

การวิเคราะห์ข้อมูลทางเศรษฐกิจและสังคม ของครัวเรือนด้วยวิธีการแบ่งกลุ่ม และหาความสัมพันธ์ สำหรับการทำให้มองข้อมูล

ชฎารัตน์ พิพัฒน์นันท์* และ สิริธร เจริญรัตน์**

บทคัดย่อ

งานวิจัยนี้เป็นการศึกษาและวิเคราะห์ข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน ประจำปี พ.ศ. 2552 โดยนำความรู้ทางด้านการทำเหมืองข้อมูลมาใช้ในการจัดกลุ่มข้อมูลตามการใช้จ่ายของประชากรโดยใช้อัลกอริทึมเคมีนสำหรับหากลุ่มของข้อมูลที่เหมาะสมจากค่า DB Index และ SD Validity Index พบว่า จำนวนกลุ่มที่เหมาะสม คือ 3 กลุ่ม จากนั้นใช้กฎความสัมพันธ์ในการหาความสัมพันธ์ของข้อมูลในแต่ละกลุ่ม เพื่อนำไปหาตัวแปรที่มีความสัมพันธ์กัน ผลการวิจัย พบว่า กฎความสัมพันธ์ที่ได้ในแต่ละกลุ่มมีลักษณะคล้ายคลึงกัน เช่น รายได้เฉลี่ยต่อเดือนของครัวเรือนมีความสัมพันธ์กับค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน ขนาดของครัวเรือนมีความสัมพันธ์กับจำนวนผู้หารายได้ และค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือนมีความสัมพันธ์กับจำนวนสมาชิกที่มีสิทธิในการเบิกค่ารักษาพยาบาล เป็นต้น

คำสำคัญ: อัลกอริทึมเคมีน กฎความสัมพันธ์

* คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยหอการค้าไทย
126/1 ถนนวิภาวดีรังสิต แขวงดินแดง เขตดินแดง กรุงเทพมหานคร 10400
เมลล์: chadarat_phi@utcc.ac.th

** คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยหอการค้าไทย
126/1 ถนนวิภาวดีรังสิต แขวงดินแดง เขตดินแดง กรุงเทพมหานคร 10400
เมลล์: sirithorn_jal@utcc.ac.th

Socio-Economic Household Data Analysis Using the Clustering and Association Technique for Data Mining

Chadarat Phipathananunth* and Sirithorn Jalearnrat**

Abstract

In this research, data from the Household Socio-Economic Survey 2009 of the Office of National Statistics, Thailand were studied and analyzed. A K-means algorithm was used to cluster the expenditure of the population. By using the DB Index and the SD Validity Index, it was found that the appropriate number of clusters was three. Then the association rule technique was used to determine the relationship between the variables. The results showed that the association rules were similar among clusters. For example, the average monthly household income was related to the average monthly household cost. Additionally, household size was related to the number of earners per household. Furthermore, the monthly tobacco cost for a household was associated with the number of members that were entitled to reimbursement for medical expenses.

Keywords: K-means Algorithm, Association Rule

-
- * School of Science and Technology, (UTCC)
126/1 Vibhavadi Rangsit Road, Din Daeng, Bangkok 10400, THAILAND.
E-mail: chadarat_phi@utcc.ac.th
- ** School of Science and Technology, (UTCC)
126/1 Vibhavadi Rangsit Road, Din Daeng, Bangkok 10400, THAILAND.
E-mail: sirithorn_jal@utcc.ac.th

บทนำ

การทำเหมืองข้อมูล (Data Mining) สามารถแบ่งอัลกอริทึมในการทำงานได้เป็น 2 ประเภทหลัก คือ การสร้างแบบจำลองแบบทำนาย (Predictive Model, Supervised Model) เน้นการจัดกลุ่มโดยอาศัยผลเฉลยที่มีอยู่ซึ่งประกอบด้วยการจำแนก (Classification) และการถดถอย (Regression) ส่วนประเภทที่สอง คือ การสร้างแบบจำลองเชิงพรรณนา (Descriptive Model, Unsupervised Model) ในที่นี้อาจเป็นการหาความสัมพันธ์ (Association) หรือการจัดกลุ่มข้อมูล (Clustering) หลักการทำเหมืองข้อมูลได้นำไปประยุกต์ใช้กับงานหลากหลายด้าน ได้แก่ ด้านธุรกิจ การแพทย์ การเงิน การธนาคาร การตลาด เป็นต้น

คณะผู้วิจัยได้เล็งเห็นประโยชน์จากการทำเหมืองข้อมูล จึงได้ทำวิจัยเรื่อง “การวิเคราะห์ข้อมูลทางเศรษฐกิจและสังคมของครัวเรือนด้วยวิธีการจัดกลุ่มและหาความสัมพันธ์ สำหรับการทำเหมืองข้อมูล” โดยนำข้อมูลทางเศรษฐกิจและสังคมของครัวเรือนจากศูนย์วิจัยมหาวิทยาลัยชิคาโก-มหาวิทยาลัยหอการค้าไทย (UC-UTCC Research Center) ที่ได้ทำการจัดเก็บข้อมูลภาวะเศรษฐกิจและสังคมของครัวเรือน (Household Socio-Economic Survey) โดยสำนักงานสถิติแห่งชาติ ประจำปี พ.ศ. 2552 จำนวน 39,633 รายการ ประกอบด้วยข้อมูลเกี่ยวกับรายได้ ค่าใช้จ่าย ภาวะหนี้สิน และทรัพย์สินของครัวเรือน ตลอดจนลักษณะที่อยู่อาศัย จากครัวเรือนตัวอย่างในทุกจังหวัดทั่วประเทศ ทั้งในเขตและนอกเขตเทศบาล มาวิเคราะห์เพื่อจัดกลุ่มข้อมูลภาวะเศรษฐกิจและสังคมของครัวเรือนโดยใช้อัลกอริทึมเคมีน (K-Means Algorithm) จัดกลุ่มข้อมูลที่มีลักษณะใกล้เคียงกันให้อยู่ในกลุ่มเดียวกัน จากนั้นใช้กฎความสัมพันธ์ (Association Rule) ในการหาความสัมพันธ์ของข้อมูลในแต่ละกลุ่ม เพื่อนำไปใช้สำหรับหาตัวแปรที่มีความสัมพันธ์กันต่อไป

ทฤษฎีที่เกี่ยวข้อง

การจัดกลุ่ม (Clustering Method)

การจัดกลุ่มข้อมูล คือ การรวมกลุ่มของข้อมูลที่มีลักษณะเหมือนกัน รูปแบบและแนวโน้มที่เหมือนกัน โดยเริ่มจากการหาตัวแทนของกลุ่ม จากนั้นทำการเปรียบเทียบข้อมูลกับตัวแทนของแต่ละกลุ่ม ถ้าข้อมูลคล้ายคลึงกับตัวแทนของกลุ่มไหนก็就会被จัดให้อยู่ในกลุ่มนั้น วิธีในการจัดกลุ่มข้อมูลแบ่งออกเป็น 2 ประเภทใหญ่ ๆ คือ (Jain *et al.*, 1999)

- 1) การจัดกลุ่มข้อมูลแบบแบ่งส่วน (Partitional Clustering) การจัดจำแนกข้อมูลออกเป็นกลุ่มย่อย ๆ ตามจำนวนกลุ่มที่กำหนดโดยพิจารณาจากความคล้ายคลึงกัน ข้อมูลที่อยู่ในกลุ่มเดียวกันจะมีลักษณะเหมือนกันหรือคล้ายกัน ส่วนข้อมูลที่อยู่ต่างกลุ่มกันจะมีลักษณะที่แตกต่างกัน เทคนิคการจัดกลุ่มข้อมูลแบบแบ่งส่วน ได้แก่ อัลกอริทึมเคมีน และฟัซซี่ซีมีน (Fuzzy C-Means)

- 2) การจัดกลุ่มข้อมูลแบบโครงสร้างลำดับขั้น (Hierarchical Clustering) การจัดกลุ่มออกเป็นกลุ่มย่อย ๆ ที่มีความสัมพันธ์กันระหว่างกลุ่มใหญ่และกลุ่มย่อย ๆ ลงไป วิธีการจัดกลุ่มข้อมูลแบบโครงสร้างลำดับขั้นมีอยู่ 2 วิธี คือ การจัดกลุ่มจากล่างขึ้นบน (Agglomerative, Bottom-up) และการจัดกลุ่มข้อมูลจากบนลงล่าง (Divisive, Top-down)

- อัลกอริทึมเคมีน (K-Means Algorithm)

การจัดกลุ่มข้อมูลอัลกอริทึมเคมีนนับเป็นการพัฒนาและนำเสนอโดย Mac Queen ในปี ค.ศ. 1967 ซึ่งแสดงให้เห็นถึงอัลกอริทึมในการจัดกลุ่มที่สมาชิกภายในกลุ่ม จะมีระยะใกล้จุดศูนย์กลางหรือตัวแทนของกลุ่ม (Mean) โดยวิธีการจัดกลุ่มข้อมูลอัลกอริทึมเคมีนนับจะประกอบด้วยกำหนัดจำนวนกลุ่มเริ่มต้น กำหนดตัวแทนกลุ่ม การจัดข้อมูลแต่ละตัวเข้ากลุ่ม และสุดท้าย คือ การปรับปรุงตัวแทนในแต่ละกลุ่ม ปัญหาที่พบ คือ การกำหนดจำนวนจุดเริ่มต้น ซึ่งส่งผลต่อประสิทธิภาพของการจัดกลุ่ม หรือการที่กลุ่มบางกลุ่มนั้นมีจำนวนสมาชิกน้อยเกินหรืออาจจะไม่มีสมาชิกในกลุ่มเลย จึงได้มีการกำหนดกฎเกณฑ์ว่ากลุ่มที่จะอยู่ในรอบถัดไปได้้นั้นจะต้องมีสมาชิกอยู่ไม่น้อยกว่าค่าคงที่ค่าหนึ่งที่กำหนดขึ้นมา ขั้นตอนวิธีการจัดกลุ่มข้อมูลอัลกอริทึมเคมีน (Tan *et al.*, 2006) มีขั้นตอน ดังนี้

- 1) กำหนดจำนวนกลุ่มที่ต้องการ โดยกำหนดให้ K เท่ากับจำนวนกลุ่มที่ต้องการจัดกลุ่ม และเลือกจุดศูนย์กลางของกลุ่มเริ่มต้น
- 2) คำนวณหาตัวแทนกลุ่ม หรือจุดศูนย์กลาง (Centroid) ในแต่ละกลุ่ม
- 3) จัดกลุ่มข้อมูลใหม่โดยพิจารณาจากค่าความใกล้ชิดหรือระยะห่างของข้อมูลในกลุ่มกับตัวแทนของกลุ่มต่าง ๆ ว่าข้อมูลนั้นสมควรที่จะอยู่กลุ่มใด
- 4) ทำการปรับปรุงการจัดกลุ่มโดยย้อนกลับไปทำข้อ 2 และจะหยุดเมื่อข้อมูลสมาชิกในกลุ่มแต่ละกลุ่มนั้นไม่มีการเปลี่ยนแปลงแล้ว

การจัดกลุ่มข้อมูลโดยอัลกอริทึมเคมีนต้องการพื้นที่ที่ใช้ในการทำงานน้อย เนื่องจากเก็บเฉพาะข้อมูล และจุดศูนย์กลางเท่านั้น ดังนั้น จึงต้องการพื้นที่เป็น $O((m + K)n)$ เมื่อ n คือ จำนวนข้อมูลทั้งหมด และ m คือ จำนวนของตัวแปรที่ใช้ ส่วนเวลาที่ใช้ในการประมวลผลจะเป็นสมการเชิงเส้นในรูปของจำนวนข้อมูล ดังนั้น เวลาที่ต้องการใช้พื้นที่เป็น $O(I * K * m * n)$ เมื่อ I เป็นจำนวนรอบที่ใช้เมื่อตัวแทนของกลุ่มมีการเปลี่ยนแปลง และ K คือ จำนวนกลุ่มที่ต้องการจัดกลุ่ม และในกรณีที่เรากำหนดค่าตัวแปรให้คงที่ เวลาที่ใช้จะเป็น $O(n)$ (Tan *et al.*, 2006)

Algorithm Basic K-Means algorithm.

- 1: Select K points as initial centroids.
- 2: **repeat**
- 3: From k clusters by assigning each point to its closest centroid.
- 4: Recomputed the centroid of each cluster.
- 5: **until** Centroids do not change.

ภาพที่ 1: แสดงขั้นตอนการทำงานของขั้นตอนวิธีการจัดกลุ่มข้อมูลอัลกอริทึมเคมีน

ที่มา: Tan *et al.* (2006)

- กฎความสัมพันธ์ (Association Rule)

กฎความสัมพันธ์เป็นเทคนิคหนึ่งของการทำเหมืองข้อมูล โดยหลักการทำงานของวิธีนี้คือการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ หรือจากการวิเคราะห์การซื้อสินค้าของลูกค้า เรียกว่า “การวิเคราะห์แบบตะกร้าสินค้า (Market Basket Analysis)” ซึ่งประเมินจากข้อมูลที่รวบรวมไว้ ผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหา ซึ่งการวิเคราะห์แบบนี้เป็นการใช้ “กฎความสัมพันธ์” เพื่อหาความสัมพันธ์ของข้อมูล

การหาความสัมพันธ์

$$I = \{i_1, i_2, \dots, i_n\}$$

$$D = \{t_1, t_2, \dots, t_m\}$$

เมื่อ I คือ เซตไอเท็ม ที่มีไอเท็มทั้งหมดเป็นสมาชิก ซึ่งในที่นี้อาจเป็นซื้อสินค้าหรือหน่วยพื้นฐานที่จะนำมาใช้

T คือ ทรานแซคชันที่เป็นเซตย่อยของเซตไอเท็ม

D คือ เซตข้อมูลที่มีทรานแซคชันทุกตัวเป็นสมาชิก

n คือ จำนวนไอเท็ม

m คือ จำนวนทรานแซคชัน

สามารถนิยามกฎความสัมพันธ์ได้ว่า $X \rightarrow Y$ เมื่อ $X \subset I$, $Y \subset I$ และ $X \cap Y = \emptyset$

นอกจากนี้ กฎความสัมพันธ์ทุกกฎจะประกอบด้วยค่าสนับสนุน (Support) และค่าความมั่นใจ (Confidence) ซึ่งมีนิยาม ดังนี้

ค่าสนับสนุนของ $X \rightarrow Y$ คือ สัดส่วนของจำนวนทรานแซคชันที่มี X และ Y ปรากฏอยู่ ต่อจำนวนทรานแซคชันทั้งหมดในชุดข้อมูล

$$\text{Support, } S(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (1)$$

ค่าความเชื่อมั่นของ $X \rightarrow Y$ คือ อัตราส่วนของจำนวนทรานแซคชันที่มีทั้ง X และ Y ต่อจำนวนทรานแซคชันที่มี X

$$\text{Confidence, } C(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2)$$

เมื่อ N คือ จำนวนทรานแซคชันทั้งหมด

$\sigma(X)$ คือ จำนวนครั้งที่เกิด X

ค่าสนับสนุนถูกนำมาใช้ เพื่อการกำจัดกฎความสัมพันธ์ที่ไม่น่าสนใจ กฎที่ประกอบด้วย ไอเท็มเซตซึ่งมีค่าสนับสนุนน้อย หมายถึง ไอเท็มเซตนั้น ๆ ไม่เป็นที่น่าสนใจ หรือเกิดในจำนวนน้อย ดังนั้น กฎความสัมพันธ์ที่มีไอเท็มเซตเหล่านั้นปรากฏอยู่ จึงไม่ถูกสนใจ ในขณะที่ค่าความเชื่อมั่นใช้ เพื่อแสดงถึงความน่าเชื่อถือของกฎ ยังมีค่าความเชื่อมั่นสูงนั้นหมายความว่าไอเท็มเซตนั้นเกิดขึ้น ร่วมกับไอเท็มเซตอีกกลุ่มหนึ่งมากขึ้นเท่านั้น

ค่าสหสัมพันธ์หรือหน่วยวัดความสัมพันธ์ (Correlation Analysis หรือ Lift) หมายถึง ค่าที่บ่งบอกถึงความสัมพันธ์ที่น่าสนใจระหว่างเหตุการณ์และผลที่ตามมา คำนวณได้ ดังนี้

$$corr_{A,B} = \frac{P(A \cup B)}{P(A)P(B)} = \frac{P(B|A)}{P(B)} \quad (3)$$

โดยที่ $P(B|A) = P(A \cup B)/P(A)$

A คือ เหตุการณ์ที่เกิดก่อน (Antecedent) หรือไอเท็มที่อยู่ทางซ้าย

B คือ สิ่งที่เกิดขึ้นตามมา (Consequence) หรือไอเท็มที่อยู่ทางขวา

$P(A)$ คือ จำนวนครั้งที่เกิดเหตุการณ์ A

$P(B)$ คือ จำนวนครั้งที่เกิดเหตุการณ์ B

ผลลัพธ์ที่ได้จากการวิเคราะห์ค่าสหสัมพันธ์ คือ

- ถ้าค่าสหสัมพันธ์มีค่าน้อยกว่า 1 หมายถึง การที่เกิดเหตุการณ์ A ไม่ได้ส่งเสริมให้เกิด B จริง
- ถ้าค่าสหสัมพันธ์มากกว่า 1 หมายถึง การที่เกิดเหตุการณ์ A ส่งเสริมให้เกิด B จริง

- ถ้าค่าสหสัมพันธ์มีเท่ากับ 1 หมายถึง การเกิดเหตุการณ์ A ไม่ได้มีความสัมพันธ์แต่อย่างใดกับการเกิด B ซึ่งถือว่าเป็นอิสระต่อกัน

ตัวอย่างและความหมายของกฎความสัมพันธ์

มูลค่าทรัพย์สินที่เป็นตัวเงินของครัวเรือน น้อยกว่า 10,001 บาท → ไม่มีสมาชิกที่มีสิทธิเบิกค่ารักษาพยาบาล ด้วยค่าสนับสนุนเท่ากับ ร้อยละ 30 ค่าความเชื่อมั่น ร้อยละ 93 หมายความว่า เมื่อครัวเรือนมีมูลค่าทรัพย์สินที่เป็นตัวเงินน้อยกว่า 10,001 บาท แล้วครัวเรือนจะไม่มีสมาชิกที่มีสิทธิเบิกค่ารักษาพยาบาล โดยโอกาสที่ครัวเรือนในกลุ่มที่ 1 นี้จะมีมูลค่าทรัพย์สินที่เป็นตัวเงินน้อยกว่า 10,001 บาท ร้อยละ 31 หากเมื่อครัวเรือนในกลุ่มที่ 1 มูลค่าทรัพย์สินที่เป็นตัวเงินน้อยกว่า 10,001 บาท ตามกฎความสัมพันธ์ที่ 1 นี้ก็จะไม่มีสมาชิกที่มีสิทธิเบิกค่ารักษาพยาบาลด้วยร้อยละ 93

งานวิจัยที่เกี่ยวข้อง

ไชยยศ เอื้อวิริยะไชยกุล และสมชาย ปราการเจริญ (2554) นำเทคนิคการทำเหมืองข้อมูลมาเพื่อศึกษาความสัมพันธ์ของนักกอล์ฟที่มาใช้บริการโดยใช้อัลกอริทึมออพริโอรีในการสร้างกฎความสัมพันธ์ จากนั้นนำไปวิเคราะห์การถดถอย (Regression Analysis) ซึ่งถูกใช้เป็นแนวทาง เพื่อการพยากรณ์การจัดสรรงบประมาณรายจ่ายประจำปี การจัดใช้ทรัพยากรต่าง ๆ ของสนามในปีต่อไป

วาทีณี นุ้ยเพียร และคณะ (2551) นำเสนอการเปรียบเทียบอัลกอริทึมการจัดกลุ่มข้อมูล โดยใช้ข้อมูลปริมาณโอโซนจากการตรวจวัดประจำวัน ซึ่งเป็นตัวแปรเชิงปริมาณแบบอัตราส่วน มีค่าต่อเนื่องและทำการคัดเลือกข้อมูล แบบ CfsSubsetEval จำนวน 6 แอททริบิวต์ และพิจารณาจากผู้วิจัยจำนวน 24 แอททริบิวต์ ผลปรากฏว่าอัลกอริทึมเคมีน และ XMeans เป็นแบบจำลองที่ให้ผลลัพธ์ดีที่สุด เมื่อเทียบกับจำนวนกลุ่มของข้อมูลจริง

สิริธร เจริญรัตน์ และชฎารัตน์ พิพัฒนนันท์ (2556) ใช้เทคนิคการทำเหมืองข้อมูล โดยใช้ อัลกอริทึมของต้นไม้การตัดสินใจ C4.5 เพื่อสร้างแบบจำลองการจำแนกข้อมูล ตามประเภทอุตสาหกรรม คือ 1) การเกษตรและประมง 2) การผลิต 3) การค้าและบริการ เพื่อหาปัจจัยที่มีผลต่อรายได้ของประชากร ผลการวิจัย พบว่า ปัจจัยที่มีผลต่อรายได้ของประชากรในกลุ่มที่ 1 และกลุ่มที่ 3 คือ อาชีพ กลุ่มที่ 2 คือ ประเภทการจ้างงาน

เสาวณีย์ ไทยรุ่งโรจน์ และวาทีตร รักษธรรม (2549) ได้นำเสนอการวิเคราะห์ และศึกษาพฤติกรรมการบริโภคของครัวเรือนไทย โดยใช้ข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน พ.ศ. 2547 ซึ่งจัดทำโดยสำนักงานสถิติแห่งชาติ พบว่า ครัวเรือนไทยมีค่าใช้จ่ายเพิ่มขึ้นเฉลี่ยร้อยละ 5.4 ต่อปี แต่พฤติกรรมการบริโภคเมื่อพิจารณาจากสัดส่วนค่าใช้จ่าย พบว่า ไม่มีการเปลี่ยนแปลงมาก ค่าใช้จ่ายด้านอาหารเครื่องดื่มยังคงมีสัดส่วนสูงสุด แต่ลดลงเมื่อเทียบกับปี พ.ศ. 2545 ขณะที่สัดส่วนค่าใช้จ่ายเกี่ยวกับยานพาหนะและค่าบริการสื่อสารมีความสำคัญเป็นลำดับ

ที่สอง นอกจากนั้น มีค่าใช้จ่ายการบริโภคเครื่องดื่มแอลกอฮอล์มีสัดส่วนเพิ่มขึ้น ส่วนค่าใช้จ่ายด้านการศึกษาและรักษาพยาบาลมีสัดส่วนลดลง

Barth, M., E. Dimitriadou, K. Hornik and E. Moser (2003) ศึกษาเปรียบเทียบวิธีการจัดกลุ่มข้อมูลโดยใช้ 8 วิธีการ ประกอบด้วย เคมีนัลกอริทึม, แผนผังการจัดระเบียบตนเอง (Self-Organizing Maps), Neural Gas, Hardel, C-Means, Maximin, CLARA และ Ufel โดยใช้ข้อมูลจากการจำลองและข้อมูลจริง การวัดประสิทธิภาพใช้ Weighted Jaccard (WJC) และ Correlation Coefficient (CC) ผลการทดลอง พบว่า เคมีนัลกอริทึม และ Neural Gas เป็นวิธีที่ดีที่สุด เนื่องจากให้ค่า WJC และ CC สูงสุด

Halkidi, M., Y. Batistakis and M. Vazirgiannis (2001) ศึกษากระบวนการในการจัดกลุ่ม หลักการพื้นฐานของการจัดกลุ่ม รวมถึงวิธีในการวัดประสิทธิภาพ โดยทำการเปรียบเทียบวิธีในการวัดประสิทธิภาพ เพื่อหาจำนวนกลุ่มที่เหมาะสม ประกอบด้วย DB Index, RMSSTD, RS และ SD Validity Index ใช้ข้อมูลทดสอบ 5 ชุด โดยมี 4 ชุดเป็นข้อมูลของข้อมูลมาตรฐาน ส่วนอีก 1 ชุดเป็นข้อมูลจริง จากผลการทดลอง พบว่า SD Validity Index สามารถให้ผลถูกต้องถึง 3 ใน 5 ของข้อมูลทดสอบ

Pakhira และคณะ (2004) นำเสนอวิธีในการวัดประสิทธิภาพที่ชื่อว่า PBM-Index โดยทำการเปรียบเทียบกับวิธี DB Index, XB-Index, Dunn's Index ใช้ข้อมูลทดสอบ 8 ชุดซึ่งเป็นข้อมูลมาตรฐานที่รู้จำนวนกลุ่มแล้ว จากผลการทดลอง พบว่า วิธีที่นำเสนอ คือ PBM-Index สามารถหาจำนวนกลุ่มที่เหมาะสมได้ถูกต้องทั้งหมด

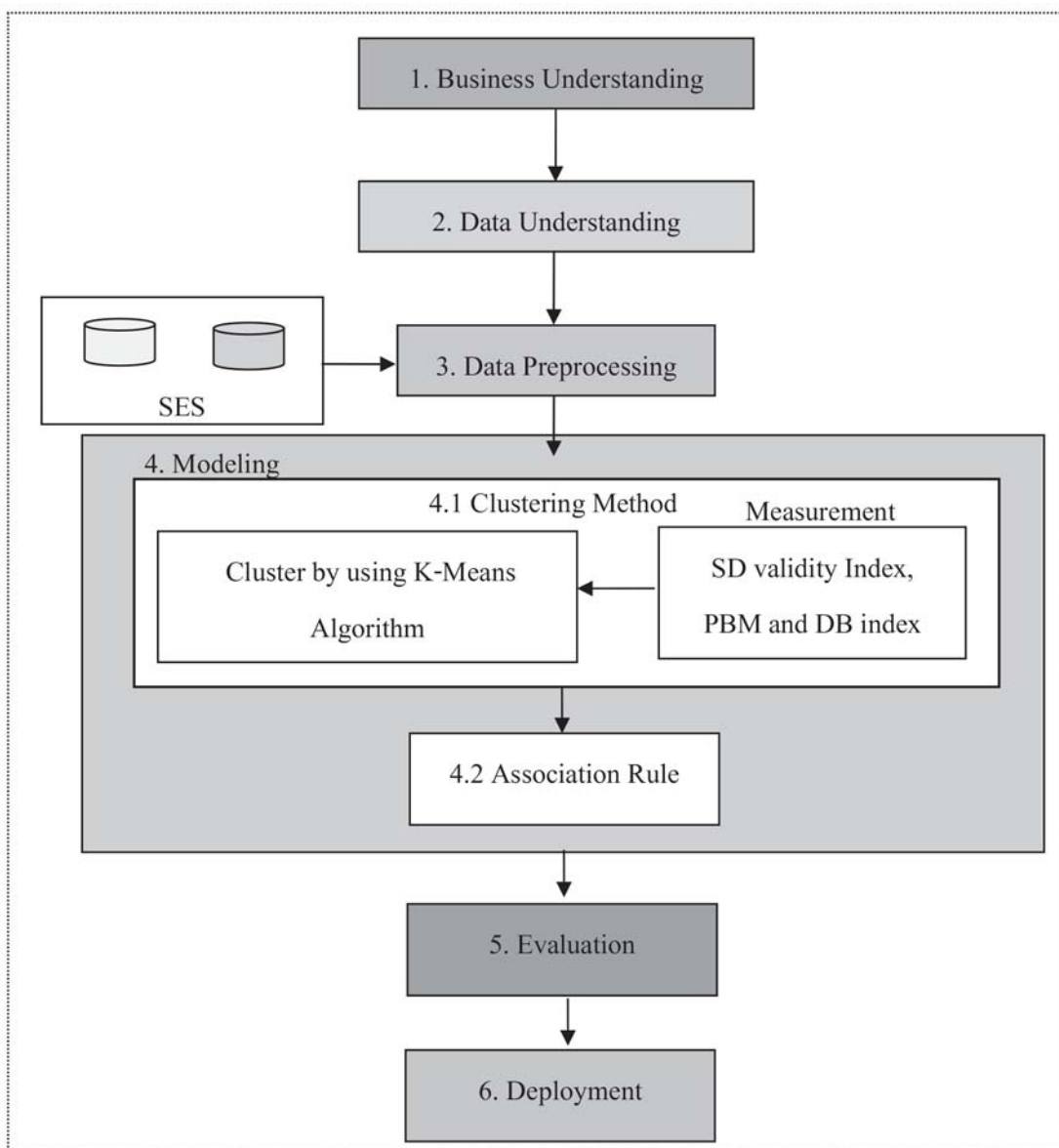
การวิเคราะห์ข้อมูลทางเศรษฐกิจและสังคมของครัวเรือนด้วยวิธีการแบ่งกลุ่มและหาความสัมพันธ์

- ข้อมูลสำหรับทำการวิจัย (Data set)

ข้อมูลสำหรับทำการวิจัยนี้ คือ ข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน ซึ่งเก็บรวบรวมข้อมูลเกี่ยวกับรายได้ ค่าใช้จ่าย ภาวะหนี้สิน และทรัพย์สินของครัวเรือน ตลอดจนลักษณะที่อยู่อาศัย โดยทำการเก็บรวบรวมข้อมูลทุกเดือน รายการจากครัวเรือนตัวอย่างในทุกจังหวัดทั้งในเขตและนอกเขตเทศบาล ปี พ.ศ. 2552 จำนวน 39,633 ครัวเรือน

- ขั้นตอนการทดลอง

ขั้นตอนการทดลองในงานวิจัยนี้ใช้เทคนิคของการทำเหมืองข้อมูลโดยใช้อัลกอริทึมเคมีนัลในการจัดกลุ่มข้อมูลที่มีลักษณะใกล้เคียงกันให้อยู่ในกลุ่มเดียวกัน จากนั้นจะใช้กฎความสัมพันธ์ในการหาความสัมพันธ์ของข้อมูลในแต่ละกลุ่ม เพื่อนำไปใช้ในการหาปัจจัยที่มีความสัมพันธ์กันต่อไป ดังแสดงในภาพที่ 2



ภาพที่ 2: แสดงภาพโครงสร้างการทำงานของงานวิจัย

1) การทำความเข้าใจธุรกิจ (Business Understanding)

ปัจจุบันนี้สภาพเศรษฐกิจของประเทศไทยอยู่ในสภาวะที่มีการแข่งขันสูง ซึ่งมีผลกระทบต่อชีวิตความเป็นอยู่ของประชากร การใช้จ่ายเป็นปัจจัยหนึ่งที่มีผลต่อการดำรงชีวิตของประชากร ในงานวิจัยนี้จะทำการรวบรวมข้อมูลเพื่อจัดกลุ่มข้อมูลครัวเรือนของประชากร และหาความสัมพันธ์ของข้อมูลเพื่อหาปัจจัยที่มีผลต่อการใช้จ่ายของประชากรในประเทศไทย

2) การทำความเข้าใจข้อมูล (Data Understanding)

คณะผู้วิจัยทำการศึกษาและทำความเข้าใจข้อมูลที่ได้ทำการจัดเก็บข้อมูลภาวะเศรษฐกิจและสังคมของครัวเรือน เช่น ชื่อตัวแปร ค่าที่จัดเก็บ ความหมาย และชนิดของข้อมูล เป็นต้น

3) การเตรียมข้อมูล (Data Preprocessing)

ขั้นตอนนี้จะทำการรวบรวมข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน ในระยะเวลา 1 ปี (พ.ศ. 2552) มาจัดเก็บลงฐานข้อมูล MySQL และทำการคัดเลือกข้อมูลที่ต้องการนำมาใช้ประโยชน์ จากนั้นจะทำการทำความสะอาดข้อมูล ทำการตัดข้อมูลที่ผิดปกติทิ้ง และแปลงข้อมูลให้อยู่ในรูปที่สามารถนำไปใช้ประมวลผลได้ ซึ่งในขั้นตอนนี้ต้องทำการแปลงข้อมูลที่จะนำไปใช้ให้อยู่ในรูปคะแนนมาตรฐาน (Standard Scores) เพื่อตรวจสอบค่าผิดปกติของข้อมูล (Outliers)

ค่าผิดปกติ คือ ข้อมูลที่มีค่าแตกต่างจากกลุ่ม เช่น ข้อมูลอายุของคนคือ 250 ปี เงินเดือนเฉลี่ยคือ 3,000,000 บาท ค่าใช้จ่ายต่อเดือนคือ 5,400,000 บาท เป็นต้น ซึ่งค่าผิดปกติมีโอกาสเกิดขึ้นได้ 2 ประการ คือ 1) การจดบันทึกหรือเก็บข้อมูลมีความคลาดเคลื่อน 2) กลุ่มตัวอย่างที่เก็บรวบรวมข้อมูล มีความแตกต่างไปจากกลุ่มอื่นจริง ซึ่งการเกิดค่าผิดปกติประการแรกนั้น สามารถเกิดขึ้นได้เสมอ จึงต้องทำการตรวจสอบข้อมูลให้ถูกต้องก่อนนำข้อมูลไปใช้ในขั้นตอนต่อไปโดยใช้คะแนนมาตรฐานซึ่งสามารถคำนวณได้ ดังนี้

$$Z_i = \frac{X_i - \bar{X}}{S.D.} \quad (4)$$

เมื่อ Z_i คือ คะแนนมาตรฐานของตัวแปร

X_i คือ ข้อมูลดิบ

$\bar{X}$ คือ ค่าเฉลี่ยของข้อมูล

$S.D.$ คือ ส่วนเบี่ยงเบนมาตรฐาน

หลังจากกำจัดข้อมูลที่มีค่าผิดปกติแล้ว ขั้นตอนที่ต่อไปจะทำการแปลงข้อมูลเพื่อเปลี่ยนค่าข้อมูลให้อยู่ในรูปมาตรฐาน (Data Normalization) เพื่อลดความเหลื่อมล้ำของข้อมูล และทำให้ข้อมูลอยู่ในช่วงเดียวกัน คณะผู้วิจัยเลือกใช้วิธี Min-Max Normalization ซึ่งเป็นเทคนิคที่ต้องรู้ค่าสูงสุดและค่าต่ำสุดของข้อมูล และจะทำให้ข้อมูลที่ได้นั้นอยู่ในช่วงใหม่ที่กำหนด โดยปกติแล้วจะกำหนดให้อยู่ในช่วง 0 และ 1

$$new_v = \frac{v - \min_x}{\max_x - \min_x} \cdot (new_max_x - new_min_x) + new_min_x \quad (5)$$

เมื่อ v คือ ค่าของข้อมูล

$\min_x$ คือ ค่าที่น้อยที่สุดของชุดข้อมูลเดิม

$\max_x$ คือ ค่าที่มากที่สุดของชุดข้อมูลเดิม
 new_min_x คือ ค่าที่น้อยที่สุดช่วงชุดข้อมูลใหม่
 new_max_x คือ ค่าที่มากที่สุดช่วงชุดข้อมูลใหม่

หลังจากทำความสะอาดและแปลงข้อมูลเรียบร้อยแล้วจะได้ข้อมูลที่นำมาใช้ในงานวิจัยนี้
 จำนวนทั้งหมด 39,633 รายการ

4) การพัฒนาตัวแบบ (Modeling)

- การจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมเคมีน

ขั้นตอนนี้จะทำการจัดกลุ่มข้อมูลโดยคณะผู้วิจัยได้เลือกใช้เทคนิคการจัดกลุ่มข้อมูล
 อัลกอริทึมเคมีนเข้ามาใช้ในการทดลอง เนื่องจากอัลกอริทึมเคมีนเหมาะสำหรับการจัดกลุ่มเมื่อมีข้อมูล
 จำนวนมาก เป็นอัลกอริทึมที่ง่ายต่อความเข้าใจ และใช้เวลาในการประมวลผลข้อมูลไม่มากนัก ขั้นตอนที่
 ในการทำงานจะหาจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่มตั้งแต่ 2-10 กลุ่ม และทำการหาจำนวนกลุ่ม
 ที่ดีที่สุดโดยใช้ค่า SD Validity Index, PBM-Index และ DB index เป็นตัวตัดสิน จากนั้นจึงนำ
 จำนวนกลุ่มที่ได้ไปเป็นข้อมูลนำเข้าเพื่อใช้ในการจัดกลุ่มข้อมูลต่อไป

- การหาความสัมพันธ์โดยใช้เทคนิคกฎความสัมพันธ์

ในขั้นตอนนี้จะนำข้อมูลที่ผ่านมาการจัดกลุ่มโดยอัลกอริทึมเคมีน มาทำการหาความสัมพันธ์
 โดยใช้เทคนิคกฎความสัมพันธ์ที่ชื่อว่าอัลกอริทึมเอพริโอรีในโปรแกรม Weka ซึ่งกฎความสัมพันธ์ที่ได้
 พิจารณาจากค่าสนับสนุน ค่าความมั่นใจ และค่าสหสัมพันธ์ จากนั้นนำข้อมูลที่ได้ไปวิเคราะห์หา
 ความสัมพันธ์ของการใช้จ่ายของประชากรต่อไป และก่อนที่จะนำข้อมูลเข้าสู่โปรแกรม Weka เพื่อใช้
 ในการสร้างกฎความสัมพันธ์

5) การแปลผลและการประเมินผล (Evaluation) เป็นการประเมินประสิทธิภาพของผลลัพธ์
 จากแบบจำลองวิเคราะห์ข้อมูลว่าครอบคลุม สามารถตอบเป้าหมายและวัตถุประสงค์ที่ตั้งไว้ใน
 ขั้นตอนแรกหรือไม่

6) การนำไปใช้ (Deployment) เป็นการนำผลลัพธ์ที่ได้จากการงานวิจัยไปใช้ประโยชน์
 ต่อไป

ตัวแปรที่ใช้ในงานวิจัย

การจัดกลุ่มและหาความสัมพันธ์ของข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน
 ใช้ตัวแปร ดังต่อไปนี้

ตารางที่ 1: ตัวแปรที่ใช้ในการจัดกลุ่มและหาความสัมพันธ์ของข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน

รายละเอียดของค่าในแต่ละตัวแปร	
ตัวแปรที่ 1 ภาค (Reg) แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ กรุงเทพมหานครและปริมณฑล กลุ่มที่ 2 คือ ภาคกลาง กลุ่มที่ 3 คือ ภาคเหนือ กลุ่มที่ 4 คือ ภาคใต้ กลุ่มที่ 5 คือ ภาคตะวันออกกึ่งยงเหนือ	ตัวแปรที่ 2 ระดับการศึกษาสูงสุดของหัวหน้าครัวเรือน (Education-Level) แบ่งออกเป็น 6 กลุ่ม กลุ่มที่ 1 คือ ต่ำกว่ามัธยมศึกษา กลุ่มที่ 2 คือ มัธยมศึกษา กลุ่มที่ 3 คือ อนุปริญญา กลุ่มที่ 4 คือ ปริญญาตรี กลุ่มที่ 5 คือ สูงกว่าปริญญาตรี กลุ่มที่ 6 คือ การศึกษาอื่น ๆ
ตัวแปรที่ 3 ขนาดของครัวเรือน (Household-Size) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ 1-2 คน กลุ่มที่ 2 คือ 3-4 คน กลุ่มที่ 3 คือ 5-7 คน กลุ่มที่ 4 คือ 8 คน ขึ้นไป	ตัวแปรที่ 4 จำนวนผู้ทำงานหารายได้ (No-Earners) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มี กลุ่มที่ 2 คือ 1 คน กลุ่มที่ 3 คือ 2-3 คน กลุ่มที่ 4 คือ 4 คน ขึ้นไป
ตัวแปรที่ 5 จำนวนสมาชิกที่ทำงานรับค่าจ้างและเงินเดือน (No-Wages) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มี กลุ่มที่ 2 คือ 1 คน กลุ่มที่ 3 คือ 2 คน กลุ่มที่ 4 คือ 3 คน ขึ้นไป	ตัวแปรที่ 6 จำนวนสมาชิกที่มีสิทธิเบิกค่ารักษาพยาบาล (Welfare) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มี กลุ่มที่ 2 คือ 1 คน กลุ่มที่ 3 คือ 2-3 คน กลุ่มที่ 4 คือ มากกว่า 3 คน ขึ้นไป
ตัวแปรที่ 7 จำนวนสมาชิกที่ได้รับบัตรประกันสุขภาพ (Universal-Health) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มี กลุ่มที่ 2 คือ 1 คน กลุ่มที่ 3 คือ 2-3 คน กลุ่มที่ 4 คือ มากกว่า 3 คนขึ้นไป	ตัวแปรที่ 8 จำนวนสมาชิกที่มีบัตรรับรองสิทธิการรักษาพยาบาล โดยใช้บัตรประกันสังคม (Medical-Card) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มี กลุ่มที่ 2 คือ 1 คน กลุ่มที่ 3 คือ 2-3 คน กลุ่มที่ 4 คือ มากกว่า 3 คนขึ้นไป

ตารางที่ 1: ตัวแปรที่ใช้ในการจัดกลุ่มและหาความสัมพันธ์ของข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน (ต่อ)

รายละเอียดของค่าในแต่ละตัวแปร	
ตัวแปรที่ 9 สถานภาพทางเศรษฐกิจสังคมของครัวเรือน (ชั้นทางเศรษฐกิจ) (Socio-Economic) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ ผู้ถือครองทำการเกษตร เช่น ปลุกพืช/เลี้ยงสัตว์/เพาะเลี้ยง ผู้ทำประมง ป่าไม้ ล่าสัตว์ เก็บของป่า บริการทางการเกษตร กลุ่มที่ 2 คือ ผู้ดำเนินธุรกิจของตนเองที่ไม่ใช่เกษตรกร กลุ่มที่ 3 คือ ลูกจ้าง เช่น ผู้ปฏิบัติงานวิชาชีพ นักวิชาการและนักบริหาร คนงานเกษตร คนงานทั่วไป เสมียน พนักงานขาย และให้บริการ ผู้ปฏิบัติงานในกระบวนการผลิต กลุ่มที่ 4 คือ ผู้ไม่ได้ปฏิบัติงานเชิงเศรษฐกิจ	ตัวแปรที่ 10 รายได้เฉลี่ยต่อเดือนของครัวเรือน (Income) แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ รายได้ต่ำกว่า 10,001 บาท กลุ่มที่ 2 คือ รายได้ 10,001-20,000 บาท กลุ่มที่ 3 คือ รายได้ 20,001-50,000 บาท กลุ่มที่ 4 คือ รายได้ 50,001-100,000 บาท กลุ่มที่ 5 คือ รายได้มากกว่า 100,000 บาท
ตัวแปรที่ 11 ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน (Month-Ex) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ ค่าใช้จ่ายต่ำกว่า 10,001 บาท กลุ่มที่ 2 คือ ค่าใช้จ่าย 10,001-20,000 บาท กลุ่มที่ 3 คือ ค่าใช้จ่าย 20,001-50,000 บาท กลุ่มที่ 4 คือ ค่าใช้จ่ายมากกว่า 50,000 บาท	ตัวแปรที่ 12 ค่าใช้จ่ายอาหารและเครื่องดื่มต่อเดือนของครัวเรือน (Food-Ex) แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ น้อยกว่า 2,001 บาท กลุ่มที่ 2 คือ 2,001-4,500 บาท กลุ่มที่ 3 คือ 4,501-7,000 บาท กลุ่มที่ 4 คือ 7,001-10,000 บาท กลุ่มที่ 5 คือ มากกว่า 10,000 บาท
ตัวแปรที่ 13 ค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือน (Tobacco-Ex) ประกอบด้วย ยาสูบ หมาก ยานัตถ์ และอื่น ๆ แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มีค่าใช้จ่าย กลุ่มที่ 2 คือ 1-200 บาท กลุ่มที่ 3 คือ 201-500 บาท กลุ่มที่ 4 คือ มากกว่า 500 บาท	ตัวแปรที่ 14 ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือน (Conex-Ex) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 คือ น้อยกว่า 5,001 บาท กลุ่มที่ 2 คือ 5,001-12,000 บาท กลุ่มที่ 3 คือ 12,001-24,000 บาท กลุ่มที่ 4 คือ มากกว่า 24,000 บาท
ตัวแปรที่ 15 ค่าใช้จ่ายเพื่อที่อยู่อาศัย เครื่องแต่งบ้านต่อเดือนของครัวเรือน (House-Ex) แบ่งออกเป็น 4 กลุ่ม คือ กลุ่มที่ 1 น้อยกว่า 2,001 บาท กลุ่มที่ 2 คือ 2,001-4,000 บาท กลุ่มที่ 3 คือ 4,001-8,000 บาท กลุ่มที่ 4 คือ มากกว่า 8,000 บาท	ตัวแปรที่ 16 ค่าใช้จ่ายเวชภัณฑ์และค่าตรวจรักษาพยาบาลต่อเดือนของครัวเรือน (Medical-Ex) แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มีค่าใช้จ่าย กลุ่มที่ 2 คือ 1-500 บาท กลุ่มที่ 3 คือ 501-1,000 บาท กลุ่มที่ 4 คือ 1,001-1,500 บาท กลุ่มที่ 5 คือ มากกว่า 1,500 บาท

ตารางที่ 1: ตัวแปรที่ใช้ในการจัดกลุ่มและหาความสัมพันธ์ของข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน (ต่อ)

รายละเอียดของค่าในแต่ละตัวแปร	
ตัวแปรที่ 17 ค่าใช้จ่ายการเดินทางและการสื่อสารต่อเดือนของครัวเรือน (Transport-Ex) แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มีค่าใช้จ่าย กลุ่มที่ 2 คือ 1-2,000 บาท กลุ่มที่ 3 คือ 2,001-4,000 บาท กลุ่มที่ 4 คือ 4,001-8,000 บาท กลุ่มที่ 5 คือ มากกว่า 8,000 บาท	ตัวแปรที่ 18 ค่าใช้จ่ายเรื่องการศึกษาต่อเดือนของครัวเรือน (Education-Ex) แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ ไม่มีค่าใช้จ่าย กลุ่มที่ 2 คือ 1-400 บาท กลุ่มที่ 3 คือ 401-800 บาท กลุ่มที่ 4 คือ 801-1,600 บาท กลุ่มที่ 5 คือ มากกว่า 1,600 บาท
ตัวแปรที่ 19 ค่าใช้จ่ายเบ็ดเตล็ด ต่อเดือนของครัวเรือน (Special-Ex) เป็นผลรวมของค่าใช้จ่ายเครื่องนุ่งห่มและรองเท้า ค่าใช้จ่ายส่วนบุคคล ค่าใช้จ่ายการบันเทิง การอ่านและกิจกรรมทางศาสนา และค่าใช้จ่ายการจัดงานพิธี ในโอกาสพิเศษ แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ น้อยกว่า 500 บาท กลุ่มที่ 2 คือ 501-1,000 บาท กลุ่มที่ 3 คือ 1,001-1,500 บาท กลุ่มที่ 4 คือ 1,501-2,000 บาท กลุ่มที่ 5 คือ มากกว่า 2,000 บาท	ตัวแปรที่ 20 ค่าใช้จ่ายที่ไม่เกี่ยวกับการอุปโภคบริโภคต่อเดือนของครัวเรือน (NonCon-Ex) ประกอบด้วยภาษี/ธรรมเนียม/ค่าปรับ ค่าสมาชิก กลุ่มอาชีพ เงิน/สิ่งของที่ส่งให้บุคคลนอกครัวเรือน บริจาคเงิน/สิ่งของให้แก่องค์กรต่าง ๆ เงินทำบุญ/เงินช่วยเหลืออื่น ๆ ค่าเบี้ยประกันภัย/ทรัพย์สิน/ประกันชีวิต/เงินฌาปนกิจศพ เงินสมทบประกันสังคม ค่าซื้อสลากกินแบ่ง/หวยของรัฐ/และการพนันอื่น ๆ ดอกเบี้ยจ่าย/ดอกเบี้ยแชร ค่าใช้จ่ายอื่น ๆ แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ มีค่าใช้จ่าย กลุ่มที่ 2 คือ 1-1,000 บาท กลุ่มที่ 3 คือ 1,001-3,000 บาท กลุ่มที่ 4 คือ 3,001-6,000 บาท กลุ่มที่ 5 คือ มากกว่า 6,000 บาท
ตัวแปรที่ 21 มูลค่าทรัพย์สินที่เป็นตัวเงินของครัวเรือน (Financial-Asset) แบ่งออกเป็น 5 กลุ่ม คือ กลุ่มที่ 1 คือ น้อยกว่า 10,001 บาท กลุ่มที่ 2 คือ 10,001-30,000 บาท กลุ่มที่ 3 คือ 30,001-50,000 บาท กลุ่มที่ 4 คือ 50,001-100,000 บาท กลุ่มที่ 5 คือ มากกว่า 100,000 บาทขึ้นไป	

* หมายเหตุ: - การแบ่งกลุ่มตัวแปรในงานวิจัยนี้ใช้หลักเกณฑ์ตามสำนักงานสถิติแห่งชาติ

- ตัวแปรที่ทางคณะผู้วิจัยนำข้อมูลของสำนักงานสถิติแห่งชาติมาปรับใช้ เพื่อให้เหมาะสมกับงานวิจัย โดยพิจารณาจากฮิสโตแกรม (Histograms) เพื่อหาการกระจายของข้อมูล ตัวแปรเหล่านี้ ได้แก่ อายุของหัวหน้าครัวเรือน รายได้เฉลี่ยต่อเดือนของครัวเรือน ค่าใช้จ่ายต่าง ๆ และมูลค่าทรัพย์สินที่เป็นตัวเงินของครัวเรือน

ผลการวิจัย

การจัดกลุ่มข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน ตามค่าใช้จ่ายของครัวเรือน ขั้นตอนนี้จะทำการจัดกลุ่มข้อมูลภาวะเศรษฐกิจและสังคมของครัวเรือน ตามค่าใช้จ่ายของครัวเรือนโดยใช้อัลกอริทึมเคมีน ซึ่งมีข้อมูลนำเข้าจำนวน 39,633 รายการ โดยแบ่งออกเป็น 2 ขั้นตอนย่อย คือ ขั้นตอนการหาจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่ม และขั้นตอนการจัดกลุ่มข้อมูล

ขั้นตอนที่ 1 การหาจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่ม

ตารางที่ 2: ผลการวัดประสิทธิภาพการจัดกลุ่มข้อมูลภาวะเศรษฐกิจและสังคมของครัวเรือน โดยใช้ อัลกอริทึมเคมีนตั้งแต่ 2-10 กลุ่ม

จำนวนกลุ่ม (Cluster)	SD Validity Index	PBM Index	DB Index
2	97.909	0.0037	1.636
3	84.464	0.0002	1.131
4	99.103	0.0020	1.656
5	93.014	0.0001	1.389
6	92.189	0.0002	1.188
7	94.357	0.0001	1.121
8	99.761	0.0000	1.149
9	145.625	0.0001	1.239
10	137.605	0.0001	1.270

จากตารางที่ 2 แสดงให้เห็นถึงค่า SD Validity Index ค่า PBM Index และค่า DB Index เมื่อทำการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมเคมีน คณะผู้วิจัยเลือกจำนวนกลุ่มที่เหมาะสมเท่ากับ 3 กลุ่ม เนื่องจากมีตัววัดประสิทธิภาพถึง 2 ใน 3 ตัวให้ผลดีเมื่อจำนวนกลุ่มเท่ากับ 3 กรณีที่มีค่า SD Validity และค่า DB Index มีค่าน้อย แสดงให้เห็นถึงการจัดกลุ่มที่ดี ในทางตรงกันข้าม ค่า PBM-Index ควรจะมีค่าสูง แสดงถึงการจัดกลุ่มที่ดี

ขั้นตอนที่ 2 การจัดกลุ่มข้อมูล

เมื่อหาจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่มได้เท่ากับ 3 กลุ่ม จากนั้นจึงทำการจัดกลุ่มข้อมูล โดยใช้ 8 ตัวแปร ซึ่งเป็นค่าใช้จ่ายหลักจากตารางที่ 1 มาใช้ในการจัดกลุ่ม ผลลัพธ์ที่ได้แสดงดังตารางที่ 3

ตารางที่ 3: แสดงรายละเอียดผลการจัดกลุ่ม โดยจำแนกตามตัวแปรหลัก 8 ตัวแปรที่ใช้ในการจัดกลุ่ม

ตัวแปรที่ใช้จัดกลุ่ม		จำนวนครัวเรือนในแต่ละกลุ่ม (ครัวเรือน(ร้อยละ))		
		กลุ่มที่ 1	กลุ่มที่ 2	กลุ่มที่ 3
		3,084(7.78)	27,923(70.45)	8,626(21.77)
ค่าใช้จ่ายอาหารและเครื่องดื่ม ต่อเดือนของครัวเรือน	น้อยกว่า 2,001 บาท	40(1.30)	3,110(11.14)	16(0.19)
	2,001-4,500 บาท	710(23.02)	14,856(53.20)	886(10.27)
	4,501-7,000 บาท	1,219(39.53)	8,160(29.22)	2,826(32.76)
	7,001-10,000 บาท	791(25.65)	1,699(6.08)	3,086(35.78)
	มากกว่า 10,000 บาท	324(10.51)	98(0.35)	1,812(21.01)
ค่าใช้จ่ายยาสูบต่อเดือน ของครัวเรือน	ไม่มี	0(0.00)	19,341(69.27)	7,257(84.13)
	1-200 บาท	0(0.00)	7,445(26.66)	852(9.88)
	201-500 บาท	1,127(36.54)	1,137(4.07)	488(5.66)
	มากกว่า 500 บาท	1,957(63.46)	0(0.00)	29(0.34)
ค่าใช้จ่ายที่ไม่เกี่ยวกับ การอุปโภคบริโภคต่อเดือน ของครัวเรือน	ไม่มี	19(0.62)	411(1.47)	9(0.10)
	1-1,000 บาท	1,370(44.42)	20,010(71.66)	1,570(18.20)
	1,001-3,000 บาท	1,082(35.08)	5,621(20.13)	3,030(35.13)
	3,001-6,000 บาท	431(13.98)	1,552(5.56)	2,219(25.72)
	มากกว่า 6,000 บาท	182(5.90)	329(1.18)	1,798(20.84)
ค่าใช้จ่ายเพื่อที่อยู่อาศัย เครื่องแต่งบ้านต่อเดือน ของครัวเรือน	น้อยกว่า 2,001 บาท	998(32.36)	15,546(55.67)	766(8.88)
	2,001-4,000 บาท	1,471(47.70)	10,158(36.38)	3,659(42.42)
	4,001-8,000 บาท	556(18.03)	2,007(7.19)	3,357(38.92)
	มากกว่า 8,000 บาท	59(1.91)	212(0.76)	844(9.78)
ใช้จ่ายเวชภัณฑ์ และค่าตรวจรักษาพยาบาล ต่อเดือนของครัวเรือน	ไม่มีค่าใช้จ่าย	1,207(39.14)	13,177(47.19)	3,427(39.73)
	1-500 บาท	1,554(50.39)	12,774(45.75)	3,595(41.68)
	501-1,000 บาท	197(6.39)	1,113(3.99)	715(8.29)
	1,001-1,500 บาท	53(1.72)	382(1.37)	297(3.44)
	มากกว่า 1,500 บาท	73(2.37)	477(1.71)	592(6.86)
ค่าใช้จ่ายการเดินทาง และการสื่อสารต่อเดือน ของครัวเรือน	ไม่มี	24(0.78)	1,612(5.77)	1(0.01)
	1-2,000 บาท	1,296(42.02)	19,748(70.72)	818(9.48)
	2,001-4,000 บาท	907(29.41)	4,911(17.59)	1,808(20.96)
	4,001-8,000 บาท	520(16.86)	1,404(5.03)	2,726(31.60)
	มากกว่า 8,000 บาท	337(10.93)	248(0.89)	3,273(37.94)

ตารางที่ 3: แสดงรายละเอียดผลการจัดกลุ่ม โดยจำแนกตามตัวแปรหลัก 8 ตัวแปรที่ใช้ในการจัดกลุ่ม (ต่อ)

ตัวแปรที่ใช้จัดกลุ่ม		จำนวนครัวเรือนในแต่ละกลุ่ม (ครัวเรือน(ร้อยละ))		
		กลุ่มที่ 1	กลุ่มที่ 2	กลุ่มที่ 3
		3,084(7.78)	27,923(70.45)	8,626(21.77)
ค่าใช้จ่ายเรื่องการศึกษา	ไม่มีค่าใช้จ่าย	1,537(49.84)	15,734(56.35)	3,437(39.84)
ต่อเดือนของครัวเรือน	1-400 บาท	1,064(34.50)	9,871(35.35)	2,092(24.25)
	401-800 บาท	248(8.04)	1,358(4.86)	1,051(12.18)
	801-1,600 บาท	164(5.32)	757(2.71)	1,071(12.42)
	มากกว่า 1,600 บาท	71(2.30)	203(0.73)	975(11.30)
ค่าใช้จ่ายเบ็ดเตล็ดต่อเดือนของครัวเรือน	น้อยกว่า 500 บาท	580(18.81)	11,858(42.47)	364(4.22)
	501-1,000 บาท	870(28.21)	8,323(29.81)	1,068(12.38)
	1,001-1,500 บาท	604(19.58)	3,887(13.92)	1,321(15.31)
	1,501-2,000 บาท	380(12.32)	1,939(6.94)	1,284(14.89)
	มากกว่า 2,000 บาท	650(21.08)	1,916(6.86)	4,589(53.20)

ผลการวิเคราะห์ความสัมพันธ์ของตัวแปรในครัวเรือนทั้ง 3 กลุ่ม

ในขั้นตอนการหาความสัมพันธ์ของตัวแปรในครัวเรือน ได้ทำการกำหนดค่าสนับสนุนขั้นต่ำ ค่าความเชื่อมั่นขั้นต่ำในแต่ละกลุ่ม และพิจารณาเฉพาะหน่วยวัดความสัมพันธ์ที่มีค่ามากกว่า 1 ขึ้นไป ผลการวิจัย พบว่า กลุ่มที่ 1 ได้กฎความสัมพันธ์ทั้งสิ้นจำนวน 66 กฎ กลุ่มที่ 2 ได้กฎความสัมพันธ์ทั้งสิ้นจำนวน 60 กฎ และกลุ่มที่ 3 ได้กฎความสัมพันธ์ทั้งสิ้นจำนวน 61 กฎ โดยได้นำเสนอกฎความสัมพันธ์ที่น่าสนใจดังแสดงในตารางที่ 4 ตารางที่ 5 และตารางที่ 6

ตารางที่ 4: ตัวอย่างกฎความสัมพันธ์ของตัวแปรในครัวเรือนกลุ่มที่ 1 (กำหนดค่าสนับสนุนขั้นต่ำเท่ากับร้อยละ 30 และค่าความเชื่อมั่นขั้นต่ำเท่ากับร้อยละ 60)

กฎ (Rule)	ค่าสนับสนุน (Support) (ร้อยละ)	ค่าความเชื่อมั่น (Confidence) (ร้อยละ)
1 มูลค่าทรัพย์สินที่เป็นตัวเงินของครัวเรือนน้อยกว่า 10,001 บาท → ไม่มีสมาชิกที่มีสิทธิเบิกค่ารักษาพยาบาล	31	93
2 ค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือน 201-500 บาทต่อเดือน → ไม่มีสมาชิกที่มีสิทธิเบิกค่ารักษาพยาบาล	32	87
3 รายได้เฉลี่ยต่อเดือนของครัวเรือน คือ 10,001-20,000 บาทต่อเดือน → ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 10,001-20,000 บาทต่อเดือน	33	79
4 ขนาดของครัวเรือน 5-7 คน → จำนวนผู้ทำงานหารรายได้ คือ 2-3 คน	35	79
5 ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือน 5,001-12,000 บาทต่อเดือน → ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน	37	75
6 ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือน 5,001-12,000 บาทต่อเดือน → จำนวนผู้ทำงานหารรายได้ คือ 2-3 คน	34	69
7 ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 10,001-20,000 บาทต่อเดือน → จำนวนผู้ทำงานหารรายได้ คือ 2-3 คน	37	67
8 ค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือนมากกว่า 500 บาทต่อเดือน → ไม่มีสมาชิกที่มีบัตรรับรองสิทธิการรักษาพยาบาล	42	66
9 ค่าใช้จ่ายเพื่อที่อยู่อาศัย เครื่องแต่งบ้าน 2,001-4,000 บาทต่อเดือน → ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือน 5,001-12,000 บาทต่อเดือน	30	64
10 ค่าใช้จ่ายเพื่อที่อยู่อาศัย เครื่องแต่งบ้าน 2,001-4,000 บาทต่อเดือน → ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 10,001-20,000 บาทต่อเดือน	30	63

ตารางที่ 5: ตัวอย่างกฎความสัมพันธ์ของตัวแปรในครัวเรือนกลุ่มที่ 2 (กำหนดค่าสนับสนุน ขั้นต่ำเท่ากับ ร้อยละ 50 และค่าความเชื่อมั่นขั้นต่ำเท่ากับร้อยละ 60)

	กฎ (Rule)	ค่าสนับสนุน (Support) (ร้อยละ)	ค่าความเชื่อมั่น (Confidence) (ร้อยละ)
1	ไม่มีจำนวนสมาชิกที่ทำงานรับค่าจ้างและเงินเดือน → ไม่มีสมาชิกที่มีบัตรรับรองสิทธิการรักษาพยาบาล	53	99
2	ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือนต่ำกว่า 10,001 บาทต่อเดือน → ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือนน้อยกว่า 5,001 บาทต่อเดือน	53	93
3	ค่าใช้จ่ายที่ไม่เกี่ยวกับการอุปโภคบริโภค 1-1,000 บาทต่อเดือน → ไม่มีสมาชิกที่มีบัตรรับรองสิทธิการรักษาพยาบาล	65	91
4	ระดับการศึกษาสูงสุดต่ำกว่ามัธยมศึกษา → ไม่มีสมาชิกที่มีสิทธิเบิกค่ารักษาพยาบาล	61	91
5	ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือนน้อยกว่า 5,001 บาทต่อเดือน → ค่าใช้จ่ายการเดินทางและการสื่อสาร 1-2,000 บาทต่อเดือน	55	86
6	ค่าใช้จ่ายการเดินทางและการสื่อสาร 1-2,000 บาทต่อเดือน → ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือนน้อยกว่า 5,001 บาทต่อเดือน	55	77
7	ค่าใช้จ่ายการเดินทางและการสื่อสาร 1-2,000 บาทต่อเดือน → ระดับการศึกษาสูงสุดต่ำกว่ามัธยมศึกษา	50	71

ตารางที่ 6: ตัวอย่างกฎความสัมพันธ์ของตัวแปรในครัวเรือนกลุ่มที่ 3 (กำหนดค่าสนับสนุน ขั้นต่ำเท่ากับร้อยละ 35 และค่าความเชื่อมั่นขั้นต่ำเท่ากับร้อยละ 45)

กฎ (Rule)	ค่าสนับสนุน (Support) (ร้อยละ)	ค่าความเชื่อมั่น (Confidence) (ร้อยละ)
1 ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือน 12,001-24,000 บาทต่อเดือน → ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน	51	96
2 ค่าใช้จ่ายการเดินทางและการสื่อสารมากกว่า 8,000 บาทต่อเดือน → ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน	36	94
3 ค่าใช้จ่ายเพื่อที่อยู่อาศัย เครื่องแต่งบ้าน 4,001-8,000 บาทต่อเดือน → ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน	35	91
4 ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน และค่าใช้จ่ายเบ็ดเตล็ดมากกว่า 2,000 บาทต่อเดือน → ไม่มีค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือน	36	86
5 ขนาดของครัวเรือน 3-4 คน → ไม่มีค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือน	42	86
6 รายได้เฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน → ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน	56	84
7 ไม่มีค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือน และค่าใช้จ่ายเบ็ดเตล็ดมากกว่า 2,000 บาทต่อเดือน → ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน	36	81
8 ขนาดของครัวเรือน 3-4 คน → จำนวนผู้ทำงานหารรายได้ 2-3 คน	36	74
9 ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือน 12,001-24,000 บาทต่อเดือน → รายได้เฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน	38	71
10 ค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน 20,001-50,000 บาทต่อเดือน → ค่าใช้จ่ายการเดินทางและการสื่อสารมากกว่า 8,000 บาทต่อเดือน	36	45

สรุปผลการวิจัย

งานวิจัยนี้เป็นการศึกษาและวิเคราะห์ข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน โดยนำความรู้ทางด้านการทำเหมืองข้อมูล มาใช้เพื่อจัดกลุ่มข้อมูลตามการใช้จ่ายของประชากร และหาความสัมพันธ์ของข้อมูลเพื่อศึกษาปัจจัยที่มีผลต่อการใช้จ่ายเงินของประชากรในประเทศไทย

สำหรับงานวิจัยนี้ขั้นตอนแรกนั้น ได้นำข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือนมาทำการจัดกลุ่มโดยใช้อัลกอริทึมเคมีน จัดกลุ่มตั้งแต่ 2-10 กลุ่ม เพื่อหาจำนวนกลุ่มที่เหมาะสม ซึ่งพบว่า เมื่อกำหนดจำนวนกลุ่มเท่ากับ 3 จะให้ค่า DB Index และ SD Validity Index ดีที่สุด ตัวแปรค่าใช้จ่ายที่นำมาใช้ในการจัดกลุ่มประกอบด้วย 8 ตัวแปร คือ 1) ค่าใช้จ่ายอาหารและเครื่องดื่มต่อเดือนของครัวเรือน 2) ค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือน 3) ค่าใช้จ่ายที่ไม่เกี่ยวกับการอุปโภคบริโภค 4) ค่าใช้จ่ายเพื่อที่อยู่อาศัย เครื่องแต่งบ้านต่อเดือนของครัวเรือน 5) ค่าใช้จ่ายเวชภัณฑ์และค่าตรวจรักษาพยาบาล 6) ค่าใช้จ่ายการเดินทางและการสื่อสาร 7) ค่าใช้จ่ายเรื่องการศึกษา และ 8) ค่าใช้จ่ายเบ็ดเตล็ด

ลักษณะของครัวเรือนในกลุ่มที่ 1 เป็นครัวเรือนที่มีรายได้เฉลี่ยต่อเดือนของครัวเรือนอยู่ในช่วงปานกลาง และมีค่าใช้จ่ายอยู่ในช่วงปานกลางด้วย ไม่ว่าจะเป็นค่าใช้จ่ายอาหารและเครื่องดื่ม ค่าใช้จ่ายเพื่อที่อยู่อาศัย เครื่องแต่งบ้านต่อเดือนของครัวเรือน แต่มีค่าใช้จ่ายเกี่ยวกับยาสูบต่อเดือนของครัวเรือนอยู่ในระดับที่สูง มีจำนวนสมาชิกครัวเรือนอยู่ในช่วง 3-4 คน ครัวเรือนในกลุ่มนี้ส่วนใหญ่มีมูลค่าทรัพย์สินทางการเงินอยู่ในช่วงน้อย สามารถหากฎความสัมพันธ์ได้ทั้งสิ้น 66 กฎ โดยกำหนดค่าสนับสนุนร้อยละ 30 และค่าความเชื่อมั่นร้อยละ 60

ลักษณะของครัวเรือนในกลุ่มที่ 2 ซึ่งมีจำนวนมากที่สุด คิดเป็นร้อยละ 70.45 เป็นครัวเรือนที่มีรายได้เฉลี่ยต่อเดือน และค่าใช้จ่ายของครัวเรือนน้อยที่สุด ขนาดของครัวเรือนอยู่ในช่วง 1-2 คน มากที่สุด และสมาชิกส่วนใหญ่ไม่ได้ทำงานรับค่าจ้างและเงินเดือน สามารถหากฎความสัมพันธ์ได้ทั้งสิ้น 60 กฎ โดยกำหนดค่าสนับสนุนร้อยละ 50 และค่าความเชื่อมั่นร้อยละ 60

ลักษณะของครัวเรือนในกลุ่มที่ 3 เป็นครัวเรือนที่มีรายได้เฉลี่ยต่อเดือน และค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือนสูงที่สุดเมื่อเปรียบเทียบกับทุกกลุ่ม คือ มีค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือนอยู่ในช่วง 20,001-50,000 บาท หัวหน้าครัวเรือนมีระดับการศึกษาสูงกว่ากลุ่มอื่น คือ ระดับปริญญาตรีและสูงกว่าปริญญาตรีมากกว่ากลุ่มอื่น สามารถหากฎความสัมพันธ์ได้ทั้งสิ้น 61 กฎ โดยกำหนดค่าสนับสนุนร้อยละ 35 และค่าความเชื่อมั่นร้อยละ 45

จากการวิจัย พบว่า กฎความสัมพันธ์ที่ได้มีลักษณะคล้ายคลึงกัน เช่น ค่าใช้จ่ายเพื่อการอุปโภคบริโภคต่อเดือนของครัวเรือนมีความสัมพันธ์กับค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน รายได้เฉลี่ยต่อเดือนของครัวเรือนมีความสัมพันธ์กับค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน ขนาดของครัวเรือน

มีความสัมพันธ์กับจำนวนผู้หารายได้ และค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือนมีความสัมพันธ์กับจำนวนสมาชิกที่มีสิทธิในการเบิกค่ารักษาพยาบาล เป็นต้น

ข้อเสนอแนะ

ผู้วิจัยคาดว่าหากสามารถประยุกต์โดยนำเทคนิคหรือวิธีอื่นมาใช้ร่วมกับอัลกอริทึมเดิมนี้จะทำให้การจัดกลุ่มข้อมูลมีประสิทธิภาพดีขึ้น นอกจากนี้ หากสามารถนำข้อมูลมากกว่า 1 ปีขึ้นไปมาวิเคราะห์ อาจทำให้ทราบถึงความสัมพันธ์ของตัวแปรที่มีผลต่อค่าใช้จ่ายประเภทต่าง ๆ ของครัวเรือนในประเทศไทยชัดเจนยิ่งขึ้น ในทางด้านเศรษฐศาสตร์ผู้ที่เกี่ยวข้องสามารถนำผลการวิจัยไปประยุกต์ใช้ในด้านเศรษฐศาสตร์จุลภาค (Microeconomics) ซึ่งเป็นการศึกษาพฤติกรรมของหน่วยทางเศรษฐกิจหน่วยย่อย เช่น ภาครัฐกิจ ครัวเรือน ผู้ผลิต ผู้บริโภค และเจ้าของปัจจัยการผลิต ในการตัดสินใจเกี่ยวกับปัญหาทางเศรษฐกิจต่าง ๆ เช่น ศึกษาว่าผู้บริโภคแต่ละคนมีพฤติกรรมในการบริโภคสินค้าและบริการอย่างไรเพื่อให้บรรลุความพอใจสูงสุดภายใต้รายได้ที่มีอยู่อย่างจำกัด

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณมหาวิทยาลัยหอการค้าไทยที่สนับสนุนทุนวิจัย และผู้มีส่วนเกี่ยวข้องทุกท่านที่ทำให้งานวิจัยนี้สำเร็จลุล่วงด้วยดี

เอกสารอ้างอิง

- ไชยยศ เอื้อวิริยะไชยกุล และสมชาย ปราการเจริญ. (2554). *การศึกษาพฤติกรรมการใช้บริการสนามของนักกอล์ฟ โดยใช้อัลกอริทึมของกฎความสัมพันธ์ กรณีศึกษาสนามกอล์ฟคริสตัลเบย์ กอล์ฟคลับ*. การประชุมทางวิชาการด้านเทคโนโลยีคอมพิวเตอร์และระบบสารสนเทศ ประยุกต์ระดับชาติครั้งที่ 3.
- เสาวณีย์ ไทยรุ่งโรจน์ และวาทีตร รัชชธรรม. (2549). การวิเคราะห์พฤติกรรมการบริโภคมวลรวมและฟังก์ชันการบริโภคของครัวเรือนไทย. *วารสารวิชาการมหาวิทยาลัยหอการค้าไทย*, 26(2): 39-70.
- สิริธร เจริญรัตน์ และชฎารัตน์ พืพฒบณันท์. (2556). การศึกษาปัจจัยที่มีผลต่อรายได้ของประชากรในประเทศไทยโดยใช้เทคนิคการทำเหมืองข้อมูล. *วารสารวิชาการมหาวิทยาลัยหอการค้าไทย*, 33(1): 132-152.
- วาทีณี น้อยเพียร และคณะ. (2551). *การเปรียบเทียบอัลกอริทึมการจัดกลุ่มข้อมูล Ozone day โดยใช้เทคนิคการทำเหมืองข้อมูล*. The 5th National Conference on Computing and Information Technology: 911-916.

- Barth, M., E. Dimitriadou, K. Hornik and E. Moser. (2003). *Comparison of Clustering Methods in fMRI Analysis by Ranking Association Coefficients*. Scientific Meeting of the International Society of Magnetic Resonance in Medicine 11. ed.
- Halkidi, M., Y. Batistakis and M. Vazirgiannis. (2001). *Clustering Algorithms and Validity Measures*. 3-22. International Conference on Scientific and Statistical Database Management 13. ed. IEEE Computer Society, Washington, DC, USA.
- Jain, A.K. and R.C. Dubes. (1988). *Algorithms for Clustering Data*. Jain, A.K. and Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Jain, A.K., M.N. Murty and P.J. Flynn. (1999). *Data Clustering: a Review*. 164-323. ACM Computing Surveys 31(3).
- Kovács, F., C. Legány and A. Babos. (2005). *Cluster Validity Measurement Techniques*. International Symposium of Hungarian Researches on Computational Intelligence Hungarian Researches on Computational Intelligence 6 ed., Hungary.
- MacQueen, J.B. (1967). *Some Methods for Classification and Analysis of Multivariate Observations*. 281-297. Berkeley Symposium on Mathematical Statistics and Probability 5. ed. Berkeley, University of California.
- Pakhira, M.K., Bandyopadhyay, S. and Maulik, U. (2004). *Validity Index for Crisp and Fuzzy Clusters*. 487-501. Pattern Recognition 37.
- Tan, P.N., M. Steinbach and V. Kumar. (2006). *Introduction to Data Mining*. Pearson Education, Inc., USA.

Translated Thai References

- Chaiyos Aurviriyachaikul, and Somchai Prakanchareon. (2011). *A Study on Course Services of Golfer Behaviors by Using Association Rules and Regression Analysis Case Study: Crystal Bay Golf Club* .The 3rd Conferences of Applied Computer Technology and Information System.
- Sauwanee Thairongroj, and Watit Ruktham. (2006). *An Analysis of Thai Households Aggregate Consumption and the Relevant Behavioral Functions*. *University of the Thai Chamber of Commerce Journal*, 26(2): 39-70.

- Sirithorn Jalearnrat, and Chadarat Phipathananunth. (2013). The study of Factors Affecting the Income of the Population in Thailand by Using Data Mining Techniques. *University of the Thai Chamber of Commerce Journal*, 33 (1): 132-152.
- Vatinee Nuiapiern, et al. (2008). *Ozone day Comparison Clustering with Data Mining Technique*. The 5th National Conference on Computing and Information Technology: 911-916.