบรวมข้อมูลสำหรับนำมาวิเคราะห์เพื่อแก้ปัญหาที่ต้องการ



3) Data Preparation เป็นการนำข้อมูลที่รวบรวมได้มาทำการกลั่นกรองข้อมูล ทำความสะอาดข้อมูล การเลือกข้อมูล หรือลดมิติข้อมูล และทำการแปลงข้อมูล เพื่อให้ได้ข้อมูลที่เหมาะสมในการสร้างรูปแบบหรือการพยากรณ์ข้อมูล ซึ่งขั้นตอนนี้จะใช้เวลานานในการวิเคราะห์ข้อมูล

4) Modeling เป็นขั้นตอนการสร้างรูปแบบหรือตัวแบบการพยากรณ์ หรือการจัดกลุ่ม ตามลักษณะข้อมูลที่เหมาะสมกับเทคนิคที่เลือกใช้ในการสร้างรูปแบบ

5) Evaluation เป็นขั้นตอนในการประเมินประสิทธิภาพหรือประเมินผลตัวแบบหรือรูปแบบที่ได้สร้างขึ้น ซึ่งลักษณะการประเมินขึ้นอยู่กับประเภทของเทคนิคที่เลือกใช้

6) Deployment เป็นขั้นตอนในการนำตัวแบบที่ได้ไปประยุกต์ใช้งานจริง

## 2. การแบ่งกลุ่มข้อมูล (Data clustering)

การแบ่งกลุ่มข้อมูล เป็นวิธีการจัดกลุ่มข้อมูล การรวมกลุ่มข้อมูลที่มีลักษณะเหมือนกัน รูปแบบและแนวโน้มที่เหมือนกันไว้ในกลุ่มเดียวกัน โดยเริ่มจากการหาตัวแทนกลุ่ม จากนั้นทำการเปรียบเทียบข้อมูลกับตัวแทนของแต่ละกลุ่ม ถ้าข้อมูลคล้ายคลึงกับตัวแทนกลุ่มไหนก็จะถูกจัดเข้ากลุ่มนั้น ซึ่งการแบ่งกลุ่มข้อมูล เป็นการเรียนรู้แบบไม่มีผู้สอน ขั้นตอนวิธีการแบ่งกลุ่ม ได้แก่ K-means clustering, Hierarchical clustering, Self-Organizing Map ซึ่งในงานวิจัยนี้จะใช้วิธี K-means clustering (ชฎารัตน์ และศิริธร, 2559)

### 2.1 เทคนิคการจัดกลุ่มด้วย K-means

K-means คือ วิธีการหนึ่งใน Data mining อยู่ในกลุ่มของ Unsupervised Learning คือการเรียนรู้แบบไม่ต้องสอน ที่สามารถเข้าใจได้ง่าย โดยจะแบ่งกลุ่มข้อมูลที่มีลักษณะของข้อมูลที่คล้ายกันรวมไว้ด้วยกันเป็นกลุ่ม ๆ โดยลักษณะความคล้ายวัดจากระยะห่างแบบยูคลิด (Euclidean Distance) จากจุดศูนย์กลางกลุ่มไปยังข้อมูลแต่ละตัว (สำราญ และรจนา, 2563) โดยมีขั้นตอนดังนี้

- 1) กำหนดจุดศูนย์กลางเริ่มต้นแบบสุ่มจำนวน k กลุ่มจุด
- 2) จัดข้อมูลเป็นกลุ่มด้วยการหาระยะห่างระหว่างข้อมูลกับจุดศูนย์กลางกลุ่ม จัดให้ข้อมูลอยู่กลุ่มที่ข้อมูลมีระยะห่างจากจุดศูนย์กลางน้อยที่สุด ด้วยการวัดแบบยูคลิด ดังสมการ (1)

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

- 3) ปรับปรุงจุดศูนย์กลางกลุ่มใหม่ด้วยการหาค่าเฉลี่ยจากข้อมูลทุกตัวภายในกลุ่ม
- 4) ทำซ้ำตั้งแต่ข้อ 2) จนกระทั่งค่าเฉลี่ยของจุดศูนย์กลางในแต่ละกลุ่มไม่มีการเปลี่ยนแปลง

## 3. งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ทำการค้นคว้างานวิจัยด้านเหมืองข้อมูลที่เกี่ยวข้องวิธีการนำมาใช้ในการวิจัย เพื่อนำมาเป็นข้อมูลในการศึกษาและอ้างอิง ผู้วิจัยได้ศึกษางานวิจัยที่มีความเกี่ยวข้อง ดังต่อไปนี้

พรนภา ชุมเชื้อ (พรนภา, 2562) ได้ศึกษาการใช้เทคนิคการจัดกลุ่มร่วมกับเทคนิคการจำแนกประเภทข้อมูล โดยประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล ประกอบด้วยเทคนิคการจัดกลุ่ม โดยใช้เทคนิค K-Means เพื่อจัดกลุ่มผู้เรียนและสร้างคลาสคำตอบเพื่อหาจำนวนกลุ่มที่เหมาะสม จากการทดลองจัดกลุ่มผู้เรียนสามารถจัดกลุ่มข้อมูลได้ 2 กลุ่ม คือ กลุ่มปกติ และกลุ่มเสี่ยง



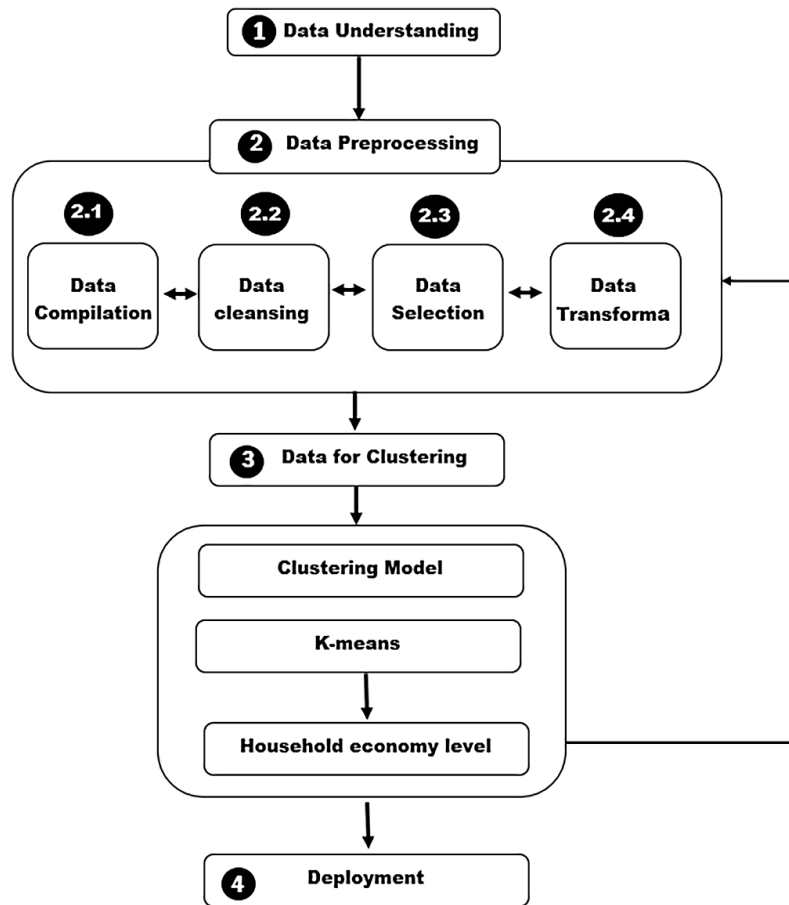
โดยใช้โปรแกรม Rapid Miner สำหรับจำแนกประเภทผู้เรียนโดยให้ค่าความถูกต้อง 98.30% ค่าความคลาดเคลื่อน RMSE เท่ากับ 0.069 โดยใช้ภาษา PHP ร่วมกับ WEKA จากผลการประเมินระดับความคิดเห็นของผู้ใช้งานมีค่าเฉลี่ย 4.34 ซึ่งอยู่ในเกณฑ์มาก

ปารณีย์ พึ่งวิชา อานนท์ ทับเที่ยง และธัญญา สัตยาภิธาน (ปารณีย์, อานนท์ และธัญญา, 2562) ได้ศึกษาการแบ่งกลุ่มพฤติกรรมของผู้บริโภคที่ซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์ ในเขตกรุงเทพมหานคร เก็บรวบรวมข้อมูล จากกลุ่มตัวอย่างจำนวน 400 คน ทำการวิเคราะห์แบ่งกลุ่มด้วยวิธี K-Means ซึ่งแบ่งกลุ่มผู้บริโภคเป็น 2 กลุ่ม ซึ่งมีลักษณะเฉพาะในแต่ละกลุ่ม คือ กลุ่มที่ 1 กลุ่มกระเป๋านักจ่ายได้ ส่วนใหญ่เป็นคนในช่วงอายุ 35 -45 ปี มีระดับการศึกษาสูงกว่าปริญญาตรี รายได้ต่อเดือนอยู่ในช่วง 30,001 - 45,000 บาท และ กลุ่มที่ 2 กลุ่มวัยสะสม ส่วนใหญ่เป็นคนในช่วงอายุ 25-34 ปี มีระดับการศึกษาปริญญาตรี รายได้ต่อเดือนอยู่ในช่วง 15,000 - 30,000 บาท มี ผลการวิจัยจะช่วยให้สามารถแบ่งตลาดออกเป็นกลุ่ม (Segment) เป็นประโยชน์ต่อการกำหนดตลาดเป้าหมายวางกลยุทธ์ทางการตลาด และการทำการตลาดออนไลน์

ปาริษา สุรัตน์เมธากุล และสายชล สินสมบุรณ์ทอง (ปาริษา และสายชล, 2564) ได้ศึกษาปัจจัยที่มีอิทธิพลต่อค่าใช้จ่ายของครัวเรือนในภาคกลางตามกลุ่มรายได้ โดยใช้ข้อมูลโครงการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน พ.ศ. 2562 ของสำนักงานสถิติแห่งชาติ ซึ่งมีจำนวนครัวเรือนในภาคกลางทั้งสิ้น 12,932 ครัวเรือน ใช้การวิเคราะห์การถดถอยพหุคูณเพื่อหาปัจจัยที่มีอิทธิพลต่อค่าใช้จ่ายของครัวเรือน การแบ่งครัวเรือนนั้นจะแบ่งตามเกณฑ์ควินไทล์ แบ่งครัวเรือนออกเป็น 5 กลุ่ม คือ กลุ่ม 1 ครัวเรือนที่มีรายได้น้อย กลุ่มที่ 2-4 ครัวเรือนที่มีรายได้ปานกลาง และกลุ่ม 5 ครัวเรือนที่มีรายได้สูง ผลการวิจัยพบว่าปัจจัยที่มีอิทธิพลต่อค่าใช้จ่ายของครัวเรือนทั้ง 3 กลุ่มรายได้คือ สมาชิกในครัวเรือน ผลตอบแทนทางอ้อม และสมาชิกครัวเรือนสูงอายุที่ได้รับสวัสดิการรัฐ

## วิธีการวิจัย

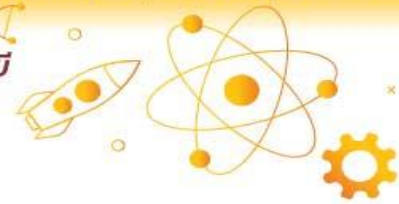
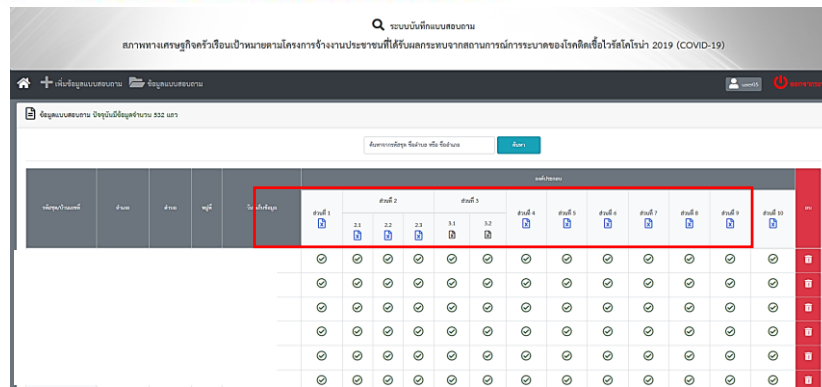
วิธีการดำเนินการวิจัย มีขั้นตอนการดำเนินงาน โดยการปรับปรุงตามกระบวนการ CRIPS-DM ได้ 4 ขั้นตอน ได้แก่ 1) การทำความเข้าใจข้อมูล 2) การเตรียมข้อมูลสำหรับการจัดกลุ่ม 3) การจัดกลุ่มข้อมูลด้วยเทคนิค K-means 4) การนำไปใช้งาน ดังภาพที่ 2 และมีรายละเอียดแต่ละขั้นตอนดังนี้



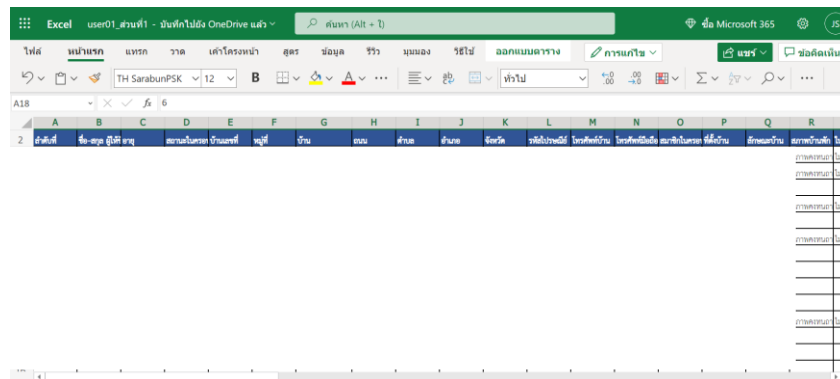
ภาพที่ 2 ขั้นตอนการดำเนินการวิจัย

### 1. การทำความเข้าใจข้อมูล (Data Understanding)

การเก็บรวบรวมข้อมูลเศรษฐกิจครัวเรือนได้มาจากข้อมูลประชากรจากภาคครัวเรือนเฉพาะครัวเรือน ของจังหวัด สกลนคร ซึ่งมี 12 หมู่บ้าน โดยช่วงเวลาที่ทำการเก็บรวบรวมข้อมูล คือ ปี พ.ศ. 2563 – 2564 และจากฐานข้อมูลตำบลและ ฐานข้อมูลขนาดใหญ่สำหรับการพัฒนาท้องถิ่นอย่างยั่งยืน (มหาวิทยาลัยราชภัฏสกลนคร, 2563) ประชากรที่ใช้ในการศึกษา ได้แก่ ครัวเรือนใน 12 หมู่บ้าน จำนวน 2,909 ครัวเรือน มีทั้งหมด 17,933 ระเบียบ ซึ่งมีการเก็บข้อมูลออกเป็น 10 ส่วน คือ 1) ข้อมูล ทั่วไปครัวเรือน 2) ทรัพย์สินของครัวเรือน 3) อาชีพและรายได้ของครัวเรือน 4) รายจ่ายของครัวเรือน 5) หนี้สินของครัวเรือน 6) ผลกระทบจากสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 7) การใช้เทคโนโลยีสารสนเทศ 8) การเข้าร่วมการละเล่น การฟ้อน การรำ พิธีกรรมตามวิถีวัฒนธรรมชุมชน 9) การเข้าร่วมโครงการที่ผ่านมาย้อนหลัง 3 ปี และ 10) ข้อคิดเห็นและ ข้อเสนอแนะ แสดงดังภาพที่ 3 ซึ่งข้อมูลแต่ละส่วนสามารถดาวน์โหลดออกมาในรูปแบบของไฟล์นามสกุล .xlsx ได้

ภาพที่ 3 ระบบฐานข้อมูลเศรษฐกิจครัวเรือน

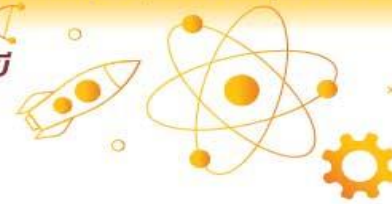


ภาพที่ 4 ระบบฐานข้อมูลเศรษฐกิจครัวเรือน

จากภาพที่ 4 เป็นการดาวน์โหลดข้อมูลออกมาจากระบบฐานข้อมูลเศรษฐกิจครัวเรือน อยู่ในรูปแบบไฟล์นามสกุล .xlsx ซึ่งข้อมูลยังไม่มี การนำมาวิเคราะห์ ผู้วิจัยจึงมีความสนใจที่จะนำข้อมูลชุดนี้มาทำการวิเคราะห์ เพื่อใช้ในการเตรียมข้อมูลสำหรับการจัดกลุ่ม

## 2. การเตรียมข้อมูลสำหรับการจัดกลุ่ม (Data Preprocessing)

จากข้อมูลในฐานข้อมูลที่ทำกรรวบรวมข้อมูลเศรษฐกิจครัวเรือน (มหาวิทยาลัยราชภัฏสกลนคร, 2563) มาจัดเก็บให้อยู่ในรูปแบบของไฟล์ .xlsx จากนั้นจะทำการทำความสะอาดข้อมูล ลดมิติข้อมูล และคัดเลือกข้อมูล จากข้อมูลทั้งหมด 17,933 ระเบียบ ปรับลดลงเหลือ จำนวนทั้งหมด 1,751 (ตารางที่ 2) ระเบียบ 16 ปัจจัย (ตารางที่ 1)



ตารางที่ 1 รายละเอียดของตัวแปรที่เป็นคุณลักษณะของกลุ่มตัวอย่างเศรษฐกิจครัวเรือน

ลำดับ	คุณลักษณะ	รายละเอียด	ชนิดข้อมูล
1	Education	วัยเรียน	Numeric
2	Working	วัยทำงาน	Numeric
3	Old	วัยสูงอายุ	Numeric
4	Occupation	อาชีพ	Numeric
5	AverageInY	รวมรายได้เฉลี่ย/ปี	Numeric
6	AssetValue	มูลค่าทรัพย์สิน	Numeric
7	AnimalHus	วัตถุประสงค์การเลี้ยงสัตว์	Numeric
8	Area	พื้นที่ก่อให้เกิดรายได้	Numeric
9	ProductCos	ต้นทุนการผลิตทำการเกษตร	Numeric
10	Product	ผลผลิตที่ได้จากการทำเกษตร	Numeric
11	TotalExpY	รวมค่าใช้จ่าย/ปี	Numeric
12	LoanB	หนี้ในระบบ	Numeric
13	LoanS	หนี้นอกระบบ	Numeric
14	TotalLia	รวมปริมาณหนี้สิน	Numeric
15	InternetUse	การใช้อินเทอร์เน็ตที่ก่อให้เกิดรายได้	Numeric
16	SalesCha	ช่องทางการขายสินค้าที่ก่อให้เกิดรายได้	Numeric

ตารางที่ 2 ข้อมูลเศรษฐกิจครัวเรือนที่ผ่านการทำความสะอาด

ลำดับ	Education	Working	Old	Occupation	AssetValue	AnimalHus	Area	ProductCos	Product	LoanB	LoanS	TotalExpY	TotalLia	AverageInY	InternetUse	SalesCha
1	0	3	0	1	1567650	0	1200	11400	2800	1	0	145560	120000	444000	1	1
2	0	2	0	1	299750	1	1600	11400	2800	1	0	93600	140000	484000	1	1
3	1	4	1	1	437650	0	800	2400	2100	1	0	241880	90000	84000	1	1
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1749	2	2	0	1	1115500	0	600	7600	0	1	0	224400	0	590000	1	0
1750	0	1	0	1	381500	1	20000	16100	35200	0	0	39000	60000	108000	1	0
1751	0	1	0	1	310500	0	300	3000	9200	1	0	56400	0	272000	1	0

จากตารางที่ 1 และ ตารางที่ 2 ปัจจัยที่ได้ จำนวน 16 ปัจจัย เกิดจากการคัดเลือกเฉพาะข้อมูลครัวเรือนที่สมบูรณ์ จากนั้นผู้วิจัยได้ทำการแปลงข้อมูล ปัจจัยที่มีลักษณะตัวเลขจำนวนมาก เพื่อให้อยู่ในรูปแบบที่เหมาะสมในการทดลอง โดยทำการปรับค่าคุณลักษณะของข้อมูลให้เป็นปกติแบบ 0-1 (0-1 Normalization) (นิภาพร, 2560) โดยใช้สมการที่ 2 ทำให้ได้ข้อมูลที่แสดงในตารางที่ 3 ซึ่งเป็นข้อมูลที่ได้ใช้ในการทดลองครั้งนี้



$$V_i = \frac{v_i - \text{MIN}_v}{\text{MAX}_v - \text{MIN}_v} \quad (2)$$

แทนค่าในสูตร โดยยกตัวอย่างข้อมูลในปัจจุบัน รวบรวมได้เฉลี่ย/ปี (AverageInY) แสดงวิธีการดังนี้

$$V_i = \frac{444000 - 20250}{947200 - 20250}$$

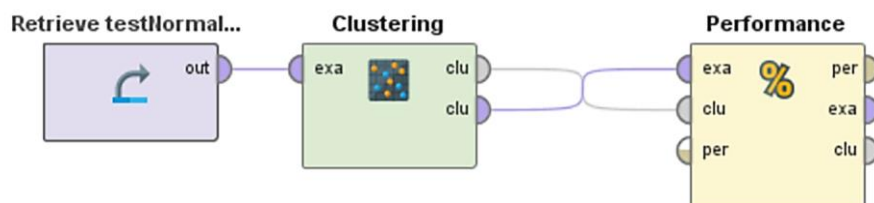
$$V_i = 0.457$$

ตารางที่ 3 ปัจจัยที่เลือกนำมาจัดกลุ่ม ในรูปแบบคุณลักษณะเป็น Numeric

ลำดับ	Education	Working	Old	Occupation	AssetValue	AnimalHus	Area	ProductCos	Product	LoanB	Loans	TotalExpY	TotalIia	AverageInY	InternetUse	SalesCha
1	0	3	0	1	0.129	0	0.010	0.070	0.001	1	0	0.027	0.017	0.457	1	1
2	0	2	0	1	0.025	1	0.020	0.070	0.001	1	0	0.017	0.019	0.500	1	1
3	1	4	1	1	0.036	0	0.010	0.020	0.001	1	0	0.045	0.013	0.069	1	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1749	2	2	0	1	0.092	0	0.040	0.000	0.000	1	0	0.014	0.024	0.494	1	0
1750	0	1	0	1	0.031	1	0.010	0.050	0.000	0	0	0.042	0.000	0.615	1	0
1751	0	1	0	1	0.026	0	0.220	0.100	0.015	1	0	0.007	0.008	0.095	1	0

### 3. การจัดกลุ่มข้อมูลด้วยเทคนิค K-means Clustering

ผู้วิจัยได้เลือกใช้เทคนิคการจัดกลุ่มด้วยเทคนิค K-means เนื่องจากเป็นวิธีการจัดกลุ่มที่ง่ายและเหมาะสมกับข้อมูลที่มีลักษณะเป็นกลุ่ม คือกำหนดจำนวนกลุ่มได้ชัดเจน (ถาวรีย์, อูร์รัฐ และจักริน, 2560) งานวิจัยนี้ได้ใช้โปรแกรม Rapid Miner เวอร์ชัน 9.10 (RapidMiner Inc., 2020) ในการจัดกลุ่มด้วยเทคนิค K-means โดยกำหนดค่าพารามิเตอร์ ทำการเลือก Add cluster attribute ทำการกำหนดจำนวนกลุ่มในการแบ่งโดยให้ k เท่ากับ 3 (ชฎารัตน์ และสิริธร, 2559)



ภาพที่ 5 ขั้นตอนการจัดกลุ่มโดยใช้ K-means



จากภาพที่ 5 แสดงให้เห็นถึงขั้นตอนการจัดกลุ่ม ภายในโอเปอเรเตอร์ Retrieve ใช้สำหรับดึงข้อมูลออกมาจาก Repositories โอเปอเรเตอร์ Clustering ใช้สำหรับแบ่งกลุ่มข้อมูลด้วยเทคนิค K-Means โดยมีการกำหนด (K=3) โอเปอเรเตอร์ Performance ใช้สำหรับคำนวณค่า Average within distance

#### 4. การนำไปใช้งาน

หลังจากทำการจัดกลุ่มข้อมูลและได้กลุ่มข้อมูลที่เหมาะสมแล้วทำการระบุระดับความเร็วในแต่ละกลุ่ม โดยพิจารณาจากระดับรายได้ 0-100,000 บาท อยู่ในระดับเศรษฐกิจน้อย ระดับรายได้ 100,000-400,000 บาท อยู่ในระดับเศรษฐกิจปานกลาง และ ระดับรายได้ มากกว่า 400,001 บาท อยู่ในระดับเศรษฐกิจสูง (สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ สำนักนายกรัฐมนตรี, 2559) สามารถนำข้อมูลระดับเศรษฐกิจเร็วเร็วที่ได้มาใช้ประโยชน์จริงในการจัดกลุ่มข้อมูลเศรษฐกิจเร็วเร็ว สำหรับสนับสนุนหรือเป็นข้อมูลประกอบการตัดสินใจในการวิจัยในลำดับต่อไป

### ผลการวิจัย

การจัดกลุ่มข้อมูลเศรษฐกิจเร็วเร็ว ด้วยเทคนิค K-means ข้อมูลจำนวน 1,751 ระเบียบ 16 ปัจจัย ผู้วิจัยได้ทำการจัดกลุ่มข้อมูลออกเป็น 3 กลุ่ม คือ กลุ่มระดับเศรษฐกิจน้อย กลุ่มระดับเศรษฐกิจปานกลาง กลุ่มระดับเศรษฐกิจสูง เพื่อจัดกลุ่มข้อมูลเศรษฐกิจเร็วเร็ว โดยมีการจัดกลุ่มเศรษฐกิจเร็วเร็ว ดังต่อไปนี้

#### ผลการจัดกลุ่มเศรษฐกิจเร็วเร็ว

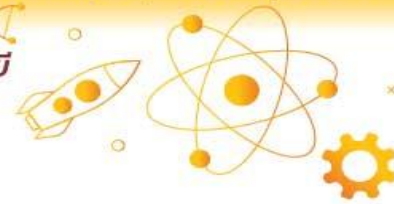
ผู้วิจัยได้ใช้ข้อมูลเศรษฐกิจเร็วเร็ว มาจัดกลุ่มระดับเศรษฐกิจ โดยใช้เทคนิค K-Means ในการจัดกลุ่มเศรษฐกิจเร็วเร็ว ได้ค่า K เท่ากับ 3 กลุ่ม

### PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -2.608
Avg. within centroid distance_cluster_0: -2.597
Avg. within centroid distance_cluster_1: -3.375
Avg. within centroid distance_cluster_2: -2.355
Davies Bouldin: -1.578
```

#### ภาพที่ 6 ค่า Centroid ของคลัสเตอร์

จากภาพที่ 6 ค่า Avg. within centroid distance คือ ค่าเฉลี่ยระยะห่างของทุกจุดจากจุดศูนย์กลางของคลัสเตอร์ ใช้เป็นเกณฑ์ที่ใช้ในการวัดคุณภาพการจัดกลุ่ม ดังนั้นจะแสดงลักษณะของข้อมูลที่มีค่าเฉลี่ยระยะห่างจาก Centroid กับข้อมูลแต่ละ cluster เท่ากับ -2.608 และ ค่า Avg. within centroid distance ของแต่ละกลุ่ม ได้ ค่า Avg. within centroid distance ของ Cluster 0 เท่ากับ -2.597 Cluster 1 เท่ากับ -3.375 และ Cluster 2 เท่ากับ -2.355



## Cluster Model

Cluster 0: 544 items

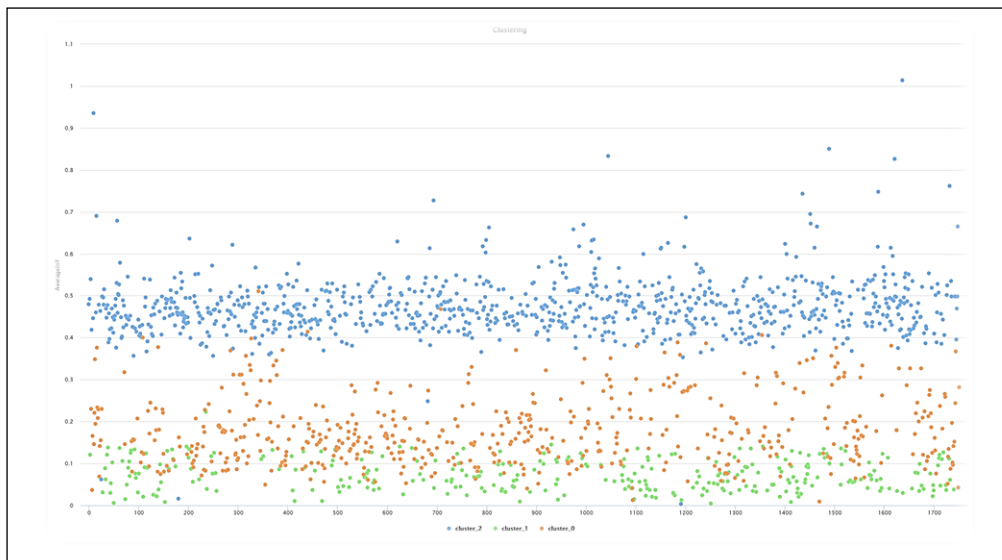
Cluster 1: 304 items

Cluster 2: 903 items

Total number of items: 1751

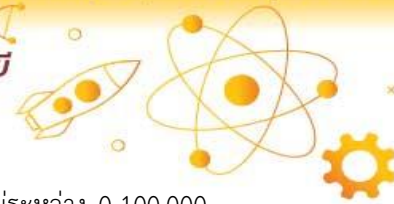
ภาพที่ 7 แสดงจำนวนข้อมูลเศรษฐกิจครัวเรือนที่จัดอยู่ในแต่ละกลุ่ม

จากภาพที่ 7 โมเดลการจัดกลุ่มของเทคนิค K-mean Clustering ที่ค่า K=3 นั้น สามารถระบุจำนวนครัวเรือนที่มีลักษณะเศรษฐกิจครัวเรือนคล้ายๆ กัน ให้อยู่ในกลุ่มเดียวกัน นั่นคือ Cluster 0 ได้ครัวเรือนที่มีลักษณะเหมือนกัน จำนวน 544 ครัวเรือน Cluster 1 ได้ครัวเรือนที่มีลักษณะเหมือนกัน จำนวน 304 ครัวเรือน และ Cluster 2 ได้ครัวเรือนที่มีลักษณะเหมือนกัน จำนวน 903 ครัวเรือน สามารถแสดงกราฟเปรียบเทียบได้ดังภาพที่ 8



ภาพที่ 8 กราฟแสดงผลการจัดกลุ่มระดับเศรษฐกิจครัวเรือน 3 กลุ่ม

จากภาพที่ 7 และ 8 สามารถอธิบายแปลผล ได้ดังนี้ กลุ่มที่ 1 จะแสดงด้วยจุดกลุ่มสีฟ้าที่ประกอบด้วยจำนวน 544 ครัวเรือน เป็นกลุ่มที่มีรายได้อยู่ระหว่าง 100,000-400,000 บาท (สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ สำนักนายกรัฐมนตรี, 2559) ลักษณะของครัวเรือนกลุ่มนี้ เป็นครัวเรือนที่มีรายได้เฉลี่ย/ปีของครัวเรือน ค่าใช้จ่ายอาหารและค่าใช้จ่ายเพื่อที่อยู่อาศัย หนี้สินของครัวเรือนอยู่ในระดับปานกลาง มีจำนวนสมาชิกในครัวเรือนมากกว่า 2 คน ครัวเรือนในกลุ่มนี้มีมูลค่าทรัพย์สินทางการเงินในช่วงปานกลาง ส่วนใหญ่ในครัวเรือนกลุ่มนี้จะมีอาชีพเกษตรกร ครัวเรือนกลุ่มนี้จะอยู่ในระดับเศรษฐกิจปานกลาง



กลุ่มที่ 2 จะแสดงด้วยจุดกลุ่มสีเขียวที่ประกอบด้วยจำนวน 304 คริวเรือน เป็นกลุ่มที่มีรายได้อยู่ระหว่าง 0-100,000 บาท (สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ สำนักนายกรัฐมนตรี, 2559) ลักษณะของคริวเรือนกลุ่มนี้เป็นคริวเรือนที่มีรายได้รายได้เฉลี่ย/ปีของคริวเรือน อยู่ในช่วงรายได้ต่ำ ค่าใช้จ่ายของคริวเรือนค่อนข้างมาก ในกลุ่มนี้จะเป็นกลุ่มที่มีหนี้สินของคริวเรือนค่อนข้างมาก มีจำนวนสมาชิกในคริวเรือนอยู่ใน 1-6 คน ส่วนใหญ่ในคริวเรือนกลุ่มนี้สมาชิกส่วนใหญ่ไม่ได้ทำงานรับค่าจ้างและเงินเดือน คริวเรือนกลุ่มนี้จะอยู่ในระดับเศรษฐกิจต่ำ

กลุ่มที่ 3 จะแสดงด้วยจุดกลุ่มสีส้ม ประกอบด้วยจำนวน 903 คริวเรือน เป็นกลุ่มที่มีรายได้ มากกว่า 400,001 บาท (สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ สำนักนายกรัฐมนตรี, 2559) ลักษณะของคริวเรือนกลุ่มนี้เป็นคริวเรือนที่มีรายได้รายได้เฉลี่ย/ปีของคริวเรือน ค่าใช้จ่าย/ปีของคริวเรือนสูงที่สุดเมื่อเปรียบเทียบกับทุกกลุ่ม ระดับการศึกษา คือปริญญาตรีและสูงกว่าปริญญาตรีมากกว่ากลุ่มอื่น คริวเรือนในกลุ่มนี้มีมูลค่าทรัพย์สินทางการเงินอยู่ในช่วงสูง ส่วนใหญ่ในคริวเรือนกลุ่มนี้จะมีอาชีพนอกภาคเกษตรกรรม คริวเรือนกลุ่มนี้จะอยู่ในระดับเศรษฐกิจสูง

เมื่อสามารถระบุว่าคริวเรือนแต่ละคริวเรือนจัดอยู่ในกลุ่มระดับเศรษฐกิจคริวเรือนได้แล้ว ผู้วิจัยได้ทำการระบุระดับเศรษฐกิจคริวเรือนให้แต่ละระเบียบ โดยใช้ชื่อปัจจัยนี้ว่า ระดับเศรษฐกิจ (Economic Level) ซึ่งมี 3 ระดับ คือ ระดับเศรษฐกิจต่ำ (Low Economic Level) ระดับเศรษฐกิจปานกลาง (Middle Economic Level) และ ระดับเศรษฐกิจสูง (High Economic Level) ดังแสดงในตารางที่ 4 คอลัมน์สุดท้าย

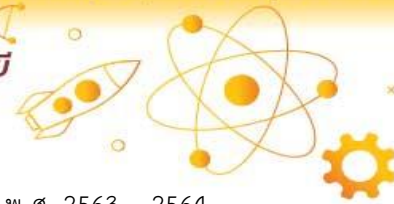
ตารางที่ 4 ข้อมูลเศรษฐกิจคริวเรือนที่ถูกระบุระดับเศรษฐกิจคริวเรือน

ลำดับ	Education	Working	Old	Occupation	AssetValue	AnimalHus	Area	ProductCos	Product	LoanB	LoanS	TotalExpY	TotalLia	AverageInY	InternetUse	SalesCha	Economic Level
1	0	3	0	1	0.129	0	0.010	0.00	0.001	1	0	0.02	0.01	0.45	1	1	High econ lv
2	0	2	0	1	0.025	1	0.020	0.00	0.001	1	0	0.01	0.019	0.500	1	1	High econ lv
3	1	4	1	1	0.036	0	0.010	0.020	0.001	1	0	0.045	0.013	0.069	1	1	Low econ lv
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
1749	2	2	0	1	0.092	0	0.040	0.000	0.000	1	0	0.014	0.024	0.494	1	0	High econ lv
1750	0	1	0	1	0.031	1	0.010	0.050	0.000	0	0	0.042	0.000	0.615	1	0	Middle econ lv
1751	0	1	0	1	0.026	0	0.220	0.100	0.015	1	0	0.007	0.008	0.095	1	0	Middle econ lv

จากตารางที่ 4 ทำการระบุข้อมูลแต่ละระเบียบให้อยู่ในกลุ่มระดับเศรษฐกิจคริวเรือน ซึ่งข้อมูลที่ได้สามารถนำไปประกอบการทำวิจัยหรือประกอบการตัดสินใจในการพัฒนาเศรษฐกิจคริวเรือนของแต่ละกลุ่มได้

### สรุปผลการวิจัย

งานวิจัยนี้ได้นำเสนอการจัดกลุ่มข้อมูลเศรษฐกิจคริวเรือน เพื่อทำการจัดกลุ่มข้อมูลเศรษฐกิจคริวเรือน และทำการเปรียบเทียบข้อมูลคริวเรือนในแต่ละกลุ่ม และพิจารณาจำนวนกลุ่มที่เหมาะสม สำหรับสนับสนุนหรือเป็นข้อมูลประกอบการตัดสินใจในการพัฒนาชุมชนท้องถิ่น และนำเสนอรายงานในรูปแบบข้อมูลภาพ โดยใช้ข้อมูลจากโครงการศาสตร์พระราชา ที่ทำการ



เก็บรวบรวมจากฐานข้อมูลตำบลและฐานข้อมูลขนาดใหญ่สำหรับการพัฒนาท้องถิ่นอย่างยั่งยืน ตั้งแต่ปี พ.ศ. 2563 – 2564 (มหาวิทยาลัยราชภัฏสกลนคร, 2563) ซึ่งมีการเก็บข้อมูล 10 ส่วน รวมทั้งหมด 178 ปัจจัย ได้มา 2,909 ครั้วเรือน คัดเลือกข้อมูลที่ส่งผลกระทบต่อเศรษฐกิจครัวเรือน ในการจัดกลุ่มข้อมูลปัจจัย 16 ปัจจัย จำนวน 1,751 ระเบียบ (ครั้วเรือน) คือ 1) วัยเรียน 2) วัยทำงาน 3) วัยสูงอายุ 4) อาชีพ 5) รวมรายได้เฉลี่ย/ปีของครั้วเรือน 6) มูลค่าทรัพย์สิน 7) วัตถุประสงค์การเลี้ยงสัตว์ 8) พื้นที่ก่อให้เกิดรายได้ 9) ต้นทุนการผลิต การทำการเกษตร 10) ผลผลิตที่ได้จากการทำการเกษตร 11) รวมค่าใช้จ่าย/ปีของครั้วเรือน 12) หนี้ในระบบ 13) หนี้นอกระบบ 14) รวมปริมาณหนี้สินของครั้วเรือน 15) การใช้อินเทอร์เน็ตที่ก่อให้เกิดรายได้ และ 16) ช่องทางการขายสินค้าที่ก่อให้เกิดรายได้ โดยข้อมูลเป็นตัวเลข เพื่อพิจารณาจำนวนกลุ่มที่เหมาะสม โดยใช้ K-Means ในการการจัดกลุ่มข้อมูลเศรษฐกิจครั้วเรือน จึงได้กำหนดค่า K เท่ากับ 3 (ชฎารัตน์ และสิริธร, 2559) เมื่อได้นำ K-Means มาวิเคราะห์การจัดกลุ่ม เพื่อทำการเปรียบเทียบในแต่ละกลุ่ม และพิจารณาจำนวนกลุ่มที่เหมาะสม

จากการวิจัยพบว่าการจัดกลุ่มเศรษฐกิจครั้วเรือน แสดงการวิเคราะห์ผล โดยทำการพิจารณาเลือกใช้การจัดกลุ่มเศรษฐกิจครั้วเรือน 3 กลุ่ม เพื่อแสดงการเปรียบเทียบข้อมูลครั้วเรือนในแต่ละกลุ่ม โดยจัดกลุ่มได้ทั้งหมด 3 กลุ่ม คือ ระดับเศรษฐกิจน้อย มีครั้วเรือนทั้งหมด จำนวน 544 ครั้วเรือน ระดับเศรษฐกิจปานกลาง มีครั้วเรือนทั้งหมด จำนวน 304 ครั้วเรือน และระดับเศรษฐกิจสูง มีครั้วเรือนทั้งหมด จำนวน 903 ครั้วเรือน

### ข้อเสนอแนะ

ข้อเสนอแนะการทำวิจัยครั้งต่อไป ควรมีการเปรียบเทียบประสิทธิภาพของจำนวน K ที่แตกต่างกัน และเปรียบเทียบการจัดกลุ่มข้อมูลด้วยเทคนิคแบบอื่น ๆ และควรนำข้อมูลมากกว่า 1 ปีขึ้นไป มาทำการวิเคราะห์

### กิตติกรรมประกาศ

ขอบคุณสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏสกลนคร อนุเคราะห์ข้อมูลจากฐานข้อมูลตำบลและฐานข้อมูลขนาดใหญ่สำหรับการพัฒนาท้องถิ่นอย่างยั่งยืน และขอบคุณทุนสนับสนุนการวิจัยและพัฒนาสำหรับนักศึกษา (ภาคปกติ) มหาวิทยาลัยราชภัฏสกลนคร จากงบประมาณเงินรายได้ ประจำปีงบประมาณ พ.ศ. 2566

### เอกสารอ้างอิง

- ชฎารัตน์ พิพัฒน์นันท์ และสิริธร เจริญรัตน์. (2559). การวิเคราะห์ข้อมูลทางเศรษฐกิจและสังคมของครั้วเรือนด้วยวิธีการแบ่งกลุ่มและหาความสัมพันธ์ สำหรับการทำให้เหมือนข้อมูล. *วารสารพัฒนาบริหารศาสตร์*, 56(3), 139-161.
- ชัตชัย แก้วตา และชนัญญาญจน์ แสงประสาน. (2564). การเลือกพืชสวนเศรษฐกิจเพื่อการเพาะปลูกที่เหมาะสม ในจังหวัดอุบลราชธานีโดยใช้เทคนิคการทำเหมือนข้อมูล. *วารสารวิทยาศาสตร์บูรพา*, 26(1), 593-609.
- ดารณี พิมพ์ช่างทอง. (2561). การวิเคราะห์จัดกลุ่มเพื่อการรณรงค์ทางการตลาดด้วยการใช้เครือข่ายสังคมออนไลน์. *วารสารวิชาการ RMUTT Global Business and Economics Review*, 13(1), 140-150.



- ถาวรีย์ สุขเกษม, อรุณีรัฐ สุขสวัสดิ์ชน, และจักริน สุขสวัสดิ์ชน. (2560). การวิเคราะห์สารอาหารเพื่อหลีกเลี่ยงโรคเบาหวานโดยใช้ขั้นตอนการแบ่งกลุ่มข้อมูล. *วารสารเทคโนโลยีสารสนเทศ*, 13(1), 1-9.
- นิภาพร ชนะมาร. (2560). การประยุกต์ใช้การเรียนรู้ของเครื่องสำหรับวางแผนการเรียนของนักศึกษาเพื่อเพิ่มผลสัมฤทธิ์ทางการเรียน. (วิทยานิพนธ์ปรัชญาดุษฎีบัณฑิต), มหาวิทยาลัยนเรศวร, พิษณุโลก.
- ปารณีย์ พึ่งวิชา, อานนท์ ทับเที่ยง, และธัญญา สัตยาภิธาน. (2562). ศึกษาการแบ่งกลุ่มพฤติกรรมของผู้บริโภคที่ซื้อเครื่องประดับผ่านเครือข่ายสังคมออนไลน์. *วารสารมนุษยศาสตร์และสังคมศาสตร์มหาวิทยาลัยราชพฤกษ์*, 7(5), 213-224.
- ปาริชา สุรัตน์เมธากุล และสายชล สนิสมบูรณ์ทอง. (2564). ปัจจัยที่มีอิทธิพลต่อค่าใช้จ่ายของครัวเรือนในภาคกลางตามกลุ่มรายได้ข้อมูล. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 10(2), 1-15.
- พรนภา ชุมเชื้อ. (2562). การใช้เทคนิคการจัดกลุ่มร่วมกับเทคนิคการจำแนกประเภท สำหรับการคัดกรองเด็กที่มีความบกพร่องทางการเรียนรู้จากบริบทพฤติกรรม. (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต), มหาวิทยาลัยทักษิณ, สงขลา.
- มหาวิทยาลัยราชภัฏสกลนคร. (2563). ระบบบันทึกแบบสอบถามสภาพทางเศรษฐกิจครัวเรือนเป้าหมายตามโครงการจ้างงานประชาชนที่ได้รับผลกระทบจากสถานการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19). ค้นจาก <http://qcovid.snru.ac.th/>.
- วนิษา แผลงรักษา และนิเวศ จิระวิชิตชัย. (2562). การแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์. *วารสารวิชาการชาชนนเทคโนโลยี มจร.ภูเก็ต*, 3(2), 1-10.
- สำนักงานคณะกรรมการพัฒนาการเศรษฐกิจและสังคมแห่งชาติ สำนักงานนายกรัฐมนตรี. (2559). *แผนพัฒนาเศรษฐกิจและสังคมแห่งชาติฉบับที่สิบสอง พ.ศ. 2560 – 2564*, ม.ป.ท.: สถาบัน
- สำราญ วานนท์ และรจนา เมืองแสน. (2563). การศึกษาและพัฒนาตัวแบบพยากรณ์คุณลักษณะความเหมาะสมสำหรับการเลือกสมัครสาขาวิชาเรียนโดยใช้เทคนิคเหมืองข้อมูล. *วารสารวิทยาการจัดการ*, 7(1), 135-151.
- Big Data. (2020). *ประเภทของการแบ่งกลุ่มข้อมูล (Clustering)*. ค้นจาก <https://bigdata.go.th/big-data-101/4-types-of-clustering/>.
- RapidMiner Inc. (2020). *RapidMiner Studio Developer v9.10.0 โปรแกรมวิเคราะห์ข้อมูลทำเหมืองข้อมูล*. ค้นจาก <https://iloadzone.com/home/rapidminer>.
- Thaidatascience. (2017). *รู้จัก Data Mining*. ค้นจาก <https://thaidatascience.wordpress.com/data-mining>.
- Thapanee, B. (2020). *เข้าใจ CRISP-DM ฉบับเร่งรัด*. ค้นจาก <https://kamboonchob.medium.com>.