

# การวิเคราะห์และทำนายพฤติกรรมการทำประกันชีวิต

## The Analysis and Forecasting of Life Insurance Behavior

จันทร์พร ดาวลอย<sup>1</sup> และ เทพฤทธิ์ บัณฑิตวัฒนวงศ์

<sup>1,2</sup>หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ

มหาวิทยาลัยศรีปทุม 61 ถ.พหลโยธิน จตุจักร กรุงเทพฯ 10900

Email: <sup>1</sup>aom\_kidkid@hotmail.com

### บทคัดย่อ

งานวิจัยนี้นำเสนอกลุ่มแบบจำลองการวิเคราะห์และทำนายพฤติกรรมการทำประกันชีวิตด้วยเทคนิคการทำเหมืองข้อมูล โดยแบบจำลองแรกสร้างด้วยวิธีการจัดหมวดหมู่ประเภทข้อมูลโดยใช้เทคนิคต้นไม้ตัดสินใจเพื่อวิเคราะห์หาปัจจัยที่มีผลต่อการเลือกประเภทกรมธรรม์ที่ถูกค่าต้องการทำประกันซึ่งได้แก่ อายุ เพศ ภูมิภาคของผู้ทำประกันชีวิต อาชีพ และประเภทของประกันภัย ปัจจัยเหล่านี้ถูกนำมาเข้าสู่กระบวนการวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอนด้วยหลักการ K-means Clustering เพื่อค้นหาพฤติกรรมของผู้ทำประกันภัยสำหรับใช้เป็นแบบจำลองที่สอง ผลการวิจัยโดยใช้ข้อมูลจริงแสดงให้เห็นว่าแบบจำลองแรกสามารถจัดหมวดหมู่การทำประกันภัยของผู้บริโภคได้ค่าความถูกต้องสูงสุด 60.40% ส่วนแบบจำลองที่สองสามารถแบ่งกลุ่มผู้ทำประกันภัยที่ดีที่สุด 2 กลุ่มได้แก่ กลุ่มผลิตภัณฑ์ประกันภัยอุบัติเหตุส่วนบุคคล ซึ่งเป็นผู้ชาย อายุระหว่าง 41-60 ปี ไม่ทราบอาชีพ อาศัยอยู่ภาคกลาง กลุ่มนี้มีขนาด 69% และกลุ่มผลิตภัณฑ์ประกันภัยสำหรับเจ้าบ้าน ซึ่งเป็นผู้หญิง อายุระหว่าง 41-60 ปี ไม่ทราบอาชีพ อาศัยอยู่ภาคตะวันออกเฉียงเหนือ กลุ่มนี้มีขนาด 31%

**คำสำคัญ**— เหมืองข้อมูล, การจัดหมวดหมู่, ต้นไม้ตัดสินใจ, การแบ่งกลุ่ม

### ABSTRACT

This research presents a set of models to analyze and predict the behavior of life insurance with data mining techniques. The first model has been created by means of decision tree-based classification to analyze factors affecting the selection of insurance policy desired by customers such as age, sex, region and occupation. The factors were brought into a nonhierarchical cluster analysis process based on K-means clustering technique to find out the behavior groups of the customers

serving as the second model. The results based on real data sets have showed that the first model can classify the insurance consumers with the highest accuracy of 60.40% while the second model can best cluster the customers into two groups: the first group is 69% of size contains personal accident insurance of 41 to 60 years-old males with unidentified occupations living in central region and the other group with the size of 31% is residential insurance held by 41 to 60 years-old females with unidentified occupations living in northeast region.

**Key words** – Data mining, Classification, Decision Tree, K-means Clustering

### 1. บทนำ

ในปัจจุบันนี้อัตราการเกิดอุบัติเหตุในประเทศไทยมีมากขึ้นทั้งในวงเทศกาลวันหยุดต่างๆและวันทำงานทั่วไป ทำให้เกิดการสูญเสียอวัยวะต่างๆ ของร่างกาย สูญเสียชีวิตและทรัพย์สิน ซึ่งสาเหตุหลักเกิดจากความประมาทในการดำเนินชีวิตประจำวัน การทำประกันภัยจึงได้รับความสนใจเพิ่มมากขึ้นจากข้อมูลสถิติของสำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย(คปภ.) [1] การทำประกันภัยประเภทสามัญที่ประกอบไปด้วย การทำประกันภัยอุบัติเหตุ ประกันภัยสุขภาพ และประกันภัยอื่นๆ ของปี 2555 มีตัวเลขโดยรวมเพิ่มขึ้นจากปี 2554 และมีแนวโน้มที่จะเพิ่มมากขึ้นเรื่อยๆ อันเนื่องมาจากปัจจัยหลายๆอย่างประกอบกันเช่น ปัญหาด้านสุขภาพ ปัญหาของประชากรในประเทศไทยที่ยังมีรายได้น้อย ปัญหาจากเศรษฐกิจและการเมืองในประเทศไทย ปัญหาจากภัยธรรมชาติที่รุนแรงเพิ่มมากขึ้น ฯลฯ ทำให้บุคคลเหล่านั้นหันมามุ่งเน้นการหาหลักประกันภัย เพื่อสร้างความมั่นคงให้กับชีวิต ทรัพย์สินของคนในครอบครัวและให้กับบริษัท ซึ่งปัจจุบันนี้มีบริษัทที่รับทำประกันภัยในประเทศไทยหลากหลายบริษัททั้งในและต่างประเทศ ในแต่ละบริษัทจะมี

เงื่อนไขการคุ้มครองหรือผลตอบแทนที่ดึงดูดใจลูกค้าที่ต้องการทำประกันภัยที่หลากหลายด้วยเช่นกัน

การนำเทคโนโลยีสารสนเทศมาช่วยในการวิเคราะห์ข้อมูลที่มีมากมายมหาศาล เพื่อหาปัจจัยที่ทำให้เห็นหันมาสนใจการทำประกันภัยแต่ละประเภท ว่ามีแนวโน้มเป็นไปในรูปแบบไหน จะทำให้บริษัทที่รับทำประกันภัยนั้นได้รับประโยชน์จากรายได้และส่วนแบ่งทางการตลาดของธุรกิจการทำประกันภัยที่เพิ่มขึ้นกว่าบริษัทคู่แข่ง สามารถคาดการณ์สถานการณ์ และสามารถวางแผนการตลาดได้อย่างระมัดระวังและมีข้อมูลไปใช้ในการแข่งขัน ดังนั้นผู้วิจัยจึงได้มีแนวคิดที่จะศึกษาเทคโนโลยีสารสนเทศ เพื่อนำมาประยุกต์ใช้กับสถานะและสถานการณ์ปัจจุบัน เพื่อตอบสนองความต้องการของฝ่ายการตลาดให้มากที่สุด โดยอาศัยการพยากรณ์ด้วยเทคนิคการทำเหมืองข้อมูล (data mining) มาเป็นเครื่องมือในการช่วยวิเคราะห์ เพื่อให้ได้ผลลัพธ์ที่น่าเชื่อถือและแม่นยำในการพยากรณ์

## 2. ทฤษฎี เทคโนโลยี และงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้ได้นำการประยุกต์ใช้เทคนิคเหมืองข้อมูล เพื่อพัฒนาสร้างตัวแบบการวิเคราะห์ความสัมพันธ์ของปัจจัยต่างๆกับพฤติกรรมการทำประกันชีวิตของผู้บริโภค โดยมีการเลือกใช้เทคนิคต้นไม้ตัดสินใจ และเลือกใช้การวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน โดยหลักการของ K-means Clustering [2] มาช่วยในการจัดกลุ่มข้อมูลของลูกค้าที่ต้องการ ที่มีลักษณะเดียวกันไว้ด้วยกัน และเลือกแบบที่มีประสิทธิภาพดีที่สุด

การทำเหมืองข้อมูล หมายถึง กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น โดยใช้หลักการหรือขั้นตอนการจำแนกประเภทและการทำนาย (Classification and Prediction) ซึ่งการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในปัจจุบันกับงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม ขั้นตอนการทำเหมืองข้อมูล มีดังนี้คือ

1. การทำข้อมูลให้สมบูรณ์ (Data Cleaning) เป็นขั้นตอนสำหรับการเติมค่าของข้อมูลหรือการคัดข้อมูลที่ ไม่เกี่ยวข้องออกไปวิธีการจัดการค่าที่ขาดหาย [3]

1.1 คัดรายการที่มีข้อมูลสูญหายทิ้ง นิยมใช้กับการทำเหมืองข้อมูลแบบจำแนกประเภท (Classification) ในกรณีที่ค่า แอททริบิวต์ (Attribute) ขาดหายไปเป็นจำนวนมาก

1.2 เติมค่าที่ขาดหายด้วยมือ วิธีนี้ไม่เหมาะสมกรณีที่ชุดข้อมูลมีขนาดใหญ่ และมีข้อมูลขาดหายจำนวนมาก

1.3 เติมค่าของข้อมูลที่ขาดหายไปทุกค่าด้วยค่าคงที่ค่าหนึ่ง เช่น ไม่รู้ค่า (Unknown)

1.4 ใช้ค่าเฉลี่ยของแอททริบิวต์ (Attribute) เติมค่าข้อมูลที่ขาดหาย เช่น ถ้าทราบว่าลูกค้าที่รายได้เฉลี่ยเดือนละ 28,000 บาท จะใช้ค่านี้นี้แทนค่ารายได้ของลูกค้าที่ขาดหาย

1.5 ใช้ค่าเฉลี่ยของแอททริบิวต์ของตัวอย่างที่จัดอยู่ในประเภทเดียวกัน เพื่อเติมค่าข้อมูลที่ขาดหาย เช่น เติมค่าความเสี่ยงด้านเครดิตด้วยค่าเฉลี่ยของรายได้จากลูกค้าที่อยู่ในกลุ่มอาชีพเดียวกัน

1.6 ใช้ค่าที่ความน่าจะเป็นไปได้มากที่สุดในการเติมแทนค่าข้อมูลที่ขาดหาย เช่น ค่าที่ได้จากสมการความถดถอย (Regression) ค่าที่ได้จากการอนุมาน โดยใช้สูตรของเบย์ (Bayesian formula) หรือต้นไม้ตัดสินใจ (Decision tree) เช่น ใช้ข้อมูลลูกค้า มาสร้างต้นไม้ตัดสินใจ เพื่อทำนายรายได้ของลูกค้า แล้วนำไปแทนค่าที่ขาดหาย วิธีนี้นิยมกันแพร่หลาย เนื่องจากทำนายค่าข้อมูลที่ขาดหาย โดยพิจารณาจากค่าของข้อมูลชุดปัจจุบัน และความสัมพันธ์ระหว่างคุณลักษณะในชุดข้อมูล

2. การผสานข้อมูล (Data Integration) เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกันเป็นการผสานข้อมูลจากแหล่งต่างๆ เพื่อช่วยลดหรือหลีกเลี่ยงความซ้ำซ้อนของข้อมูล ซึ่งจะนำไปสู่ปัญหาความไม่สอดคล้องกันของข้อมูลเพื่อเพิ่มความเร็วและคุณภาพในการทำเหมืองข้อมูล

งานที่ต้องทำในการผสานข้อมูล

2.1 การผสานโครงสร้างการเก็บข้อมูล (Schema Integration) โดยใช้ข้อมูลช่วยในการบ่งชี้คุณลักษณะในแหล่งเก็บข้อมูลต่างๆ เช่น คุณลักษณะรหัสลูกค้า (Customer\_id) ในแหล่งข้อมูล A เป็นคุณลักษณะเดียวกับ รหัสลูกค้า (Cust\_number) ในแหล่งข้อมูล B หรือไม่

2.2 ตรวจสอบและแก้ไขค่าข้อมูลที่ขัดแย้ง เช่น ค่าคุณลักษณะเดียวกัน แต่ใช้หน่วยวัดต่างกัน ในแต่ละแหล่งข้อมูล

2.3 การกำจัดค่าข้อมูลซ้ำซ้อน เช่น ค่าคุณลักษณะในแหล่งข้อมูลหนึ่ง อาจเป็นค่าที่ได้จากการคำนวณในอีกแหล่งข้อมูลหนึ่ง อาทิ ค่ารายได้รายปี

2.4 ค่าสหสัมพันธ์เป็น 0 หมายถึง ตัวแปรทั้งสองเป็นอิสระ ไม่มีความสัมพันธ์ระหว่างกัน

3. การคัดเลือกข้อมูล (Data Selection) เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึก โดยการกำหนดขนาดของกลุ่มตัวอย่าง (sample size) หมายถึง จำนวนข้อมูลกลุ่มตัวอย่าง ซึ่งผู้วิจัยจะกำหนดกลุ่มตัวอย่างว่าจะใช้จำนวนเท่าใด เนื่องจากข้อมูลตัวอย่างที่นำมามีขนาดของกลุ่มตัวอย่างจำนวนมาก

หลักการทฤษฎีของ ทาโร ยามาเน่ [4] จะทำให้ผลการวิเคราะห์ข้อมูลมีความเชื่อมั่นสูงโอกาสที่จะเกิดความคลาดเคลื่อนมีน้อย เนื่องจากทฤษฎีนี้ได้ใช้กันมากในงานวิจัยต่างๆ โดยมีวัตถุประสงค์เพื่อประมาณค่าสัดส่วนจากประชากรที่สนใจและมีจำนวนจำกัด (finite population) ซึ่งหลักทางทฤษฎีของทาโร ยามาเน่ ได้เสนอสูตรการคำนวณขนาดตัวอย่างสัดส่วน 1 กลุ่มโดยสมมุติค่าสัดส่วนเท่ากับ 0.5 และที่ระดับความเชื่อมั่น 95%

$$n = \frac{N}{1+Ne^2}$$

โดยที่

e คือ ความคลาดเคลื่อนของการเลือกตัวอย่าง

N คือ ขนาดของประชากร

n คือ ขนาดของกลุ่มตัวอย่าง

4. การแปลงข้อมูล (Data Transformation) [5] เป็นขั้นตอนการแปลงข้อมูลในขั้นตอนการคัดเลือกให้เหมาะสำหรับขั้นตอนการทำเหมืองข้อมูล การแปลงข้อมูลที่พบบ่อยในการทำเหมืองข้อมูลคือ การทำนอร์มอลไลซ์ (Normalization) โดยแปลงค่าข้อมูลให้อยู่ในช่วงสั้นๆ ที่ขั้นตอนวิธีการทำเหมืองข้อมูลสามารถนำไปใช้ประมวลผลได้

5. เหมืองข้อมูล (Data Mining) เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่ ซึ่งใช้เทคนิคของการทำเหมืองข้อมูลดังต่อไปนี้ การจัดหมวดหมู่ (Classification)

เป็นขั้นการสร้างตัวแบบโดยการเรียนรู้จากข้อมูลที่ได้กำหนดคลาส (Class) ไว้เรียบร้อยแล้วซึ่งตัวแบบที่ได้อาจแสดงในรูปแบบดังต่อไปนี้

#### 5.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree Technique)

การเรียนรู้แบบต้นไม้ตัดสินใจ [5] เป็นวิธีการเรียนรู้ที่ถูกใช้มากที่สุดในเรื่องของการอนุมาน โดยการประมาณสมการซึ่งแสดงในรูปแบบของต้นไม้หรือรูปแบบของกลุ่มของกฎ (if-then rules) ต้นไม้ตัดสินใจจะทำการแบ่งตัวอย่างโดยเรียงจากบนลงล่าง ซึ่งค่าของแต่ละบึงจ้ยรวมทั้งค่าของตัวอย่างจะมีลักษณะที่แน่นอน แบ่งเป็นกลุ่มต่างๆ และเป็นค่าที่ไม่ต่อเนื่อง (discrete value)

#### 5.2 การจัดกลุ่มฐานข้อมูล (Database Clustering)

เป็นเทคนิคการลดขนาดของข้อมูลด้วยการรวมกลุ่มตัวแปรที่มีลักษณะเดียวกันไว้ด้วยกันการแบ่งกลุ่มข้อมูล หรือ Clustering เป็นเทคนิคการแบ่งกลุ่มข้อมูลขนาดใหญ่ให้กลายเป็นกลุ่มย่อยๆ เราเรียกกลุ่มย่อยที่ถูกแบ่งออกนี้ว่า คลัสเตอร์ โดยใช้ความคล้ายคลึงกันของข้อมูลเป็นเกณฑ์ในการแบ่ง โดยข้อมูลที่ถูกรวมแยกออกจากกันจะถือว่ามีความคล้ายคลึงกันน้อย ส่วนข้อมูลที่อยู่กลุ่มเดียวกันจะถือว่ามีความคล้ายคลึงกันมาก การแบ่งกลุ่มข้อมูลเป็นเทคนิคที่ใช้แบ่งกลุ่มคน สัตว์ สิ่งของ องค์กร ฯลฯ ที่เหมือนกันหรือคล้ายกันไว้ในกลุ่มเดียวกัน แต่ต่างกันจะต้องอยู่คนละกลุ่ม การศึกษาความคล้ายพิจารณาจากตัวแปรที่ใช้ในการแบ่งกลุ่ม

6 รูปแบบการประเมินค่า (Pattern Evaluation) เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล ว่าขั้นตอนการประเมินผลของแบบจำลองที่ได้เกิดความผิดพลาดมากน้อยเพียงใด ผลลัพธ์จากการทำนายที่ได้นั้นเป็นไปตามความต้องการหรือไม่ โดยใช้เทคนิควิธีการประเมินความแม่นยำด้วยวิธีการไขว้ข้าม (k-fold cross-validation) เป็นวิธีการในการตรวจสอบค่าความผิดพลาดในการคาดการณ์ของแบบจำลองข้อมูล

7 การแทนความรู้ (Knowledge Representation) เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบ โดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ เมื่อทำตามขั้นตอนอย่างถูกต้องก็จะส่งผลให้ข้อมูลที่สืบค้นจากเหมืองข้อมูลมีความถูกต้องแม่นยำและสามารถใช้ช่วยสนับสนุนการตัดสินใจจริงได้ตามสิ่งที่ธุรกิจนั้นต้องการ

เทคโนโลยีที่ใช้ในการทำสารสนเทศ คือ ซอฟต์แวร์เวก้า (Waikato Environment for Knowledge Analysis) [6,7] ซึ่งเขียนมาโดยเน้นกับงานทางด้านการศึกษาด้วยเครื่อง (Machine Learning) และการทำเหมืองข้อมูล ซอฟต์แวร์จะประกอบไปด้วยโมดูลย่อยๆ สำหรับใช้ในการจัดการข้อมูล

เป็นซอฟต์แวร์ที่สามารถใช้ส่วนติดต่อผู้ใช้แบบกราฟิก (Graphic User Interface) เป็นซอฟต์แวร์ที่มีการสนับสนุนสำหรับกระบวนการของการทดลองทำเหมืองข้อมูล เช่น การเตรียมการของการนำเข้าข้อมูล การประเมินผลทางสถิติของรูปแบบการเรียนรู้ การแสดงการนำเข้าข้อมูลและผลของการเรียนรู้

ซึ่งผู้วิจัยจะนำซอฟต์แวร์เวก้า มาช่วยด้านเทคนิคการทำเหมืองข้อมูลแบบการจำแนก เทคนิคการจำแนกที่ใช้คือ J48 เป็นขั้นตอนวิธีในการจำแนกโดยใช้ต้นไม้การตัดสินใจ C4.5 วิธีนี้จะใช้ข้อมูลในการสร้างต้นไม้ตัดสินใจ โดยที่แต่ละโหนด หมายถึง แอททริบิวต์ แต่ละกิ่งของต้นไม้เป็นผลในการทดสอบและโหนดใบแสดงคลาส เพื่อวิเคราะห์การหาบึงจ้ยและแนวทางการวิเคราะห์ประเภทกรรมธรรม์ที่ถูกคัดต้องการทำประกันภัยและเพื่อพัฒนาสร้างตัวแบบการวิเคราะห์ความสัมพันธ์ของบึงจ้ยต่างๆกับพฤติกรรมการทำประกันชีวิตของผู้บริโภค

งานวิจัยที่เกี่ยวข้อง

กฤษณพร สุริยะบรรเทิง, ดร.กมล เกียรติเรืองภมลา (2555) ศึกษาเรื่อง การสร้างแบบจำลองการขายผลิตภัณฑ์ประกันภัยให้กับลูกค้าผู้สูงอายุกลุ่ม บัญชีออมทรัพย์โดยการทำเหมืองข้อมูล งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองที่สามารถสนับสนุนกิจกรรมทางการตลาดระหว่างสององค์กร โดยใช้เทคนิคการทำเหมืองข้อมูล ตามกรอบการทำเหมืองข้อมูล CRISP-DM (Cross-Industry Standard Process for Data Mining) โดยการสร้างแบบจำลองการแบ่งกลุ่มด้วยขั้นตอนวิธี SimpleKMeans และสร้างแบบจำลองการหาความสัมพันธ์ด้วยขั้นตอนวิธี Apriori ซึ่งผลวิจัยที่ได้สามารถนำไปใช้สนับสนุนกิจกรรมทางการตลาดขององค์กรได้ เช่น การสร้างโปร โมชั่น หรือการออกผลิตภัณฑ์ใหม่ที่เหมาะสมกับลูกค้าแต่ละกลุ่ม เพื่อให้เกิดความพึงพอใจสูงสุด และความภักดีต่อองค์กร รวมถึงการกระตุ้นการซื้อผลิตภัณฑ์ของลูกค้า เป็นการศึกษาในรูปแบบวิจัยเชิงทดลอง (Experimental Research) โดยการนำข้อมูลลูกค้าผู้สูงวัยกลุ่มบัญชีออมทรัพย์ธนาคารพาณิชย์แห่งหนึ่ง และข้อมูลผลิตภัณฑ์ของบริษัทประกันแห่งหนึ่ง จากฐานข้อมูลที่มีอยู่แล้ว ซึ่งเป็นข้อมูลทุติยภูมิ (Secondary Data) มาศึกษาค้นหาความรู้ ความสัมพันธ์ของข้อมูลที่แฝงอยู่ ด้วยเทคนิคการทำเหมืองข้อมูลตามขั้นตอนของ CRISP-DM เพื่อสร้างรูปแบบจำลองผลิตภัณฑ์ใหม่ให้กับผู้สูงอายุซึ่งเป็นลูกค้าบัญชีออมทรัพย์ธนาคาร จากขั้นตอนการเก็บรวบรวมข้อมูลได้เพิ่มข้อมูลที่เกี่ยวข้อง ข้อมูลที่ใช้ในการสร้างแบบจำลองจะใช้ข้อมูลการซื้อผลิตภัณฑ์ของลูกค้าบัญชีออมทรัพย์ของธนาคาร โดยข้อมูลที่นำมาใช้ในการทำเหมืองข้อมูล คือ เพศของลูกค้า อายุของลูกค้า สถานภาพของลูกค้า รายได้ของลูกค้า ระดับการศึกษาของลูกค้า ยอดเงินในบัญชีของลูกค้า ระยะเวลาในการเป็นลูกค้าธนาคาร ภูมิภาค ประเภทผลิตภัณฑ์

ศิริกาญจนา พิลาบุตร (2551) ศึกษาเรื่อง การสร้างกฎข้อบังคับของฐานข้อมูลโดยการทำเหมืองข้อมูล งานวิจัยนี้มีจุดมุ่งหมายที่จะพัฒนากฎที่ได้จากการทำเหมืองข้อมูลเพื่อนำมาใช้สร้างกฎข้อบังคับการเปลี่ยนแปลงฐานข้อมูลหรือที่รู้จักกันในชื่อว่าค่าเบสทริกเกอร์ (Database trigger) โดยโปรแกรมที่พัฒนาขึ้นจะใช้ภาษา JAVA ในการพัฒนาและนำเอา

เทคโนโลยีทางด้านเหมืองข้อมูลที่เริ่มเป็นที่รู้จักกันแพร่หลายนี้มาประยุกต์ใช้ในการค้นหารูปแบบของข้อมูล เพื่อนำมาสร้างเป็นกฎข้อบังคับโดยขั้นตอนวิธีที่ใช้ในการค้นหารูปแบบของข้อมูลเพื่อนำมาใช้ในการสร้างเป็นกฎข้อบังคับนี้คือขั้นตอนวิธี C4.5 ซึ่งเป็นขั้นตอนวิธีที่ใช้ในการทำเหมืองข้อมูลประเภทการจำแนกโดยมุ่งเน้นให้ขั้นตอนวิธีสามารถสร้างกฎการจำแนกที่ได้ให้อยู่ในรูปอย่างง่ายต่อการนำไปใช้ในการสร้างกฎข้อบังคับของข้อมูล รายละเอียดในเนื้อหางานวิจัยนี้ประกอบด้วยขั้นตอนการวิจัย กระบวนการทำงานของระบบสร้างค่าเบสทริกเกอร์ วิธีการค้นหากฎการจำแนกด้วยโปรแกรม WEKA และขั้นตอนวิธีการแปลงกฎการจำแนกเป็นค่าเบสทริกเกอร์เพื่อใช้ประโยชน์ในการควบคุมความถูกต้องของข้อมูลในฐานะข้อมูล

บุญเสริม กิจศิริกุล(2546) ศึกษาเรื่อง โครงการวิจัยร่วมภาครัฐและเอกชน ปีงบประมาณ 2545 โครงการย่อยที่ 7 ขั้นตอนวิธีการทำเหมืองข้อมูล งานวิจัยนี้ทำการศึกษาวิจัยขั้นตอนวิธีการทำเหมืองข้อมูลและการสร้างโปรแกรมสำหรับการทำงานเหมืองข้อมูล อธิบายรายละเอียดขั้นตอนการทำเหมืองข้อมูลด้วยเทคนิคต่างๆที่ใช้ในการทำเหมืองข้อมูล เช่น เทคนิคต้นไม้ตัดสินใจและวิธีการเรียนรู้ต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคการเรียนรู้เบย์ (Bayesian) เทคนิคการค้นหากฎความสัมพันธ์ (Association Rule) ซึ่งได้อธิบายเกี่ยวกับขั้นตอนเทคนิคของการทำเหมืองข้อมูลไว้อย่างละเอียด และมีการยกตัวอย่างประกอบการอธิบายความแตกต่างของแต่ละเทคนิคในการทำเหมืองข้อมูลสามารถนำมาประกอบการทำวิจัยเกี่ยวกับการทำเหมืองข้อมูลได้เป็นอย่างดี

ธัญพร ชัยสุข โภสกล (2553) ศึกษาเรื่อง การวิเคราะห์ข้อมูลผู้บริโภคเพื่อทำนายพฤติกรรมการซื้อสินค้า กรณีศึกษา “ร้านค้าสวัสดิการ” งานวิจัยนี้นำเสนอการวิเคราะห์ความสัมพันธ์ของโรคไขมันและหัวใจกับพฤติกรรมผู้บริโภคในการเลือกซื้อสินค้าโดยประเมินค่าจากการวัดประเภทระดับไขมันในเลือด 4 ประเภทคือ HDL, LDL, คอเลสเตอรอล และไตรกลีเซอไรด์ โดยใช้สองเทคนิคในการทำเหมืองข้อมูล ได้แก่ เทคนิคกฎความสัมพันธ์เพื่อหาความสัมพันธ์ของผู้บริโภคที่เกิดความเสี่ยงต่อโรคกับพฤติกรรมการซื้อสินค้า และการแบ่งกลุ่มเพื่อจัดกลุ่มผู้บริโภคจากลักษณะประจำหรือพฤติกรรมผู้บริโภค จากการศึกษาพบว่าการใช้เทคนิคกฎความสัมพันธ์สามารถทำนายพฤติกรรมและลักษณะผู้บริโภคที่เกิดความเสี่ยงต่อโรคได้ถูกต้อง 74% สำหรับการแบ่งกลุ่มผู้บริโภคนั้นใช้วิธีการแบ่งกลุ่มแบบ 2 ขั้นตอน คือใช้ Hierarchical Clustering เพื่อหาค่า k ที่เหมาะสมที่สุด หลังจากนั้นจึงนำค่า k มากำหนดใน K-means Clustering

จากข้อเสนอแนะจากผู้ทำวิจัยทั้งหมดข้างต้น ทำให้ทราบแนวทางในการวิจัยที่จะนำมาประยุกต์ใช้กับงานวิจัยของผู้วิจัย อาทิเช่น การเลือกเทคนิคการสร้างแบบจำลอง ผู้วิจัยได้เลือกใช้ขั้นตอนวิธี SimpleKMeans สำหรับการแบ่งกลุ่มลูกค้าที่ถือครองผลิตภัณฑ์ ในการทำเหมืองข้อมูลรวมทั้งเทคนิคต่างๆ ไม่ว่าจะเป็น เทคนิคการรวบรวมข้อมูล เทคนิคการเตรียมข้อมูล เทคนิคการพัฒนาแบบจำลอง ที่จะนำมาใช้ประกอบการ

ทำงานวิจัย พร้อมทั้งได้ทราบถึงขั้นตอนการดำเนินการวิจัย เพื่อให้สามารถดำเนินได้ราบรื่นและสำเร็จไปได้ด้วยดี

### 3. วิธีดำเนินการวิจัย

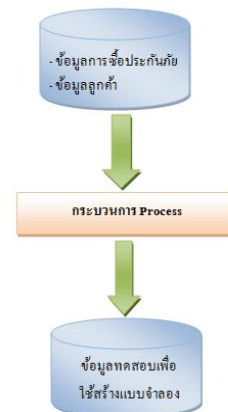
การศึกษาวิจัยการวิเคราะห์พัฒนาสร้างตัวแบบความสัมพันธ์ของปัจจัยต่างๆกับพฤติกรรมการทำประกันภัยของผู้บริโภค ผู้วิจัยได้กำหนดขั้นตอนของการดำเนินงานประกอบด้วยขั้นตอนดังนี้

1. การรวบรวมข้อมูลที่เกี่ยวข้องกับการวิจัย
2. การเตรียมข้อมูลประชากรและกลุ่มตัวอย่าง
3. ขั้นตอนวิธีที่ใช้ในดำเนินการวิจัย

1. การรวบรวมข้อมูลที่เกี่ยวข้องกับการวิจัย

ผู้วิจัยได้รวบรวมข้อมูลที่เกี่ยวข้องกับงานวิจัย ประกอบด้วยฐานข้อมูล ที่ใช้ในการวิเคราะห์ผ่านกระบวนการคลังข้อมูลเพื่อให้ได้ข้อมูลที่เหมาะสมต่อการวิจัย ดังนี้

- 1.1 ฐานข้อมูลการซื้อประกันภัย เป็นที่เก็บข้อมูลการซื้อประกันภัยในแต่ละประเภทของลูกค้าในแต่ละวัน
- 1.2 ฐานข้อมูลลูกค้า เป็นที่เก็บข้อมูลประวัติของลูกค้าที่ซื้อประกันภัย เช่น ชื่อ นามสกุล วัน เดือน ปีเกิด ที่อยู่ และอาชีพของลูกค้าที่ซื้อประกันภัย



รูปที่ 1. แหล่งที่มาของข้อมูลพฤติกรรมของผู้ทำประกันภัย

2. การเตรียมข้อมูลประชากรและกลุ่มตัวอย่างเพื่อนำมาใช้สร้างตัวแบบการวิเคราะห์ความสัมพันธ์ของปัจจัยต่างๆกับพฤติกรรมการทำประกันภัยของผู้บริโภค ผู้วิจัยได้นำข้อมูลการทำประกันภัยของลูกค้าในบริษัทประกันภัยแห่งหนึ่งมาใช้ในการวิเคราะห์ โดยข้อมูลของลูกค้าที่ทำประกันภัยนั้นเป็นข้อมูลในช่วงปี 2555-2556 ประกอบด้วย 4 ข้อมูลปัจจัย ได้แก่ ข้อมูลอายุ , เพศ ,อาชีพ และข้อมูลจังหวัดของคนที่ทำประกันภัยประเภทต่างๆ การวิเคราะห์เพื่อให้ข้อมูลเหมาะสมกับเทคนิคที่นำมาวิเคราะห์ที่เป็นขั้นตอนที่ใช้เวลานาน เนื่องจากการวิเคราะห์ข้อมูลที่มีคุณภาพ และอยู่ในรูปแบบที่เหมาะสมจะทำให้ผลลัพธ์ที่ถูกต้องและตรงกับความเป็นจริง โดยแบ่งเป็น 5 ขั้นตอนย่อยดังนี้

## 2.1 การเตรียมข้อมูล

ในงานวิจัยนี้ผู้วิจัยได้นำข้อมูลบางส่วนมาใช้ในการทดลองจำนวน 25,982 ระเบียบ ต้องทำการคัดเลือกข้อมูลที่จำเป็นของลูกค้าที่ทำประกันภัย และจัดทำเป็นคลังข้อมูลเพื่อสะดวกแก่การนำไปใช้ โดยมีรายละเอียดดังนี้

### 2.1.1 ฐานข้อมูลการซื้อประกันภัย

ในงานนี้ผู้วิจัยได้นำข้อมูลบางส่วนมาใช้ในการทดลองคือข้อมูลการซื้อขายระหว่างเดือนมกราคม 2555 ถึงธันวาคม 2556 จำนวน 25,982 ระเบียบ โดยคัดเลือกเฉพาะข้อมูลที่จำเป็นต้องใช้ในการงานวิจัย

### 2.2 การทำข้อมูลให้สมบูรณ์

การทำข้อมูลการทำประกันภัยของลูกค้าให้มีความสมบูรณ์ โดยการเติมค่าของข้อมูลหรือการคัดข้อมูลที่ไมเกี่ยวข้องออกไป เช่น อายุของคนทำประกันภัย ได้มาจากการนำปีที่ทำกรมธรรม์ลบปีพุทธศักราชที่ลูกค้าเกิด , อาชีพของคนที่ทำประกันภัยที่ไม่ระบุ ไปค่าเป็น ไม่ทราบอาชีพ , เพศของผู้ทำประกันภัย ได้จากคำนำหน้าชื่อของลูกค้า เช่น นางสาว คือ เพศหญิง

### 2.3 การผสานข้อมูล

การรวมข้อมูลการทำประกันภัยที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน เพื่อช่วยลดหรือหลีกเลี่ยงความซ้ำซ้อนของข้อมูล เพื่อเพิ่มความเร็วและคุณภาพในการทำเหมืองข้อมูล

### 2.4 การคัดเลือกข้อมูล

เนื่องจากข้อมูลที่ใช้สำหรับทำการวิจัยจากแหล่งที่บันทึกมีข้อมูลเป็นจำนวนมาก ผู้วิจัยจึงดึงจำนวนของกลุ่มตัวอย่างมาเพื่อใช้ในการวิเคราะห์ข้อมูล โดยใช้หลักการของทฤษฎี ทาโร่ ยามาเน่ จะทำให้ผลการวิเคราะห์ข้อมูลมีความเชื่อมั่นสูง โอกาสที่จะเกิดความคลาดเคลื่อนมีน้อย เนื่องจากทฤษฎีนี้ได้ใช้กันมากในงานวิจัยต่างๆ การคำนวณขนาดตัวอย่างสัดส่วน 1 กลุ่มโดยสมมุติค่าสัดส่วนเท่ากับ 0.5 และที่ระดับความเชื่อมั่น 95%

ข้อมูลที่ใช้ในการวิจัยจำนวน 25,982 ระเบียบ ถ้าอนุญาตให้เกิดความคลาดเคลื่อนในการสุ่มร้อยละ 5 เมื่อนำมาแทนค่าในสูตรการคำนวณของทาโร่ ยามาเน่ จะได้ว่า

$$\begin{aligned} N &= 25,982 \\ e &= 0.05 \\ N &= \frac{25,982}{1+25,982 (0.05)^2} \\ &= \frac{25,982}{1+64.95} \\ &= 393.96 \end{aligned}$$

กลุ่มตัวอย่างที่ใช้ในการวิจัย = 394 คน

## 2.5 การแปลงข้อมูล มีกระบวนการย่อยดังนี้

### 2.5.1 การปรับเปลี่ยนรูปแบบข้อมูล

ผู้วิจัยปรับเปลี่ยนรูปแบบข้อมูลโดยทำการนำข้อมูลมาเชื่อมต่อกันด้วยเทคนิคการทาคัดล้างข้อมูลและเปลี่ยนรูปแบบข้อมูลให้ครอบคลุมตัวชี้วัดที่กำหนด โดยงานวิจัยนี้เลือกใช้ Microsoft Access มาใช้ในการออกแบบคลังข้อมูล

### 2.5.2 การลดขนาดข้อมูล

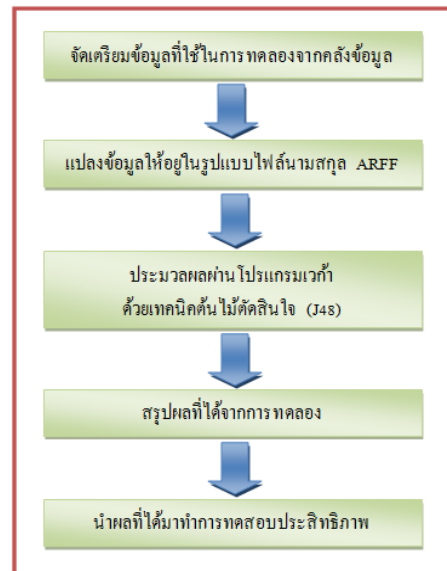
ผู้วิจัยทำการลดขนาดข้อมูล และลดจำนวนมุมมองของข้อมูลที่ไม่จำเป็น และเปลี่ยนรูปแบบข้อมูลให้ง่ายต่อการวิเคราะห์ โดยการเลือกข้อมูลเฉพาะที่สนใจ และทำการจัดกลุ่มประกันภัยใหม่เพื่อให้เหมาะสมต่องานวิจัย

### 3. ขั้นตอนวิธีที่ใช้ในดำเนินการวิจัย

งานวิจัยนี้ใช้เทคนิคการทำเหมืองข้อมูลแบบการจัดหมวดหมู่ โดยใช้เทคนิคต้นไม้ตัดสินใจ ซึ่งต้นไม้ตัดสินใจนั้น เป็นการเรียนรู้โดยการจัดประเภทข้อมูลออกเป็นกลุ่มต่างๆ โดยใช้คุณลักษณะข้อมูลในการจำแนกประเภท ต้นไม้ตัดสินใจที่ได้จากการเรียนรู้ทำให้ทราบว่าคุณลักษณะใดเป็นตัวกำหนดการจำแนกประเภท และคุณลักษณะแต่ละตัวที่มีความสำคัญมากน้อยต่างกันอย่างไร เพราะฉะนั้น การจำแนกประเภทมีประโยชน์ช่วยให้ผู้วิจัยสามารถวิเคราะห์ข้อมูลและตัดสินใจได้ถูกต้องยิ่งขึ้น เกี่ยวกับพฤติกรรมกรรมการซื้อประกันภัยในแต่ละประเภทของลูกค้า เทคนิคต้นไม้ตัดสินใจ มีขั้นตอนรายละเอียดดังต่อไปนี้

#### 3.1 เทคนิคต้นไม้ตัดสินใจ

เทคนิคต้นไม้ตัดสินใจผู้วิจัยจะนำข้อมูลการซื้อประกันภัยระหว่างปี 2555 – 2556 จำนวน 25,982 ระเบียบ และใช้สูตรการคำนวณของทาโร่ ยามาเน่ ได้จำนวนกลุ่มตัวอย่างลูกค้าที่ซื้อประกันภัยจำนวน 394 คน เป็นกลุ่มทดสอบประสิทธิภาพ โดยมีขั้นตอนการทำงานดังรูปภาพที่ 2



รูปภาพที่ 2. ลำดับขั้นตอนการทำงานด้วยเทคนิคต้นไม้ตัดสินใจ

ในเทคนิคต้นไม้ตัดสินใจนี้ ทางผู้วิจัยได้ทำการเตรียมข้อมูลในรูปแบบไฟล์นามสกุล ARFF แล้วจึงนำไปทดสอบแบบจำลองในโปรแกรมเวก้า เวอร์ชัน 3.6.11 ซึ่งเป็นเครื่องมือที่ช่วยในการทดสอบดังรูปภาพที่ 3

```

@relation policy

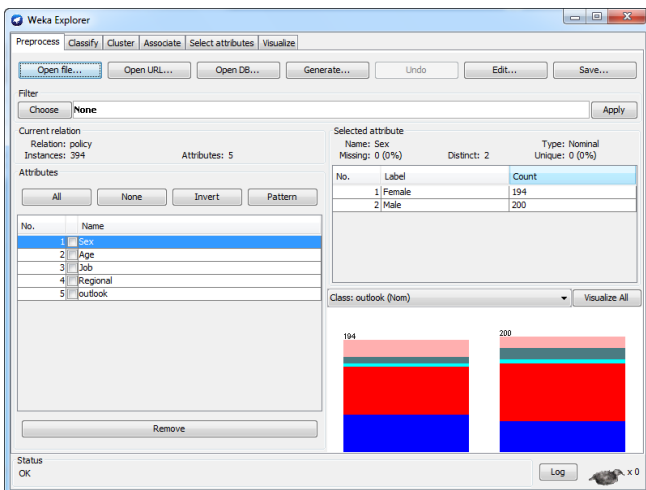
@attribute 'Sex' {Female, Male}
@attribute 'Age' {10_35, 36_60, 61_up}
@attribute 'Job' {Unknown, Risk, No-Risk}
@attribute 'Regional' {North, Central, Northeast, South}
@attribute outlook {MSHOB, MSPA, MT02, MT03, MTHC}

@data
Male, 10_35, Unknown, Northeast, MSHOB
Female, 36_60, Unknown, Central, MSHOB
Male, 10_35, Unknown, Central, MSHOB
Male, 36_60, Unknown, Central, MSHOB
Female, 36_60, Unknown, Central, MSHOB
Female, 36_60, Unknown, Central, MSHOB
Female, 36_60, Unknown, Central, MSHOB
Female, 36_60, Unknown, Central, MSHOB
Male, 36_60, Unknown, Central, MSHOB
Male, 10_35, Unknown, Central, MSHOB
Female, 36_60, Unknown, Central, MSHOB
Male, 36_60, Unknown, Central, MSHOB
Male, 36_60, Unknown, Central, MSHOB

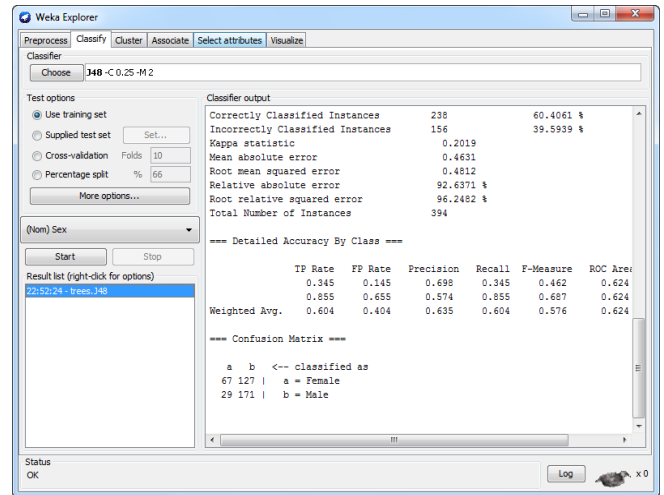
```

รูปภาพที่ 3. การเตรียมข้อมูลในรูปแบบไฟล์นามสกุล ARFF

หลังจากได้ไฟล์ข้อมูลแล้ว นำมาเข้าสู่ โปรแกรมเวก้า เพื่อทำการประมวลผลด้วยเทคนิคขั้นตอนวิธี ต้นไม้ตัดสินใจ ผลลัพธ์ที่ได้ออกมาดังรูปภาพที่ 4 และ รูปภาพที่ 5



รูปภาพที่ 4 การนำข้อมูลเข้าโปรแกรมเวก้า



รูปภาพที่ 5 การประมวลผลด้วยเทคนิคขั้นตอนวิธีต้นไม้ตัดสินใจ (J48)

### 3.2 การจัดกลุ่มฐานข้อมูล

ผู้วิจัยได้ใช้การวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน โดยการใช้หลักการของ K-Means Clustering การใช้เทคนิค K-Means Clustering เพื่อเป็นประโยชน์ในการช่วยลดขนาดข้อมูลด้วยการรวมกลุ่มตัวแปรที่มีลักษณะเดียวกันไว้ด้วยกัน โดยแบ่งกลุ่มข้อมูลขนาดใหญ่ให้กลายเป็นกลุ่มย่อยๆ โดยใช้ความคล้ายคลึงกันของข้อมูลในการกำหนดจำนวนกลุ่มที่เหมาะสมซึ่งจะช่วยให้ผู้วิจัยสามารถจัดแยกกลุ่มพฤติกรรมของลูกค้าที่จะเลือกซื้อประกันภัยในแต่ละประเภทได้ การแบ่งกลุ่มด้วยขั้นตอนวิธีแบบ K-Means Clustering ในงานวิจัยนี้เลือกใช้ซอฟต์แวร์เวอร์ก้า 3.6.11 เป็นเครื่องมือในการทดลอง มีขั้นตอนการทดลองดังนี้



รูปภาพที่ 6 ลำดับขั้นตอนการทำงานด้วยเทคนิค K-Means Clustering

ในเทคนิค K-Means Clustering ทางผู้วิจัยได้ทำการเตรียมข้อมูลในรูปแบบไฟล์นามสกุล ARFF แล้วจึงนำไปทดสอบแบบจำลองโปรแกรมเวก้า เวอร์ชัน 3.6.11 ซึ่งเป็นเครื่องมือที่ช่วยในการทดสอบผังรูปภาพที่ 7

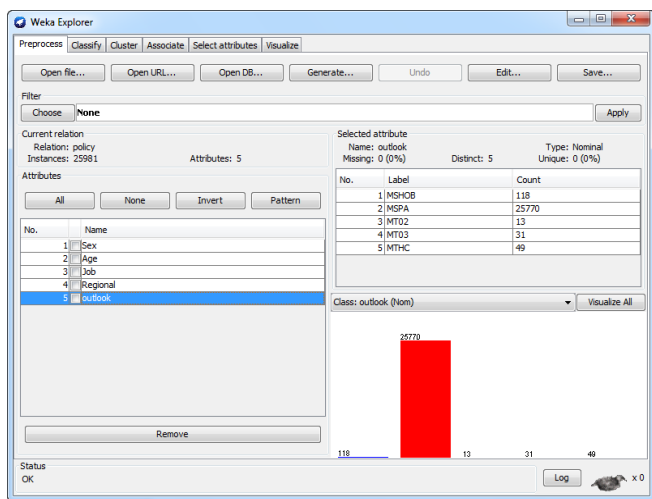
```
@relation policy

@attribute 'Sex' {Female, Male}
@attribute 'Age' {10_20, 21_40, 41_60, 61_up}
@attribute 'Job' {Unknown, Risk, No-Risk}
@attribute 'Regional' {North, Central, Northeast, South}
@attribute outlook {MSHOB, MSPA, MT02, MT03, MTHC}

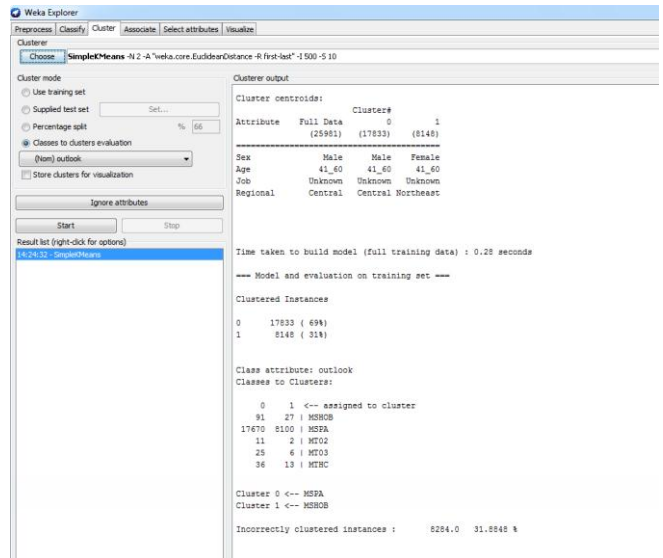
@data
Male, 21_40, Unknown, Northeast, MSHOB
Female, 41_60, Unknown, Central, MSHOB
Male, 21_40, Unknown, Central, MSHOB
Male, 41_60, Unknown, Central, MSHOB
Female, 41_60, Unknown, Central, MSHOB
Female, 41_60, Unknown, Central, MSHOB
Female, 41_60, Unknown, Central, MSHOB
Female, 41_60, Unknown, Central, MSHOB
Male, 41_60, Unknown, Central, MSHOB
Male, 21_40, Unknown, Central, MSHOB
Female, 41_60, Unknown, Central, MSHOB
Male, 41_60, Unknown, Central, MSHOB
Male, 21_40, Unknown, Central, MSHOB
Male, 41_60, Unknown, Central, MSHOB
Male, 41_60, Unknown, Central, MSHOB
Female, 41_60, Unknown, Central, MSHOB
Female, 41_60, Unknown, Central, MSHOB
Male, 61_up, Unknown, Central, MSHOB
```

รูปภาพที่ 7 การเตรียมข้อมูลเพื่อทำ K-Means Clustering ในรูปแบบไฟล์นามสกุล ARFF

หลังจากได้ไฟล์ข้อมูลแล้ว นำมาเข้าสู่โปรแกรมเวก้า เพื่อทำการประมวลผลด้วยขั้นตอนวิธี K-Means Clustering ผลลัพธ์การแบ่งกลุ่มที่ได้ ออกมาดังรูปภาพที่ 8 และรูปภาพที่ 9



รูปภาพที่ 8 การนำข้อมูลไฟล์ ARFF เข้าโปรแกรมเวก้า



รูปภาพที่ 9 ผลลัพธ์ขั้นตอนวิธีแบบ K-Means Clustering

#### 4. ผลการวิจัย

ผลการวิจัยจำนวนข้อมูลที่มีจำนวน 25,981 รายการ นำกลุ่มตัวอย่างมาใช้ในการสร้างแบบจำลองต้นไม้ตัดสินใจตามหลักทฤษฎีทาร์โรว์ ยามาเน่ จำนวน 394 รายการ ผลการทดลองจะสรุปได้ว่าค่าความถูกต้อง 60.40% ดังหน้าจอรูปภาพที่ 10 สามารถอธิบายได้ดังต่อไปนี้

- Outlook = MSHOB
    - | Regional = North: Female (3.0)
    - | Regional = Central: Male (85.0/38.0)
    - | Regional = Northeast: Female (28.0/6.0)
    - | Regional = South: Female (2.0)
  - Outlook = MSPA
    - | Job = Unknown: Male (141.0/61.0)
    - | Job = Risk
      - | | Regional = North: Male (3.0/1.0)
      - | | Regional = Central: Female (2.0/1.0)
      - | | Regional = Northeast: Male (4.0/1.0)
      - | | Regional = South: Female (12.0/2.0)
    - | Job = No-Risk: Male (21.0/9.0)
  - Outlook = MT02: Male (13.0/6.0)
  - Outlook = MT03: Male (31.0/11.0)
  - Outlook = MTHC: Female (49.0/20.0)
- Number of Leaves: 13  
Size of the tree: 17

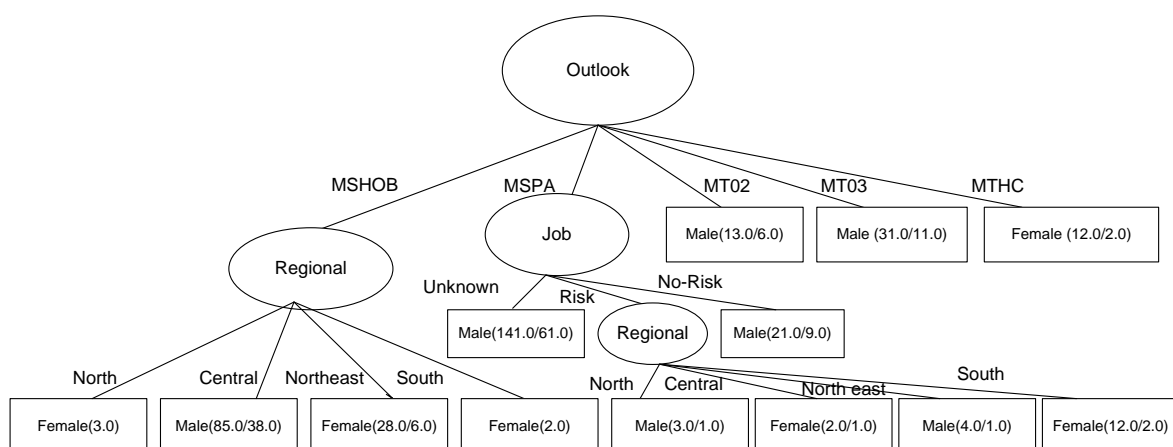
นำผลลัพธ์ที่ได้มาแสดงเป็นแบบจำลองต้นไม้ตัดสินใจพฤติกรรมกรมการท่าอากาศยานของสุวรรณภูมิ ตามรูปภาพที่ 11 สามารถอธิบายได้ดังนี้

บรรทัดที่ 1) ผลិតภัณฑ์ MSHOB  
 บรรทัดที่ 2) คนทำประกันภัยอยู่ภาคเหนือ 3 คน เป็นผู้หญิง = 3  
 บรรทัดที่ 3) คนทำประกันภัยอยู่ภาคกลาง 85 คน เป็นผู้ชาย = 47 / ผู้หญิง = 38  
 บรรทัดที่ 4) คนทำประกันภัยอยู่ภาคตะวันออกเฉียงเหนือ 28 คน เป็นผู้หญิง = 22 / ผู้ชาย = 6  
 บรรทัดที่ 5) คนทำประกันภัยอยู่ภาคใต้ 2 คน เป็นผู้หญิง = 2  
 บรรทัดที่ 6) ผลิตภัณฑ์ MSPA  
 บรรทัดที่ 7) ตำแหน่งงาน ไม่ทราบอาชีพ 141 คน เป็นผู้ชาย = 80 / ผู้หญิง = 61  
 บรรทัดที่ 8-9) ตำแหน่งงานที่มีความเสี่ยง 3 คน คนที่ทำประกันภัยอาศัยอยู่ภาคเหนือ และเป็นผู้ชาย = 2 / ผู้หญิง = 1  
 บรรทัดที่ 10) ตำแหน่งงานที่มีความเสี่ยง 2 คน คนที่ทำประกันภัยอาศัยอยู่ภาคกลาง และเป็นผู้หญิง = 1 / ผู้ชาย = 1  
 บรรทัดที่ 11) ตำแหน่งงานที่มีความเสี่ยง 4 คน คนที่ทำประกันภัยอาศัยอยู่ภาคตะวันออกเฉียงเหนือและเป็นผู้ชาย = 3 / ผู้หญิง = 1  
 บรรทัดที่ 12) ตำแหน่งงานที่มีความเสี่ยง 12 คน คนที่ทำประกันภัยอาศัยอยู่ภาคใต้ และเป็นผู้หญิง = 10 / ผู้ชาย = 2  
 บรรทัดที่ 13) ตำแหน่งงานที่ไม่มีความเสี่ยง 21 คนทำประกันเป็นผู้ชาย = 12 / ผู้หญิง = 9  
 บรรทัดที่ 14) ผลิตภัณฑ์ MT02 13 คน เป็นผู้ชาย = 7 / ผู้หญิง = 6  
 บรรทัดที่ 15) ผลิตภัณฑ์ MT03 31 คน เป็นผู้ชาย = 20 / ผู้หญิง = 11  
 บรรทัดที่ 16) ผลิตภัณฑ์ MTHC 49 คน เป็นผู้ชาย = 29 / ผู้หญิง = 20  
 จำนวนใบของต้นไม้เท่ากับ 13  
 ขนาดของต้นไม้เท่ากับ 17

## 5. สรุปผล

งานวิจัยนี้นำเสนอแบบจำลองสำหรับการทำนายพฤติกรรมการทำประกันภัยของผู้บริโภค ผลการทดลองการจัดหมวดหมู่พบว่าแบบจำลองสามารถทำนายถูกต้อง 60.40% ดังรายละเอียดคือ

- ผลิตภัณฑ์ MSPA
  - ตำแหน่งงาน ไม่ทราบอาชีพ 141 คน เป็นผู้ชาย = 80 / ผู้หญิง = 61
  - ตำแหน่งงานที่มีความเสี่ยง 3 คน คนที่ทำประกันภัยอาศัยอยู่ภาคเหนือ และเป็นผู้ชาย = 2 / ผู้หญิง = 1
  - ตำแหน่งงานที่มีความเสี่ยง 2 คน คนที่ทำประกันภัยอาศัยอยู่ภาคกลาง และเป็นผู้หญิง = 1 / ผู้ชาย = 1
  - ตำแหน่งงานที่มีความเสี่ยง 4 คน คนที่ทำประกันภัยอาศัยอยู่ภาคตะวันออกเฉียงเหนือและเป็นผู้ชาย = 3 / ผู้หญิง = 1
  - ตำแหน่งงานที่มีความเสี่ยง 12 คน คนที่ทำประกันภัยอาศัยอยู่ภาคใต้ และเป็นผู้หญิง = 10 / ผู้ชาย = 2
  - ตำแหน่งงานที่ไม่มีความเสี่ยง 21 คนทำประกันเป็นผู้ชาย = 12 / ผู้หญิง = 9
- ผลิตภัณฑ์ MT02 13 คน
  - เป็นผู้ชาย = 7 / ผู้หญิง = 6
- ผลิตภัณฑ์ MT03 31 คน
  - เป็นผู้ชาย = 20 / ผู้หญิง = 11
- ผลิตภัณฑ์ MTHC 49 คน
  - เป็นผู้ชาย = 29 / ผู้หญิง = 20



รูปภาพที่ 11 แบบจำลองต้นไม้ตัดสินใจพฤติกรรมการทำประกันภัยของผู้บริโภค



ผลการแบ่งกลุ่มโดยใช้วิธีแบบ K-Means Clustering สามารถแบ่งกลุ่มลักษณะผู้บริโภครังนี้

กลุ่มที่ 0 ผลัดกันทำงานกันภัยอุบัติเหตุส่วนบุคคล (MSPA) : กลุ่มผู้ทำประกันภัยเป็นผู้ชาย อายุระหว่าง 41-60 ปี ไม่ทราบอาชีพ อาศัยอยู่ภาคกลาง เลือกรับประกันชีวิตประเภท การประกันภัยอุบัติเหตุส่วนบุคคล (MSPA) กลุ่มนี้มีขนาดคิดเป็น 69%

กลุ่มที่ 1 ผลัดกันทำงานกันภัยสำหรับเจ้าบ้าน (MSHOB) : กลุ่มผู้ทำประกันภัยเป็นผู้หญิง อายุระหว่าง 41-60 ปี ไม่ทราบอาชีพ อาศัยอยู่ภาคตะวันออกเฉียงเหนือ เลือกรับประกันชีวิตประเภท การประกันภัยสำหรับเจ้าบ้าน (MSHOB) กลุ่มนี้มีขนาดคิดเป็น 31%

### เอกสารอ้างอิง

- [1] สำนักงานคณะกรรมการกำกับและส่งเสริมธุรกิจประกันภัย. (2557). ข้อมูลการทำประกันภัย. [ออนไลน์]. สืบค้นเมื่อ 5 พฤษภาคม 2557, จาก <http://www.oic.or.th/>
- [2] กัลยา วานิชย์บัญชา.(2548).การวิเคราะห์ข้อมูลหลายตัวแปร.กรุงเทพฯ. จุฬาลงกรณ์มหาวิทยาลัย.
- [3] Jiawei Hun and Micheline Kamber. (2001). Data Mining Concepts and Techniques. United States of America.
- [4] Yamane, Taro. (1967). Statistics, An Introductory Analysis, 2nd Ed., New York: Harper and Row.
- [5] Daniel T. Larose. (2005). Discovering Knowledge in Data : An Introduction to DATA MINING. Canada.
- [6] Quinlan, J.R. (1986). Induction of decision trees, pp. 81-106, Machine Learning 1.
- [6] มาโนช ห้วยหงษ์ทอง, นางสาววิภารัตน์ กุลที และนางสาวสรารัตน์ แก้วมานถ.(2553).คู่มือแนะนำ WEKA.ปทุมธานี: มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี.
- [7] The University of Waikato.WEKA.[ออนไลน์]. สืบค้นเมื่อ 25 พฤษภาคม 2557 , จาก<http://www.cs.waikato.ac.nz/ml/weka/>