

การหาความสัมพันธ์ของรายการสินค้าร้านกาแฟ โดยใช้เทคนิคกฎความสัมพันธ์

สุชาดา รักษาควร และ ปัทมา อินทรกำแหง

ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีและการจัดการอุตสาหกรรม
มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ วิทยาเขตปทุมธานี
Emails : 5606021610071@fitm.kmutnb.ac.th, 5606021620068@fitm.kmutnb.ac.th

บทคัดย่อ

การหาความสัมพันธ์ของรายการสินค้าโดยใช้เทคนิคกฎความสัมพันธ์ ได้มีการพัฒนาแอปพลิเคชันในการจัดเตรียมข้อมูลโดยใช้ภาษา PHP, HTML, CSS ฐานข้อมูลที่ใช้จัดการคือ SQL Server ซึ่งเป็นฐานข้อมูลที่บริษัทได้ใช้ในปัจจุบัน หลังจากนั้นได้นำข้อมูลมาทำการจัดกลุ่มข้อมูลด้วยอัลกอริทึม K-Mean ผ่านโปรแกรม SPSS เพื่อค้นหาลักษณะของข้อมูลที่มีความคล้ายคลึงกัน เมื่อได้กลุ่มข้อมูลแล้วนำข้อมูลแต่ละกลุ่มมาหาความสัมพันธ์เพื่อเป็นประโยชน์ต่อการวางแผนในการส่งเสริมการขายสินค้าของผู้บริหาร และต่อยอดทางการตลาดให้กับหน่วยงานในการผลิตสินค้าให้เพียงพอต่อความต้องการของผู้บริโภค

ABSTRACT

Discovering the Relations of Product Items using Association Rules Technique are developed using PHP, HTML, CSS and use SQL Server for Database management. Then, we use K-mean algorithm to cluster a data for finding a similar data's appearance with SPSS program. Moreover, we bring each group's data to associate the set of rules for finding the appropriate promotion. Moreover, the obtained rules can use to develop a marketing strategy and products can be enough for a customer demand.

คำสำคัญ-- เทคนิคกฎความสัมพันธ์; การจัดกลุ่มข้อมูล;
การทำเหมืองข้อมูล;

1. บทนำ

เนื่องจากในปัจจุบันธุรกิจร้านกาแฟในกรณีศึกษาได้มีการขยายกิจการและเพิ่มสาขาตามพื้นที่ต่าง ๆ เป็นจำนวนมาก และทางร้านได้มีการจำหน่ายเครื่องดื่มและขนม เช่น กาแฟ ขนมปัง บัฟเฟต์ ถั่วกรอบแก้ว เป็นต้น และมีแนวโน้มจะเพิ่มจำนวนของรายการ

สินค้าและจำนวนของลูกค้ามากขึ้น ซึ่งในปัจจุบันร้านกาแฟดังกล่าวยังไม่มียุทธศาสตร์ในการวิเคราะห์พฤติกรรมผู้บริโภค รวมถึงการหาความสัมพันธ์ของรายการสินค้า ดังนั้น ทางส่วนงานสนับสนุนข้อมูลที่จะช่วยในการตัดสินใจและจัดทำจึงได้มีแนวความคิดในการนำข้อมูลของรายการสินค้ามาดำเนินการจัดกลุ่มเพื่อวิเคราะห์หาความสัมพันธ์ของรายการสินค้า เพื่อช่วยสนับสนุนการตัดสินใจของผู้บริหารในการจัดรายการสินค้าเพื่อส่งเสริมการขาย รวมถึงส่วนงานอื่น ๆ ที่มีความเกี่ยวข้องสามารถนำข้อมูลดังกล่าวมาวิเคราะห์เพื่อต่อยอดทางการตลาดได้อย่างมีประสิทธิภาพ

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1. การจัดกลุ่ม (CLUSTERING) [1]

เป็นวิธีการที่พิจารณาข้อมูลแต่ละแถวเสมือนเป็นวัตถุ (OBJECT) ซึ่งจะมีหลักการเหมือนกับการจำแนกประเภทข้อมูล คือจะทำการแบ่งข้อมูลออกเป็นกลุ่ม (คลัสเตอร์) โดยจะจัดให้ข้อมูลที่มีความคล้ายคลึงกันอยู่ในคลัสเตอร์เดียวกัน และข้อมูลที่อยู่ต่างคลัสเตอร์กันจะมีความคล้ายคลึงกันน้อยที่สุดซึ่งความเหมือนหรือต่างกันสามารถเปรียบเทียบได้กับความใกล้ชิดกันของวัตถุใดๆ โดยใช้ระยะทางเป็นตัวชี้วัด คุณภาพของแต่ละคลัสเตอร์สามารถอธิบายได้จากเส้นผ่านศูนย์กลางของคลัสเตอร์ (DIAMETER) ซึ่งแสดงระยะห่างมากที่สุดของวัตถุสองชิ้นที่อยู่ในคลัสเตอร์เดียวกัน แต่ละคลัสเตอร์จะมีตัวแทนที่สามารถแทนวัตถุทุกชิ้นของคลัสเตอร์นั้นได้ เช่น การใช้จุดศูนย์กลางคลัสเตอร์ (Centroid) แทนคลัสเตอร์นั้น สำหรับบางเทคนิคตัวแทนของคลัสเตอร์อาจมีได้หลายตัวแทน ทั้งนี้ขึ้นอยู่กับความเหมาะสมของแต่ละเทคนิคที่เลือกใช้

2.2. Association Rules [2]

เป็นเทคนิคหนึ่งของ Data Mining ที่สำคัญและสามารถนำไปประยุกต์ใช้ได้จริงกับงานต่าง ๆ หลักการทำงานของวิธีการนี้คือการหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อ

นำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ หรือส่วนมากจะวิเคราะห์การซื้อสินค้าของลูกค้าเรียกว่า “Market Basket Analysis” ซึ่งประเมินจากข้อมูลในตารางที่รวบรวมไว้ ผลการวิเคราะห์ที่ได้จะเป็นคำตอบของปัญหา ซึ่งการวิเคราะห์แบบนี้เป็นการใช้ “กฎความสัมพันธ์” (Association Rules) เพื่อหาความสัมพันธ์ของข้อมูล

2.3. งานวิจัยที่เกี่ยวข้อง

การทำเหมืองข้อมูลกฎความสัมพันธ์ด้วยวิธี ค่าถ่วงน้ำหนักสูงสุดจากกราฟบริบูรณ์แบบสมมาตร [3] การทำเหมืองข้อมูลกฎความสัมพันธ์จากรายการซื้อขาย เพื่อใช้ในการทำการขายส่งเสริมการขายแนะนำ รายการ สินค้าที่มักจะซื้อด้วยกันบ่อย แต่การหาความสัมพันธ์ต้องอาศัยการกำหนดค่าสนับสนุนขั้นต่ำที่เหมาะสม เพราะถ้ากำหนดค่าน้อยไป จำนวนกฎที่ได้จะมากและถ้ากำหนดค่ามากไปจำนวนกฎที่ได้ก็จะน้อย ทำให้ผู้ขายยากจะหาว่าที่เหมาะสมได้ในแต่ละชุดข้อมูล จากปัญหาดังกล่าว ผู้วิจัย นำเสนอวิธีการใหม่ในการเก็บความถี่การขายในรูปแบบกราฟบริบูรณ์สมมาตร แบบมีทิศทางเพื่อนำค่าถ่วงน้ำหนักสูงสุดไป คำนวณหาความสัมพันธ์จากค่าสนับสนุนสูงสุดได้และใช้ได้กับทุกชุดข้อมูล

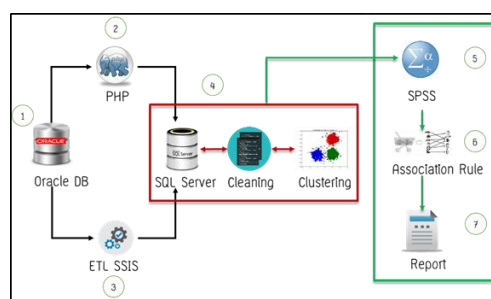
การสร้างแบบจำลองการขายผลิตภัณฑ์ประกันภัยให้กับลูกค้าผู้สูงอายุกลุ่มบัญชีออมทรัพย์ โดยการทำเหมืองข้อมูล [4] งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลอง ที่สามารถสนับสนุนกิจกรรมทางการตลาดระหว่างสององค์กร โดยใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) ตามกรอบการทำเหมืองข้อมูล CRISP-DM (Cross Industry Standard Process for Data Mining) โดยการสร้างแบบจำลองการแบ่งกลุ่ม (Clustering) ด้วยขั้นตอน วิธี Simple Means และสร้างแบบจำลองการหาความสัมพันธ์ (Association Rule) ด้วยขั้นตอนวิธี Apriori ซึ่งผลวิจัยที่ได้สามารถนำไปใช้สนับสนุนกิจกรรมทางการตลาดขององค์กรได้ เช่น การสร้างโปรโมชั่นหรือการออกผลิตภัณฑ์ใหม่ ที่เหมาะสมกับลูกค้าแต่ละกลุ่ม เพื่อให้เกิดความพึงพอใจสูงสุด และความภักดีต่อองค์กร รวมถึงการกระตุ้นการซื้อผลิตภัณฑ์ของลูกค้า

การวิเคราะห์ข้อมูลทางเศรษฐกิจและสังคมของครัวเรือน ด้วยวิธีการแบ่งกลุ่ม และหาความสัมพันธ์ สำหรับการหาเหมืองข้อมูล [5] งานวิจัยนี้เป็นการศึกษา และวิเคราะห์ข้อมูลการสำรวจภาวะเศรษฐกิจ และสังคมของครัวเรือน ประจำปี พ.ศ. 2552 โดยนำความรู้ทางด้านการทำเหมืองข้อมูลมาใช้เพื่อจัดกลุ่มข้อมูลตามการใช้จ่ายของ ประชากรโดยใช้อัลกอริทึมเคมีน สำหรับหากลุ่มของข้อมูลที่เหมาะสมจากค่า DB Index และ SD Validity Index พบว่าจำนวนกลุ่มที่เหมาะสมคือ 3 กลุ่ม จากนั้นใช้กฎความสัมพันธ์ในการหาความสัมพันธ์ของข้อมูล

ในแต่ละกลุ่ม เพื่อนำไปหาตัวแปรที่มีความสัมพันธ์กัน ผลการวิจัย พบว่า กฎความสัมพันธ์ที่ได้ในแต่ละกลุ่มมีลักษณะคล้ายคลึงกัน เช่น รายได้เฉลี่ยต่อเดือนของครัวเรือนมีความสัมพันธ์กับค่าใช้จ่ายเฉลี่ยต่อเดือนของครัวเรือน ขนาดของครัวเรือนมีความสัมพันธ์กับจำนวนผู้หารายได้และค่าใช้จ่ายยาสูบต่อเดือนของครัวเรือนมีความสัมพันธ์กับจำนวนสมาชิกที่มีสิทธิในการเบิกค่ารักษาพยาบาล เป็นต้น

3. ขั้นตอนและวิธีการดำเนินงาน

ในส่วนของขั้นตอนและวิธีการดำเนินงาน ผู้วิจัยได้ดำเนินการออกแบบกระบวนการทำงานทั้งหมด 7 ขั้นตอน ดังรูปที่ 1



รูปที่ 1. ภาพแสดงกระบวนการทำงานโดยรวมของระบบ

3.1. การพัฒนาและเลือกโปรแกรมที่ใช้ในการพัฒนา

จากภาพรวมกระบวนการทำงานของระบบ ผู้วิจัยได้แบ่งส่วนการทำงานออกเป็น 3 ส่วนหลัก ๆ เพื่อให้เห็นถึงการทำงานที่ชัดเจนและเป็นระบบ ดังนี้

ส่วนที่ 1 ส่วนของการจัดทำโปรแกรมเตรียมข้อมูลสำหรับการทำเหมืองข้อมูล (Data Mining)

ส่วนที่ 2 ส่วนของการทำเหมืองข้อมูล โดยใช้เทคนิคการจัดกลุ่มข้อมูล (Clustering Data)

ส่วนที่ 3 ส่วนของการทำเหมืองข้อมูล โดยใช้เทคนิคกฎความสัมพันธ์ (Association Rule)

โปรแกรมที่เลือกใช้ในการพัฒนาโปรแกรมคือ Sublime ซึ่งคุณสมบัติที่ทำให้เลือกนั้นได้กล่าวไปในข้างต้นแล้ว และใช้ภาษา PHP ในการพัฒนาโปรแกรม ซึ่งการออกแบบหน้าโปรแกรมนั้นจำเป็นต้องออกแบบให้การใช้งานนั้นเป็นไปตามความต้องการของผู้ใช้งาน รวมถึงการทำงานภายในนั้นจำเป็นต้องครบถ้วนตรงตามความต้องการของผู้ใช้และสามารถใช้งานได้อย่างคุ้มค่ากับเวลา และการพัฒนา ซึ่งการทำงานภายใต้ขั้นตอนนี้มีดังนี้

3.1.1. สำรวจความต้องการของผู้ใช้งาน การสำรวจความต้องการของผู้ใช้งานนั้นมีความจำเป็นอย่างยิ่งในการทำเพื่อให้ง่ายต่อการพัฒนาโปรแกรมที่ถูกต้อง และเพื่อให้โปรแกรมที่จะพัฒนาขึ้นนั้นมีความถูกต้อง และผิดพลาดน้อยที่สุด

3.1.2. ประมวลผลข้อมูลที่สำรวจจากผู้ใช้งาน ก่อนการพัฒนาโปรแกรมขึ้นมานั้นจะมีการทำการวิเคราะห์ข้อมูลที่ได้มาอีกครั้ง เพื่อลำดับความสำคัญในการพัฒนาโปรแกรมซึ่งจะทำให้การพัฒนาโปรแกรมนั้นดูมีความก้าวหน้าเหมาะสมกับเวลาในแต่ละช่วง

3.1.3. เมื่อออกแบบการทำงานของโปรแกรมตามส่วนที่ได้กล่าวไปข้างต้นรวมถึงขอคำปรึกษาและอธิบายกับผู้ใช้งานโปรแกรมแล้ว หลังจากนั้นเชื่อมต่อฐานข้อมูล Oracle เพื่อทำการดึงข้อมูลส่วนที่ต้องนำมาใช้ในการจัดทำวิจัยไปเก็บไว้ในส่วนของ SQL Server Management ของทางฝ่ายงานสนับสนุนข้อมูลหรือ MIS ซึ่งข้อมูลที่จะนำมาเก็บใน SQL Server นั้นจะผ่านการเตรียมข้อมูล (Preprocess Data) เพื่อพร้อมสำหรับการจัดกลุ่มข้อมูล (Clustering Data) ภายใต้โปรแกรมที่พัฒนาขึ้นนี้จะประกอบด้วยการทำงานดังแสดงในขั้นตอนต่อไป

3.2. ขั้นตอนการจัดกลุ่มข้อมูล (Clustering Data)

โปรแกรมที่ใช้ในการจัดกลุ่มข้อมูล (Clustering Data) เป็นโปรแกรม SPSS ซึ่งกล่าวไว้ในข้างต้นแล้ว โดยมีขั้นตอนสำหรับการจัดกลุ่มข้อมูลดังรูปที่ 2



รูปที่ 2. ขั้นตอนการจัดกลุ่มข้อมูล

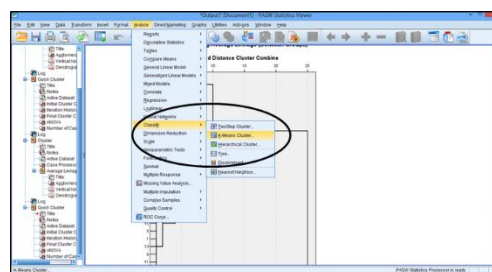
3.2.1. ศึกษาข้อมูลในระบบจัดการฐานข้อมูล (SQL Server Management) เพื่อให้ทราบว่าข้อมูลที่ต้องนำมาจัดทำวิจัยนั้นมีลักษณะเป็นอย่างไร โดยข้อมูลที่นำมาศึกษานั้น เป็นข้อมูลในส่วนข้อมูลรายการขายสินค้า และข้อมูลส่วนข้อมูลหลัก (Master Data) ซึ่งส่วนของข้อมูลหลัก (Master Data) เป็นข้อมูลสินค้าที่มีสำหรับจัดจำหน่ายภายในร้านกาแฟทุกสาขา ข้อมูลหลัก (Master Data) หรือข้อมูลสินค้าจะมีทั้งหมด 11 กลุ่ม 55 ประเภท ทั้งหมด 218 SKU

3.2.2. ทำการเตรียมข้อมูล (Preprocess Data) กระบวนการเตรียมข้อมูลจะมีการตรวจสอบสินค้าใหม่ที่มีโอกาสเพิ่มขึ้นในอนาคต (Mapping Data) เพื่อให้ได้ข้อมูลที่เหมาะสมสำหรับการจัดกลุ่มข้อมูล (Clustering Data) ซึ่งการทำงานในส่วนของการจัดกลุ่มข้อมูล จะมีการทำงานหลาย ๆ ครั้งเพื่อให้ได้ข้อมูลที่มีคุณสมบัติเหมาะสม แก่การหาความสัมพันธ์ (Association Rules) ข้อมูลที่จะนำมาใช้คือข้อมูลสินค้าที่มีทั้งหมดในร้านกาแฟทุกสินค้าซึ่งมีจำนวนทั้งหมด 218 SKU และข้อมูลในส่วนของการรายการการซื้อสินค้ามีประมาณ 200,000 แถว เพื่อ

นำไปวิเคราะห์หาความสัมพันธ์โดยผลที่จะได้จะมีผลครอบคลุมในทุกสาขาที่มีในปัจจุบัน

3.2.3. การจัดกลุ่มข้อมูล (Clustering Data) เป็นการนำเอาข้อมูลที่ผ่านการเตรียมข้อมูลแล้ว มาผ่านเทคนิคที่ใช้ในการจำแนกหรือแบ่งตัวแปรออกเป็นกลุ่มย่อย ๆ ตั้งแต่ 2 กลุ่มขึ้นไป ตัวแปรที่อยู่ในกลุ่มเดียวกันจะมีลักษณะคล้ายกันหรือเหมือนกัน ส่วนตัวแปรที่อยู่ต่างกลุ่มจะมีลักษณะที่แตกต่างกัน ดังนั้น การพิจารณาเลือกลักษณะหรือตัวแปรที่จะนำมาใช้ในการแบ่งกลุ่มจึงมีความสำคัญมาก ซึ่งจากงานวิจัยนี้ การจัดกลุ่มข้อมูลจะมีประโยชน์ในการวิเคราะห์ถึงตัวแปรที่มีอิทธิพลต่อยอดขายของร้านกาแฟ เทคนิคที่ใช้ในการทำ Clustering Analysis คือ K-mean Clustering

3.2.4. ขั้นตอนการจัดกลุ่มข้อมูล (Clustering Data) โดย โปรแกรม SPSS



รูปที่ 3. ทำ K-Means

ผลจากการทำ Clustering โดยใช้เทคนิค K-Means

| Number of Cases in each Cluster | | |
|---------------------------------|--|--------|
| Cluster | | |
| 1 | | 15.000 |
| 2 | | 2.000 |
| 3 | | 10.000 |
| 4 | | 2.000 |
| Valid | | 29.000 |
| Missing | | .000 |

รูปที่ 4. Number of Cases in each Cluster

จากรูปที่ 4 แสดงจำนวนตัวแปร หรือ สาขาของร้านกาแฟที่อยู่ในแต่ละ Cluster จะพบว่าสาขาส่วนใหญ่จะอยู่ใน Cluster ที่ 1 ซึ่งมีจำนวน 15 สาขา ส่วน Cluster ที่ 2 และ 4 มีจำนวนสาขาเท่ากันและน้อยที่สุด ซึ่งมีจำนวน 2 สาขา

3.2.5. ทดสอบ Model

| ANOVA | | | | | | |
|----------------|-------------|----|-------------|----|--------|------|
| | Cluster | | Error | | F | Sig. |
| | Mean Square | df | Mean Square | df | | |
| Coffee | 3.953E7 | 3 | 784457.060 | 25 | 50.397 | .000 |
| NonCoffee | 5725466.320 | 3 | 201277.553 | 25 | 28.446 | .000 |
| PromotionDrink | 6115.802 | 3 | 1939.869 | 25 | 3.153 | .043 |
| Food | 2945178.142 | 3 | 181953.293 | 25 | 16.186 | .000 |
| ReadyToDrink | 19069.144 | 3 | 2928.141 | 25 | 6.512 | .002 |
| Sweets | 1366704.768 | 3 | 55510.073 | 25 | 24.621 | .000 |
| Other | 38860.934 | 3 | 4526.329 | 25 | 8.586 | .000 |

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

รูปที่ 5. ตาราง ANOVA

จากตารางวัดค่าความถูกต้องโดยใช้วิธีการหาค่ากลางของความคลาดเคลื่อนกำลังสอง (Mean Square Error : MSE) พิจารณาจากสินค้าที่มีความสำคัญมากที่สุด 3 อันดับ ได้แก่อันดับที่ 1 คือ Coffee ซึ่งจากตาราง ANOVA มีความน่าเชื่อถือน้อยกว่า Non Coffee โดยมีค่าความคลาดเคลื่อนอยู่ที่ 784457.060 และอันดับที่ 2 Non Coffee มีความน่าเชื่อถือน้อยกว่า Promotion Drink โดยมีค่าความคลาดเคลื่อนอยู่ที่ 201277.533 และอันดับสุดท้าย Promotion Drink มีความน่าเชื่อถือมากที่สุด โดยมีค่าความคลาดเคลื่อนอยู่ที่ 1939.869

3.2.6. ออกรายงาน

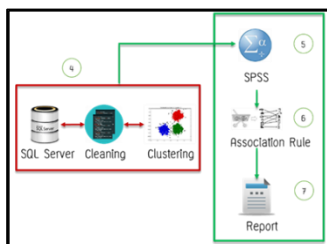
เป็นการจัดทำ Data Virtualization เพื่อแสดงรายงานการเปรียบเทียบยอดขายแต่ละกลุ่มสินค้า



รูปที่ 6. การออกรายงานเปรียบเทียบยอดขายแต่ละกลุ่มจากผลรวมค่าเฉลี่ยกลุ่มสินค้า

3.3. ขั้นตอนการหากฎความสัมพันธ์ (Association Rules)

โปรแกรมที่ใช้ในการหากฎความสัมพันธ์ (Association Rules) เป็นโปรแกรม SPSS โดยมีขั้นตอนในการหากฎความสัมพันธ์ดังรูปที่ 7



รูปที่ 7. การหาความสัมพันธ์ (Association Rules)

การหาความสัมพันธ์นี้เป็นการหาความสัมพันธ์ระหว่างสินค้าภายในร้านกาแฟ ซึ่งแบ่งตามกลุ่มของสาขาที่ได้จากการแบ่งกลุ่ม ดังนั้นจึงต้องมีการหาความสัมพันธ์ของสินค้าทั้งหมด 4 ครั้ง

3.3.1. ขั้นตอนในการเตรียมข้อมูล กลุ่มสาขาที่ทำการจัดกลุ่มข้อมูล (Clustering Data) คือกลุ่มสาขาที่มีลักษณะคล้ายกันในที่นี้คือมีจำนวนสินค้าแต่ละชนิดที่ขายได้ในปริมาณที่ใกล้เคียงกัน เมื่อพิจารณาจากข้อมูลดังกล่าวแล้วผู้จัดทำได้นำข้อมูลการขายสินค้าของแต่ละกลุ่มมาวิเคราะห์ต่อเพื่อให้ได้ข้อมูลที่เหมาะสำหรับการหาความสัมพันธ์ (Association

Rules) มากที่สุด โดยพิจารณาเลือกเฉพาะสินค้าที่มีความสำคัญต่อการวิเคราะห์จริง ๆ

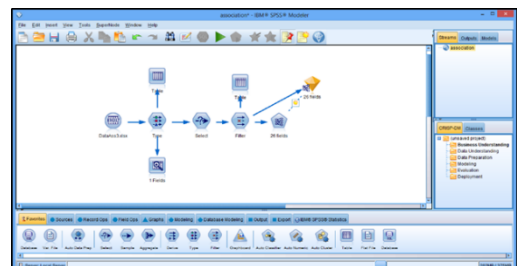
3.3.2. การคัดเลือกข้อมูล (Data Selection) เราจำเป็นต้องคัดเลือกเฉพาะข้อมูลสินค้าที่สามารถนำมาใช้ประโยชน์ได้ ซึ่งจากข้อมูลที่มีนั้นบางรายการ หรือ บางสินค้าจะไม่มีผลต่อการนำไปวิเคราะห์ต่อในส่วนของการหาความสัมพันธ์ จึงจำเป็นต้องตัดออก ซึ่งมีดังนี้

3.3.2.1. รายการการซื้อสินค้า ที่มีการส่งสินค้าแค่เมนูเดียว เช่น 1 ใบเสร็จซื้อขนมกรอบแก้ว 1 ถ้วย

3.3.2.2. สินค้าที่มีจำนวนยอดขายน้อย 450 ขึ้น ซึ่งแสดงให้เห็นว่าสินค้านั้นไม่ค่อยมีความสำคัญต่อยอดขาย

3.3.2.3. ข้อมูลที่เตรียมจะมีคอลัมน์ (Attribute) คือ เลขที่บิล (Billing_No), ชื่อสาขา (Branch_Name), เลขที่กลุ่ม (Cluster) และ คอลัมน์ชื่อสินค้า

3.3.3. ขั้นตอนการหาความสัมพันธ์ (Association Rule) โดยโปรแกรม SPSS Modeler



รูปที่ 8. ขั้นตอนการทำ Association Rules

4. ผลการดำเนินงาน

จากการดำเนินงานในการจัดทำวิจัยครั้งนี้ผลการดำเนินงานแบ่งเป็น 3 ส่วนหลัก ๆ ดังนี้

4.1. ผลจากการพัฒนาโปรแกรมเตรียมข้อมูล

ในส่วนของงานนี้ได้วางแผนงานไว้เป็นการพัฒนาโปรแกรมสำหรับเตรียมข้อมูลในการทำเหมืองข้อมูล (Data Mining) จากข้อมูลร้านกาแฟ ซึ่งจากการพัฒนาโปรแกรมในครั้งนี้ภายในจะมีการทำงาน 5 ส่วนหลัก ๆ ดังนี้

4.1.1. ส่วนของการดึงข้อมูลจากฐานข้อมูล Oracle มาเก็บไว้ใน SQL Server ของทางส่วนงานสนับสนุนข้อมูล หรือ MIS เพื่อให้ข้อมูลสำหรับนำมาทำการจัดกลุ่มข้อมูล (Clustering Data) นั้นมีความเป็นปัจจุบัน

4.1.2. ส่วนของการดูข้อมูลรายการการขายสินค้าของทางร้านกาแฟที่ผ่านการแปลงข้อมูลให้อยู่ในรูปแบบเดียวกัน (Mapping Data) เป็นที่เรียบร้อยแล้ว พร้อมทั้งจะนำไปทำการจัดกลุ่มข้อมูล (Clustering Data)

4.1.3. ส่วนของการแปลงข้อมูลให้อยู่ในรูปแบบเดียวกัน (Mapping Data) เป็นการจัดรูปแบบข้อมูลรายการสินค้านั้นให้อยู่ในรูปแบบที่ต้องการเพื่อให้ง่ายต่อการทำการจัดกลุ่มข้อมูล

4.1.4. ส่วนของการเตรียมข้อมูลในการหาความสัมพันธ์ (Association Rules)

4.1.5. ส่วนของการดูประวัติการแปลงข้อมูลให้อยู่ในรูปแบบเดียวกัน (Mapping Data) เป็นการแจ้งรายละเอียดข้อมูลเบื้องต้น เช่น วันเวลาบุคคลที่ทำการเพิ่มข้อมูล

ซึ่งในการพัฒนาโปรแกรมในส่วนนี้ได้พัฒนาด้วยภาษา PHP และมีการใช้ภาษา CSS ในการจัดรูปแบบโปรแกรมให้มีความสวยงามและมีความเป็นระเบียบง่ายต่อการจัดส่วน และการวางตำแหน่งความสำคัญของกระบวนการทำงานในส่วนต่าง ๆ ผลที่ได้เป็นไปตามที่ผู้ใช้งานต้องการง่ายต่อการใช้งานการทำงานมีความครบถ้วน และถูกต้อง

รูปที่ 9. แสดงผลลัพธ์ในการเตรียมข้อมูลสำหรับการหาความสัมพันธ์

4.2. ผลจากการทำการจัดกลุ่มข้อมูล (Clustering Data)

| ANOVA | | | | | |
|----------------|-------------|----|-------------|----|--------|
| | Cluster | | Error | | |
| | Mean Square | df | Mean Square | df | F |
| Coffee | 3.953E7 | 3 | 784457.060 | 25 | 50.397 |
| NonCoffee | 5725466.320 | 3 | 201277.553 | 25 | 28.446 |
| PromotionDrink | 6115.802 | 3 | 1939.869 | 25 | 3.153 |
| Food | 2945178.142 | 3 | 181953.293 | 25 | 16.186 |
| ReadyToDrink | 19069.144 | 3 | 2928.141 | 25 | 6.512 |
| Sweets | 1366704.768 | 3 | 55510.073 | 25 | 24.621 |
| Other | 38860.934 | 3 | 4526.329 | 25 | 8.586 |
| | | | | | Sig. |
| | | | | | .000 |
| | | | | | .000 |
| | | | | | .043 |
| | | | | | .000 |
| | | | | | .002 |
| | | | | | .000 |
| | | | | | .000 |

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

รูปที่ 10. ตาราง ANOVA

จากตาราง ANOVA พิจารณาว่าค่า F พบว่ากลุ่มสินค้าที่มีความสำคัญต่อการแบ่งกลุ่มสาขาร้านค้ากาแฟเป็นอันดับ 1 คือ Coffee อันดับ 2 คือ Non Coffee และอันดับ 3 คือ Sweets และอื่น ๆ ตามลำดับ ซึ่งส่งผลให้เกิดการแบ่งกลุ่มข้อมูลออกเป็น 4 กลุ่ม

| Cluster Number of Case | Coffee | Promotion Drink | NonCoffee | Food | Sweets | ReadyToDrink | Other |
|------------------------|---------|-----------------|-----------|---------|---------|--------------|--------|
| 1 | Mean | 2950.6 | 58.33 | 1887.53 | 1158.87 | 326.23 | 137.73 |
| | N | 15 | 15 | 15 | 15 | 15 | 15 |
| | Sum | 59259 | 875 | 28313 | 17383 | 5030 | 2066 |
| | Minimum | 2740 | 13 | 1103 | 644 | 0 | 63 |
| | Maximum | 5072 | 97 | 2475 | 1782 | 757 | 335 |
| 2 | Mean | 6370.00 | 88.50 | 2921.50 | 2216.00 | 296.00 | 218.00 |
| | N | 2 | 2 | 2 | 2 | 2 | 2 |
| | Sum | 12740 | 177 | 5843 | 4432 | 592 | 436 |
| | Minimum | 6225 | 62 | 2421 | 1707 | 61 | 213 |
| | Maximum | 6515 | 115 | 3422 | 2725 | 531 | 223 |
| 3 | Mean | 1553.90 | 27.10 | 865.30 | 443.30 | 87.50 | 66.70 |
| | N | 10 | 10 | 10 | 10 | 10 | 10 |
| | Sum | 15539 | 271 | 8653 | 4433 | 875 | 667 |
| | Minimum | 126 | 0 | 34 | 14 | 6 | 0 |
| | Maximum | 2755 | 102 | 1684 | 987 | 363 | 191 |
| 4 | Mean | 9059.00 | 119.00 | 3531.50 | 2112.50 | 1655.00 | 169.50 |
| | N | 2 | 2 | 2 | 2 | 2 | 2 |
| | Sum | 18118 | 238 | 7063 | 4225 | 3310 | 339 |
| | Minimum | 8795 | 0 | 3200 | 1293 | 1399 | 162 |
| | Maximum | 9323 | 238 | 3863 | 2932 | 1911 | 177 |
| Total | Mean | 3643.31 | 63.83 | 1710.72 | 1050.79 | 336.17 | 120.07 |
| | N | 29 | 29 | 29 | 29 | 29 | 29 |
| | Sum | 105656 | 1561 | 49872 | 30473 | 9807 | 3508 |
| | Minimum | 126 | 0 | 34 | 14 | 6 | 0 |
| | Maximum | 9323 | 238 | 3863 | 2932 | 1911 | 177 |

รูปที่ 11. ตาราง Crosstab

จากตาราง Crosstab เมื่อพิจารณาผลรวมค่าเฉลี่ยยอดขายของกลุ่มสินค้าแต่ละกลุ่ม แสดงถึงลักษณะของกลุ่มข้อมูลกลุ่มต่าง ๆ ได้ดังนี้

กลุ่มที่ 1 จัดเป็นกลุ่มที่ขายสินค้าโดยรวมอยู่ในระดับปานกลาง
กลุ่มที่ 2 จัดเป็นกลุ่มที่ขายสินค้าโดยรวมอยู่ในระดับดี
กลุ่มที่ 3 จัดเป็นกลุ่มที่ขายสินค้าโดยรวมอยู่ในระดับน้อย
กลุ่มที่ 4 จัดเป็นกลุ่มที่ขายสินค้าโดยรวมอยู่ในระดับดีมาก



รูปที่ 12. แสดงการเปรียบเทียบยอดขายแต่ละกลุ่มจากผลรวมค่าเฉลี่ยกลุ่มสินค้า

4.3. ผลจากการหาความสัมพันธ์ (Association Rules)

จากการหาความสัมพันธ์ด้วยโปรแกรม SPSS Modeler ได้กฎความสัมพันธ์ดังนี้

| กลุ่มสาขา | กฎความสัมพันธ์ | Sup % | Con % | Lift |
|-----------|--|-------|-------|------|
| Cluster 1 | Iced Espresso --> Iced Chocolate | 12.66 | 25.30 | 0.84 |
| | Iced Espresso --> Iced Americano | 12.39 | 24.46 | 0.82 |
| | Iced Espresso --> Iced Cappuccino | 23.95 | 23.11 | 0.77 |
| | Iced Espresso --> Iced Green Tea Latte | 16.09 | 22.50 | 0.75 |
| | Iced P Coffee --> Iced P Lovers Tea | 12.05 | 22.48 | 1.24 |
| | Iced Espresso --> Iced Latte | 18.05 | 22.31 | 0.74 |
| Cluster 2 | Iced Espresso --> Iced Mocha | 10.63 | 20.94 | 0.70 |
| | Iced Cappuccino --> Iced Latte | 19.71 | 23.85 | 0.99 |
| | Iced Espresso --> Iced Americano | 15.39 | 22.97 | 0.83 |
| | Iced Cappuccino --> Iced Green Tea Latte | 15.83 | 22.89 | 0.95 |
| | Iced P Coffee --> Iced P Lovers Tea | 10.35 | 22.50 | 1.32 |
| | Iced Espresso --> Iced Chocolate | 15.39 | 21.57 | 0.78 |
| Cluster 3 | Iced Espresso --> Iced Cappuccino | 23.98 | 21.40 | 0.77 |
| | Iced Espresso --> Iced Green Tea Latte | 15.83 | 20.98 | 0.76 |
| | Iced Latte --> Iced Chocolate | 15.39 | 20.73 | 1.05 |
| | Iced Cappuccino --> Iced Mocha | 14.37 | 28.81 | 0.97 |
| Cluster 4 | Iced Cappuccino --> Iced Latte | 17.55 | 26.30 | 0.88 |
| | Iced Espresso --> Iced Chocolate | 14.56 | 34.64 | 1.09 |

รูปที่ 13. ตารางกฎความสัมพันธ์ของข้อมูล

กฎที่ได้มีทั้งหมด 18 กฎ โดยที่ กลุ่มที่ 1 ได้กฎออกมา 7 กฎ, กลุ่มที่ 2 ได้กฎออกมา 8 กฎ, กลุ่มที่ 3 ได้กฎออกมา 2 กฎ และกลุ่มที่ 4 ได้กฎออกมา 1 กฎ แต่กฎที่สามารถนำไปใช้

ประโยชน์ได้จริง มีทั้งหมด 4 กฎ คือ กฎที่มีค่า Lift หรือค่าวัดความสัมพันธ์ของข้อมูล มากกว่า 1 ซึ่งบ่งบอกได้ว่าสินค้า 2 สิ่งส่งเสริมกันจริง ส่วนกฎที่มี ค่า Lift หรือค่าวัดความสัมพันธ์ของข้อมูล น้อยกว่า 1 ซึ่งบ่งบอกได้ว่าสินค้า 2 สิ่งไม่ส่งเสริมกันจริง และกฎที่มี ค่า Lift หรือค่าวัดความสัมพันธ์ของข้อมูล เท่ากับ 1 ซึ่งบ่งบอกได้ว่าสินค้า 2 สิ่ง ไม่ได้มีความเกี่ยวข้องสัมพันธ์กันเลย เราจะไม่นำกฎเหล่านั้นไปใช้ประโยชน์ในงานวิจัย

5. สรุปผลการดำเนินงาน

ในการจัดทำวิจัยนี้สามารถจัดกลุ่มสาขาร้านค้ากาแฟได้เป็น 4 กลุ่ม ซึ่งแบ่งกลุ่มตามระดับยอดขายของทางร้านจาก 4 กลุ่มดังกล่าว กลุ่มที่ 4 ซึ่งมีจำนวน 2 สาขาเป็นกลุ่มที่มียอดขายสูงสุดสามารถอธิบายได้ว่าสาขา ร้านกาแฟกลุ่มนี้เป็นกลุ่มที่ขายดีมาก และกลุ่มที่ 3 ซึ่งมีจำนวน 10 สาขาเป็นกลุ่มที่มียอดขายน้อยที่สุดสามารถอธิบายได้ว่ากลุ่มนี้เป็นกลุ่มที่ขายไม่ดี และกลุ่มที่ 2 ซึ่งมีจำนวน 2 สาขาเป็นกลุ่มที่ขายดี และกลุ่มที่ 1 ซึ่งมีจำนวน 15 สาขาเป็นกลุ่มที่ขายปานกลาง เมื่อได้ข้อมูลกลุ่มที่ผ่านการทำการจัดกลุ่มข้อมูล (Clustering Data) แล้วได้นำข้อมูลการขายสินค้าของแต่ละกลุ่มสาขามาหากฎความสัมพันธ์ (Association Rules) ต่อ มีกฎที่มีความน่าเชื่อถือทั้งหมด 4 กฎจากทั้งหมด 18 กฎที่บ่งบอกได้ว่าสินค้า 2 สิ่งส่งเสริมกันจริงดังนี้

| กลุ่มสาขา | กฎความสัมพันธ์ | Sup % | Con % | Lift |
|-----------|------------------------------------|-------|-------|------|
| Cluster 1 | Iced P Coffee -->Iced P Lovers Tea | 12.05 | 22.48 | 1.24 |
| Cluster 2 | Iced P Coffee -->Iced P Lovers Tea | 10.35 | 22.50 | 1.32 |
| | Iced Latte -->Iced Chocolate | 15.39 | 20.73 | 1.05 |
| Cluster 4 | Iced Espresso -->Iced Chocolate | 14.56 | 34.64 | 1.09 |

รูปที่ 14. ตารางกฎความสัมพันธ์ของข้อมูล

ซึ่งจากตารางดังกล่าว

กฎที่ 1 สามารถอธิบายได้ว่าลูกค้าที่ซื้อ Iced P Lovers Tea แล้วจะซื้อ Iced P Coffee ด้วย คิดเป็นค่าความเชื่อมั่น (Confidence) 22.48% โดยที่คิดเป็นค่าสนับสนุน (Support) 12.05% ของจำนวนรายการทั้งหมด ซึ่งวัดค่าความสัมพันธ์ของสินค้าได้ 1.24 ซึ่งแสดงว่าการที่ซื้อ Iced P Lovers Tea ส่งเสริมให้เกิดการซื้อ Iced P Coffee จริง เนื่องจากการลงพื้นที่จริงเพื่อสังเกตพฤติกรรมลูกค้าทำให้เห็นว่าส่วนใหญ่แล้วลูกค้าซื้อ 2 คนจะเป็นผู้หญิง และผู้ชายมาคู่กัน ซึ่งผู้หญิงจะบริโภค Iced P Lovers Tea เนื่องจากเป็นเมนูที่ประกอบด้วยชาเย็น และชาเขียวและนมซึ่งเป็นเมนูที่เป็นสูตรเฉพาะของทางร้าน ส่วนผู้ชายบริโภค Iced P Coffee ซึ่งเป็นกาแฟสูตรเฉพาะของทางร้านเช่นกันจะมีรสชาติเข้ม

กฎที่ 2 สามารถอธิบายได้ว่าลูกค้าที่ซื้อ Iced P Lovers Tea แล้วจะซื้อ Iced P Coffee ด้วย คิดเป็นค่าความเชื่อมั่น (Confidence) 22.50% โดยที่คิดเป็นค่าสนับสนุน (Support) 10.35% ของจำนวนรายการทั้งหมด ซึ่งวัดค่า

ความสัมพันธ์ของสินค้าได้ 1.32 ซึ่งแสดงว่าการที่ซื้อ Iced P Lovers Tea ส่งเสริมให้เกิดการซื้อ Iced P Coffee จริง เนื่องจากการลงพื้นที่จริงเพื่อสังเกตพฤติกรรมลูกค้าทำให้เห็นว่าส่วนใหญ่แล้วลูกค้าซื้อ 2 คนจะเป็นผู้หญิง และผู้ชายมาคู่กัน ซึ่งผู้หญิงจะบริโภค Iced P Lovers Tea เนื่องจากเป็นเมนูที่ประกอบด้วยชาเย็น และชาเขียวและนมซึ่งเป็นเมนูที่เป็นสูตรเฉพาะของทางร้าน ส่วนผู้ชายบริโภค Iced P Coffee ซึ่งเป็นกาแฟสูตรเฉพาะของทางร้านเช่นกันจะมีรสชาติเข้ม

กฎที่ 3 สามารถอธิบายได้ว่าลูกค้าที่ซื้อ Iced Chocolate แล้วจะซื้อ Iced Latte ด้วย คิดเป็นค่าความเชื่อมั่น (Confidence) 20.73% โดยที่คิดเป็นค่าสนับสนุน (Support) 15.39% ของจำนวนรายการทั้งหมด ซึ่งวัดค่าความสัมพันธ์ของสินค้าได้ 1.05 ซึ่งแสดงว่าการที่ซื้อ Iced Chocolate ส่งเสริมให้เกิดการซื้อ Iced Latte จริง เนื่องจากการลงพื้นที่จริงเพื่อสังเกตพฤติกรรมลูกค้าทำให้เห็นว่าส่วนใหญ่แล้วลูกค้าซื้อ 2 คนจะเป็นผู้หญิง และผู้ชายมาคู่กัน ซึ่งผู้หญิงจะบริโภค Iced Chocolate ส่วนผู้ชายบริโภค Iced Latte

กฎที่ 4 สามารถอธิบายได้ว่าลูกค้าที่ซื้อ Iced Chocolate แล้วจะซื้อ Iced Espresso ด้วย คิดเป็นค่าความเชื่อมั่น (Confidence) 34.64% โดยที่คิดเป็นค่าสนับสนุน (Support) 14.56% ของจำนวนรายการทั้งหมด ซึ่งวัดค่าความสัมพันธ์ของสินค้าได้ 1.09 ซึ่งแสดงว่าการที่ซื้อ Iced Chocolate ส่งเสริมให้เกิดการซื้อ Iced Espresso จริง เนื่องจากการลงพื้นที่จริงเพื่อสังเกตพฤติกรรมลูกค้าทำให้เห็นว่าส่วนใหญ่แล้วลูกค้าซื้อ 2 คนจะเป็นผู้หญิง และผู้ชายมาคู่กัน ซึ่งผู้หญิงจะบริโภค Iced Chocolate ส่วนผู้ชายบริโภค Iced Espresso ซึ่งเป็นกาแฟรสชาติเข้ม

ซึ่งผลที่ได้จากการทำเหมืองข้อมูล สามารถนำมาเป็นส่วนหนึ่งในการช่วยตัดสินใจของผู้บริหารและอาจนำมาประกอบแนวคิดในการวางแผนส่งเสริมการขายของทางร้าน

กิตติกรรมประกาศ

งานวิจัยเรื่อง “การหาความสัมพันธ์ของรายการสินค้าร้านกาแฟ โดยใช้เทคนิคกฎความสัมพันธ์” สำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับคำแนะนำจากท่านอาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.วันทนี ประจวบศุภกิจ ที่คอยติดตามและให้คำแนะนำในการทำงานวิจัยในครั้งนี้ และขอขอบคุณ คุณชวรงค์ พิทยานุรักษ์ ผู้จัดการแผนกสนับสนุนข้อมูล ที่ได้ให้คำแนะนำ ให้คำปรึกษา คอยติดตาม ให้ความช่วยเหลือ และให้ความรู้ถึงข้อมูลที่ใช้ในการทำงานวิจัยในครั้งนี้

เอกสารอ้างอิง

- [1] สมโภชน์ ศรีสมุทร. การจัดกลุ่ม (Clustering). [ออนไลน์] 2557.] สืบค้นวันที่ 22 กรกฎาคม 2559 .[จาก <http://www.saruthipong.com/port/document/299-705/299-705-10.pdf>
- [2] DATAMININGTREND . Association Rules. [ออนไลน์] 2557.] สืบค้นวันที่ 22 กรกฎาคม 2559 .[จาก <http://dataminingtrend.com/2014/association-rules>
- [3] ประมูล สุขสกาฬวงศ์ และพญ. มีสัจ. **การทำเหมืองข้อมูล กฎความสัมพันธ์ด้วยวิธีค่าถ่วงน้ำหนักสูงสุด จากกราฟ บริบูรณ์แบบสมมาตร**. คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ. [บทคัดย่อออนไลน์] 2559. [สืบค้นวันที่ 24 กรกฎาคม 2559]. จาก http://202.44.34.144/nccitedoc/admin/nccit_files/NCCI-T-20150810113437.pdf
- [4] นางสาวกฤษณพร สุริยะบรรเทิง และดร.กมล เกียรติเรือง. **กมล. การสร้างแบบจำลองการขายผลิตภัณฑ์ประกันภัย ให้กับลูกค้าผู้สูงอายุกลุ่มบัญชีออมทรัพย์โดยการทำเหมืองข้อมูล**. วิทยาลัยนวัตกรรมการบริหาร มหาวิทยาลัยธรรมศาสตร์ เขตพระนคร กรุงเทพฯ.[บทคัดย่อออนไลน์] 2559. [สืบค้นวันที่ 24 กรกฎาคม 2559]. จาก <http://www.citu.tu.ac.th/uploads/research/file/201410/08th-cdhltuyz.013pdf>
- [5] อาจารย์ชฎารัตน์ พิพัฒน์นันท์, อาจารย์สิริธร เจริญรัตน์ และนางสาวศศิธร มงคลศรีพัฒนา. **การวิเคราะห์ข้อมูลทางเศรษฐกิจและสังคมของครัวเรือน ด้วยวิธีการแบ่งกลุ่ม และหาความสัมพันธ์ สำหรับการทำการเหมืองข้อมูล**. มหาวิทยาลัยหอการค้าไทย [บทคัดย่อออนไลน์] . 2556. [สืบค้นวันที่ 24 กรกฎาคม 2559]. จาก <http://eprints.utcc.ac.th/1638/3//1638summary.pdf>