

# การวิเคราะห์การซื้อสินค้ากลุ่มผลิตภัณฑ์อาหารเสริม เครื่องสำอาง และอุปโภคบริโภค สำหรับธุรกิจขายตรง โดยวิธีกฎความสัมพันธ์ Purchasing Prediction of The Health Food, Cosmetic and Consumer Products For MLM Business by Association Rules

สิทธิชัย วรโชติกำจร<sup>1</sup> และ จรรย์ แสงราช<sup>2</sup>

Sittichai Worachotekamjorn and Charun Sanrach

## บทคัดย่อ

บทความวิจัยนี้เป็นการประยุกต์ใช้แนวคิดเรื่องการทำเหมืองข้อมูลมาวิเคราะห์พฤติกรรมของผู้บริโภคที่เป็นสมาชิกในระบบธุรกิจ MLM (Multi-Level Marketing) โดยเน้นที่ผลิตภัณฑ์ใน 3 กลุ่มหลักคือ กลุ่มอาหารเสริม กลุ่มเครื่องสำอาง และกลุ่มอุปโภคบริโภค เพื่อหาความสัมพันธ์ของการเลือกซื้อผลิตภัณฑ์ของสมาชิก โดยเน้นเฉพาะกลุ่มสมาชิกที่ดูแลเครือข่ายสายงานและยอดขายหลัก (แม่ทีม/ออฟไลน์) ซึ่งประกอบไปด้วยรายการใบสั่งซื้อรายการสินค้า โดยผลที่ได้จะมีประโยชน์ต่อการวางแผนในการส่งเสริมการขายสินค้า และผลิตสินค้าให้เพียงพอต่อความต้องการ ซึ่งฐานข้อมูลที่น่ามาใช้ในการวิเคราะห์ที่อยู่ในช่วงเดือน มกราคม ถึงธันวาคม ปี พ.ศ. 2548 จำนวนทั้งสิ้น

24,879 รายการ และใช้โปรแกรม WEKA สำหรับวิเคราะห์ในรูปแบบกฎความสัมพันธ์ (Association Rules) ด้วยวิธีอัลกอริทึม Apriori ซึ่งพบกฎความสัมพันธ์ทั้งสิ้น 6 กฎ โดยมีความเชื่อมั่น 61, 60, 59, 58, 54, 50 เปอร์เซนต์ตามลำดับ

**คำสำคัญ :** ผลิตภัณฑ์อาหารเสริม / ธุรกิจขายตรง / กฎความสัมพันธ์

## Abstract

The purpose of this research is using the data mining concept to analyze the consumer behavior of members in multi-level marketing (MLM) businesses by

<sup>1</sup> นักศึกษาปริญญาเอก สาขาวิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

<sup>2</sup> ผู้ช่วยศาสตราจารย์ สาขาภาควิชาคอมพิวเตอร์ศึกษา คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

focusing only three main groups, namely health food, cosmetic and consumer groups, to find the association between purchasing order list and items available from regular purchasing patterns of the members. The result is expected to be used for product promotion and adequate production. Data were analyzed from member database during January to December 2005 which included 24,879 items. Association rule was analyzed using WEKA program by the A-priori algorithm method. Six association rules were found at the confident level of 61, 60, 59, 58, 54 and 50 percent, respectively.

**Keywords :** Health food / MLM business / Association rules

## บทนำ

ปัจจุบันสภาวะตลาดการขายตรงยังคงมีการแข่งขันกันสูงและมีสินค้าหลากหลายทั้งสินค้าอุปโภค บริโภค เครื่องสำอาง ผลิตภัณฑ์อาหารเสริม เครื่องใช้ในครัวเรือน กลุ่มสินค้าเครื่องใช้สำนักงาน เป็นต้น แต่ละธุรกิจที่มีระบบการขายตรงต่างมุ่งสร้างความแตกต่างทั้งคุณภาพสินค้า แผนการตลาด ผลประโยชน์ ความก้าวหน้าในตำแหน่งหรือในสายงานของธุรกิจขายตรง เพื่อสร้างความน่าสนใจและดึงดูดผู้คนทั่วไปที่ต้องการมีรายได้มากกว่าอาชีพหลักที่ทำอยู่ให้มีความเป็นอยู่ที่ดีขึ้น ดังนั้นธุรกิจที่มีระบบการขายตรงจำเป็นที่จะต้องสื่อสาร หรือสร้างความเข้าใจตลอดจนสร้างภาพลักษณ์ที่ชัดเจนให้ผู้บริโภคได้เข้าใจถึงการขายตรงและสามารถทำให้ผู้บริโภคมากมายได้เปลี่ยนทิศทางการเป็นผู้ซื้อเป็นผู้ขายเพื่อเป้าหมายความสำเร็จซึ่งในส่วนผู้บริโภคที่ซื้อสินค้าในระบบขายตรงนั้นจะถูกเรียกว่าสมาชิก ซึ่งจะมีทั้งสมาชิกทั่วไป และสมาชิกที่ดูแลการขายเครือข่ายสมาชิก และดูแลยอดขายหลักในสายงานนั้นๆ ของตัวเอง (แม่ทีม/ออฟไลน์) เป็นส่วนสำคัญในการเพิ่มยอดขายให้กับบริษัท ดังนั้นผู้ประกอบการธุรกิจขายตรงจำเป็นต้องวิเคราะห์ข้อมูลต่างๆ เหล่านี้ เพื่อพัฒนา และผลิตสินค้าให้เหมาะสมกับสมาชิกมากที่สุด

สำหรับเทคนิคในการวิเคราะห์ข้อมูลจากข้อมูลการซื้อขายสินค้านั้น จะใช้เทคนิคการทำเหมืองข้อมูลซึ่งนิยมนำมาใช้ในธุรกิจขนาดใหญ่ในปัจจุบันอย่างแพร่หลาย เช่น ธุรกิจซูเปอร์มาร์เก็ต, ห้างสรรพสินค้า เป็นต้น การที่มีผู้คนมากมายเลือกซื้อสินค้าที่หลากหลาย ผู้ประกอบการจะสามารถเพิ่มยอดขายในสินค้านั้นๆ จะต้องเข้าใจในตัวผู้บริโภคว่าต้องการซื้อสินค้าประเภทไหน ซึ่งแต่ละบุคคลก็อาจชอบแตกต่างกัน ดังนั้นผู้ประกอบการจะต้องส่งเสริมการขายที่แตกต่างกันตามแต่ละบุคคลด้วยเช่น ลดราคาสินค้าต่างกัน, มีของแถมเฉพาะสินค้านั้นๆ ที่ต่างกันตามช่วงเวลา, ซื้อสินค้าชิ้นหนึ่ง เมื่อซื้ออีกชิ้นที่กำหนดจะมีสิทธิ์ซื้อในราคาพิเศษ เป็นต้น รวมถึงโอกาสที่ผู้บริโภค หรือลูกค้าคนนั้นเมื่อซื้อสินค้า 1 - 2 ชิ้นแรก มีโอกาสที่จะซื้อสินค้าอื่นๆ ประกอบด้วย ซึ่งทำให้สามารถเพิ่มยอดขายสินค้าในกลุ่มนั้นๆ ได้ การที่ลูกค้ามีพฤติกรรมการบริโภคสินค้าที่ต่างกันนั้นทางผู้ประกอบการจำเป็นต้องเก็บข้อมูลของลูกค้าในรูปแบบของฐานข้อมูลสมาชิก เมื่อมีการซื้อสินค้าจะต้องแจ้งหมายเลขสมาชิกร่วมด้วย ทำให้ผู้ประกอบการสามารถนำข้อมูลสมาชิกที่เก็บรวบรวมไว้มาวิเคราะห์ความต้องการได้ถูกต้องเที่ยงตรงมากยิ่งขึ้น ด้วยวิธีการทำเหมืองข้อมูลจากฐานข้อมูลนำมาสร้างเป็นแบบจำลองพฤติกรรมผู้บริโภคเพื่อส่งเสริมด้านการตลาดเช่นได้แก่ การนำเสนอสินค้าเมื่อลูกค้าซื้อสินค้าที่จุดขายสินค้า (Cross-Selling) ในรูปแบบวางที่เคาน์เตอร์ขาย หรือผ่านพนักงาน Sale Counter ทำการเชียร์ (cheer) สินค้า และยังสามารถนำมาใช้สร้างความสัมพันธ์ที่ดีต่อลูกค้า ในรูปแบบ CRM: (Customer Relationship Management)

จากตัวอย่างข้างต้นผู้วิจัยได้นำหลักการดังกล่าวมาวิเคราะห์ข้อมูลในการศึกษาพฤติกรรมผู้บริโภคที่เป็นสมาชิกที่ดูแลเครือข่ายสายงานและยอดขายหลัก (แม่ทีม/ออฟไลน์) ในระบบธุรกิจ MLM (Multi-Level Marketing) โดยเน้นที่ผลิตภัณฑ์ใน 3 กลุ่มหลักคือ กลุ่มอาหารเสริม กลุ่มเครื่องสำอาง และกลุ่มอุปโภคบริโภค เนื่องจากเป็นสินค้าแกนหลักของบริษัท เพื่อใช้เป็นข้อมูลประกอบการวางแผนการนำเสนอสินค้า เพื่อส่งเสริมการขาย และเพิ่มยอดขายสินค้าต่อไป

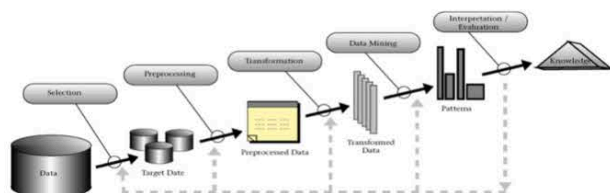
## วัตถุประสงค์ของการวิจัย

เพื่อค้นหารูปแบบกฎความสัมพันธ์ของการซื้อผลิตภัณฑ์ใน 3 กลุ่มหลักคือ กลุ่มอาหารเสริม กลุ่มเครื่องสำอาง และกลุ่มอุปโภคบริโภค เฉพาะสมาชิกที่ดูแลเครือข่ายสายงานและยอดขายหลัก (แม่ทีม/ออฟไลน์)

## แนวคิดทฤษฎีที่เกี่ยวข้อง

### 1. การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูลเป็นกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหารในด้านวิทยาศาสตร์การแพทย์ รวมทั้งในด้านเศรษฐกิจ และสังคม ซึ่งการทำเหมืองข้อมูลเปรียบเสมือนอภิวินิจฉัยการหนึ่งในการจัดเก็บ และตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในฐานข้อมูลที่สามารถดึงค่าสารสนเทศของข้อมูลมาใช้ จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล (บุญเสริม กิจศิริกุล, 2545)



รูปที่ 1. กระบวนการทำเหมืองข้อมูลในสภาพแวดล้อมแบบ KDD (Knowledge Discovery in Database)

ที่มา: O. Fayyad, (1996)

กระบวนการของการทำเหมืองข้อมูล นั้นประกอบ ด้วยขั้นตอนต่างๆ ดังนี้

1.1 การรวบรวมข้อมูล (Data integration) โดยรวมข้อมูลจากหลากหลายแหล่งเก็บข้อมูล

1.2 การคัดเลือกข้อมูล (Data selection) โดยเลือกเฉพาะข้อมูลที่ต้องการนำมาวิเคราะห์

1.3 การกลั่นกรองข้อมูล (Data cleaning) โดยนำข้อมูลที่ไม่มีค่าหรือ ข้อมูลที่ขาดหายหรือ ข้อมูลขยะ และรวมถึงข้อมูลที่มีค่าไม่แน่นอนออกไป

1.4 การแปรรูปข้อมูล (Transformation) โดยรวมกลุ่มข้อมูลและแปรข้อมูลให้อยู่ในรูปแบบที่พร้อมจะนำไปใช้ในการวิเคราะห์ หรือการทำเหมืองข้อมูล

1.5 การทำเหมืองข้อมูล (Data mining) โดยการประยุกต์เทคนิคการทำงานต่าง ๆ ของกระบวนการทำเหมืองข้อมูล เพื่อให้ได้ความรู้หรือสิ่งที่สนใจออกจากข้อมูลที่ได้แปรรูปไว้แล้ว

1.6 การประเมินรูปแบบ (Pattern evaluation) ซึ่งรูปแบบที่น่าสนใจควรเป็นสิ่งที่ผู้ใช้เชื่อถือได้ หรือยืนยันสมมติฐานที่มีเหตุผล

1.7 การนำเสนอความรู้สู่ผู้ใช้ (Knowledge presentation) เป็นการนำเอาความรู้ที่ได้จากการทำเหมืองข้อมูลไปประยุกต์ใช้ในการตัดสินใจ เพื่อนำไปวางแผน และปรับกลยุทธ์ทางธุรกิจให้มีประสิทธิภาพมากขึ้น (เรวดี ศักดิ์ดุลยธรรม, 2552)

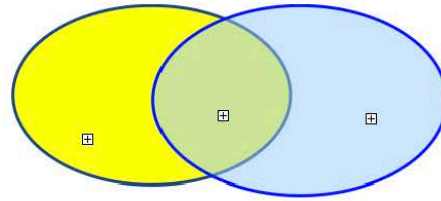
### 2. กฎความสัมพันธ์ (Association Rules)

กฎความสัมพันธ์ (Association Rules) ถูกนำมาใช้เพื่อหาความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ (Agrawal, Imielinski, and Swami, 1993) กฎความสัมพันธ์ถูกนำไปใช้อย่างแพร่หลายในด้านต่าง ๆ เช่นการวิเคราะห์พฤติกรรมของลูกค้า, การทำนายรูปแบบการเข้า

ใช้เว็บไซต์, การทดลองทางวิทยาศาสตร์, การรักษาโรคและการทำนาย ปรากฏการณ์ธรรมชาติเป็นต้น ขั้นตอนการหากฎความสัมพันธ์แบ่งออกเป็น 2 ขั้นตอนใหญ่ ขั้นตอนแรกคือ การสืบค้นเซตรายการความถี่ (Frequent Itemset Mining) ซึ่งเป็นการสืบค้นเซตรายการที่มีความถี่หรือค่าสนับสนุน (Support) มากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำ (Minimum Support) จากนั้นเซตรายการความถี่จะถูกนำไปสร้างกฎความสัมพันธ์ในขั้นตอนที่สอง ซึ่งจะเห็นว่ากฎความสัมพันธ์ที่ได้จะขึ้นอยู่กับเซตรายการความถี่ถ้าเซตรายการความถี่มีจำนวนมากกฎความสัมพันธ์ที่ได้ก็จะมีจำนวนมากตาม (พนิดา ทรงรัมย์, 2555) เช่น การค้นหาความสัมพันธ์ของลูกค้าและการซื้อสินค้า โดยมีรากฐานมาจากทฤษฎีการวิเคราะห์ตะกร้าตลาด (Market Basket Analysis) ซึ่งผลการวิเคราะห์จะอยู่ในรูปแบบกฎความสัมพันธ์ โดยมีค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) เป็นค่าที่บอกถึงความน่าเชื่อถือของกฎ โดยอยู่ในรูปแบบของการทำนายว่า หากมีเหตุการณ์ใดเกิดขึ้น แล้วจะมีเหตุการณ์อื่นใดเกิดขึ้นตามมาบ้างด้วยความมั่นใจที่เปอร์เซ็นต์ (ธีรวัฒน์ แสนปัญญา, 2554)

### 3. อัลกอริทึมเอพริออริ (Apriori Algorithm)

เป็นอัลกอริทึมพื้นฐานที่ใช้ในการหาความสัมพันธ์ของข้อมูล โดยใช้หลักการค้นหาแบบวงกว้างก่อนนับทรานแซคชัน ซึ่งจะทำการสร้างและตรวจสอบไอเท็มที่เกิดขึ้นบ่อยทีละชั้น โดยเริ่มจากเซตไอเท็มที่มีจำนวนสมาชิกเท่ากับหนึ่ง ถ้าเซตไอเท็มใดมีค่าสนับสนุนน้อยกว่าค่าที่กำหนดไว้ต่ำสุดก็จะตัดเซตไอเท็มนั้นออก ไม่นำไปสร้างเซตไอเท็มในขั้นถัดไป การกระทำของอัลกอริทึมจะวนไปเรื่อยๆ จนกระทั่งไล่ทุกระดับชั้นไม่เหลือเซตไอเท็มที่จะสร้างต่อไป ในการนับจำนวนทรานแซคชันจะไล่นับครั้งเดียวในแต่ละระดับชั้น และใช้ Hash tree ในการค้นหาไอเท็มเซตที่ปรากฏบ่อยและละเว้นไอเท็มเซตที่ปรากฏซ้ำด้วยความถี่ต่ำกว่าเกณฑ์ (ธีรวัฒน์ แสนปัญญา, 2554)



รูปที่ 2. ค่าสนับสนุน และความเชื่อมั่นจากกฎความสัมพันธ์  
ที่มา: ผศ.ดร.จิรัฐฐา ญบุญอบ (2011)

จากรูปที่ 2. อธิบายความสัมพันธ์ของกฎ เพื่อหาค่าสนับสนุน และความเชื่อมั่น ของสินค้า 2 ชนิด ดังนี้

3.1 ค่าสนับสนุน (support) ของกฎ ในการซื้อสินค้า diapers และสินค้า beer พบว่า ความน่าจะเป็นของการซื้อเป็นไปได้อันทั้ง diapers และ beer

3.2 ค่าความเชื่อมั่น (Confidence) ของกฎ ในการซื้อสินค้า diapers และสินค้า beer พบว่า ความน่าจะเป็นแบบมีเงื่อนไข ทำให้มีการซื้อสินค้า diapers พร้อมกับการซื้อ beer

### วิธีดำเนินการวิจัย

ข้อมูลที่น่านำมาใช้เป็นข้อมูลกลุ่มตัวอย่างจะเป็นข้อมูลการซื้อผลิตภัณฑ์กลุ่มอาหารเสริม กลุ่มเครื่องสำอาง และกลุ่มอุปโภคบริโภค เฉพาะสมาชิก ระดับแม่ทีมที่ดูแลยอดขาย และเครือข่ายสายงาน บริษัทเดอะไรท์พาวเวอร์ จำกัด ปี พ.ศ. 2548 จำนวนทั้งสิ้น 24,879 รายการ โดยมีสินค้าในกลุ่มอาหารเสริม กลุ่มเครื่องสำอาง และกลุ่มอุปโภคบริโภคที่จะนำมาวิเคราะห์จำนวน 108 รายการ

ซอฟต์แวร์ที่ใช้ในการจัดเก็บข้อมูล เรียกใช้ข้อมูล กลั่นกรองข้อมูล แปลงรูปข้อมูล และวิเคราะห์ข้อมูลดังนี้

(1) Microsoft SQL Server 2005 เป็นซอฟต์แวร์จัดการฐานข้อมูล

- 1.1 ใช้ในการจัดเก็บข้อมูลกลุ่มตัวอย่างที่นำมาใช้
- 1.2 ใช้ภาษา SQL ในการเรียกใช้ข้อมูล (Query) จากฐานข้อมูลตามที่กำหนด
- 1.3 ใช้ภาษา SQL ในการกลั่นกรองคัดเลือกข้อมูล
- 1.4 ใช้ในการแปลงข้อมูลเข้า และออกทั้งในรูปแบบ xls (Excel Worksheet) และ csv (Comma Separate Value)

(2) Seagate Crystal Report 9.2 เป็นซอฟต์แวร์ที่เรียกข้อมูล (SQL) จากฐานข้อมูล มาสร้างในรูปรายงานแบบการแจกแจงตารางข้อมูลแบบไขว้ (Crosstab Report) และส่งออกในรูปแบบ csv

(3) Microsoft Excel เป็นซอฟต์แวร์ที่ใช้ในการสร้างแก้ไขข้อมูลแบบ csv

(4) EditPlus เป็นซอฟต์แวร์แบบแก้ไข ค้นหา เปลี่ยนแปลงข้อมูลแบบข้อความตัวอักษรในรูปแบบ Text Format เช่น csv และ arff (attribute-relation file format)

(5) WEKA 3.7.9 เป็นซอฟต์แวร์ที่ใช้ในการวิเคราะห์ และพยากรณ์ข้อมูล รวมถึงหาค่าความสัมพันธ์ของข้อมูลในรูปแบบ Associate Rules โดยใช้ Apriori Algorithm

การเตรียมข้อมูลให้เหมาะสมสำหรับการทำเหมืองข้อมูลซึ่งแบ่งเป็นขั้นตอนย่อย 3 ขั้นตอนดังนี้

**ขั้นตอนที่ 1. การเลือกข้อมูล (Data Selection)**  
เป็นการเลือกข้อมูลที่สนใจ มาใช้ในการทำเหมืองข้อมูลในที่นี้คือ รหัสใบสั่งซื้อ (Orderno), รหัสสินค้าที่สั่งซื้อ (Itemcd), จำนวนการสั่งซื้อ (Qty) โดยมีข้อกำหนดช่วงระยะที่จะนำมาใช้คือวันสั่งซื้อ (Orderdate) ต้องอยู่ในช่วงเดือน มกราคม ถึง เดือน ธันวาคม ปี พ.ศ.2548 และเป็นข้อมูลที่สถานะใบสั่งซื้อ (Cancel) ไม่ถูกยกเลิกคือเป็น 0 และทำการเลือกเฉพาะสมาชิกที่เป็นแม่ทีม (Holdno) ต้องไม่เท่ากับ 0 ซึ่งในส่วนนี้จะใช้ภาษา SQL ในการคัดเลือกข้อมูล

**ตารางที่ 1. แสดงตัวอย่างข้อมูลการสั่งซื้อสินค้า 12 รายการ**

No	Orderno	Itemcd	Qty
1	BKO4810-034372	0301005	2
2	BKO4810-034372	0401007	1
3	BKO4810-034372	0302006	1
4	BKO4810-034373	0201009	1
5	BKO4810-034373	0101010	2
6	BKO4810-034430	0101010	2
7	BKO4810-034430	0303002	4
8	BKO4810-034430	0301001	3
9	BKO4810-034445	0401008	1
10	BKO4810-034445	0101010	2
11	BKO4810-034445	0302006	2
12	BKO4810-034445	0302004	1

**ที่มา :** ตารางจัดทำขึ้นโดยผู้วิจัย

**ขั้นตอนที่ 2. การกลั่นกรองข้อมูล (Data Cleaning)**  
เป็นขั้นตอนการกำจัดข้อมูลที่ไม่สนใจ ข้อมูลขยะ หรือไม่สมบูรณ์ออกไปในที่นี้คือ จำนวนการสั่งซื้อ (Qty) ต้องมากกว่า 0 ขึ้นขึ้นไป, รหัสสินค้า (Itemcd) ต้องไม่เป็นค่าว่าง (Null) และต้องไม่มีค่าว่างแบบเคาะ space, รายการสินค้าต้องตัดออกจากรายการคืนสินค้าในรายการลดหนี้ (Credit Note) และต้องเป็นสินค้าในกลุ่มอาหารเสริม เครื่องสำอาง และกลุ่มอุปโภคบริโภคเท่านั้น ถ้าไม่กรองข้อมูลเหล่านี้ออกจะทำให้ได้กฎความสัมพันธ์ที่ไม่ถูกต้อง โดยในส่วนนี้จะใช้ภาษา SQL ในการคัดกรองข้อมูล

**ขั้นตอนที่ 3.** การแปลงข้อมูล (Data Transformation) เป็นการแปลงข้อมูลหลังจากการกรองข้อมูลในขั้นที่ 2 แล้วแปลงให้อยู่ในรูปแบบข้อมูลตารางแบบไขว้ (Crosstab Report)

(1) ใช้ภาษา SQL ในรูปแบบ Stored Procedure ติดต่อกับโปรแกรม Seagate Crystal Report ในรูปแบบ OLEDB ADO for Microsoft SQL Server เพื่อดึงข้อมูลเข้าจนครบ และอยู่ในรูปแบบ Crosstab

(2) ทำการแปลงข้อมูลแบบ Crosstab ที่ได้ออกเป็น xls ในรูปแบบ Microsoft Excel

(3) ใช้โปรแกรม Microsoft Excel ทำการแก้ไขข้อมูล และแปลงให้อยู่ในรูปแบบ csv (Comma Separate Value)

(4) ใช้โปรแกรม EditPlus ในการแก้ไขเปลี่ยนแปลงข้อมูลให้อยู่ในรูปแบบไฟล์ arff ซึ่งเป็นไฟล์นำเข้าข้อมูลในการวิเคราะห์ของโปรแกรม WEKA

โครงสร้างของไฟล์ arff จะแบ่งออกเป็น 3 ส่วนดังนี้

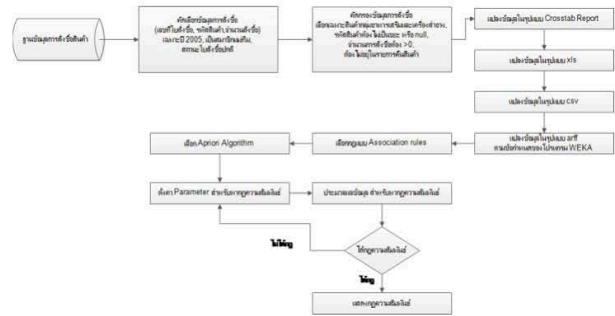
ส่วนที่ 1 @relation ใช้สำหรับกำหนดโครงสร้างชื่อของข้อมูลตาราง

ส่วนที่ 2 @attribute ใช้สำหรับกำหนดโครงสร้าง Attribute สินค้าทั้งหมด

ส่วนที่ 3 @data ใช้สำหรับกำหนดข้อมูลที่แปลงรูปแล้ว เพื่อนำมาวิเคราะห์ ในโปรแกรม WEKA

จากข้อมูลให้นำมาวิเคราะห์กำหนดชื่อโครงสร้างตารางคือ mlmHold2005 โดยมี attribute ทั้งหมด 108 attributes และมีข้อมูลทั้งหมดหลังจากการแปลงรูปแล้ว 5,606 รายการ ทำการวิเคราะห์ผลโดยใช้โปรแกรม WEKA โดยใช้กฎความสัมพันธ์ Associate และอัลกอริทึม Apriori โดยทำการทดสอบปรับเปลี่ยนค่า parameter เพื่อหากฎ ซึ่งจะปรับค่าในส่วน lowerBoundMinSupport เริ่มต้นที่ 0.1 และ ค่า minMetric เริ่มต้นที่ 0.9 หมายถึงตั้งค่าการสนับสนุนที่ 10% และความเชื่อมั่นที่ 90% ซึ่งไม่พบกฎ

ใด ๆ ทดลองปรับค่าไปเรื่อยๆ ที่ lowerBoundMinSupport เริ่มต้นที่ 0.05 และ ค่า minMetric ที่ 0.5 ซึ่งจะพบกฎทั้งหมด 6 กฎ สำหรับการตั้งค่าการสนับสนุนที่ 5% และความเชื่อมั่นที่ 50%



รูปที่ 3. ลำดับขั้นตอนในการทำเหมืองข้อมูล โดยการค้นหากฎความสัมพันธ์

## ผลการวิจัย

ข้อมูลการสั่งซื้อสินค้าที่ใช้ในการนำเสนออยู่ในช่วงเดือน มกราคม ถึง เดือนธันวาคม ปี พ.ศ. 2548 จำนวน 24,879 รายการ กำหนดให้เป็น Training set ในส่วน Test set อยู่ระหว่างเดือน พฤศจิกายน ถึง เดือนธันวาคม ปี พ.ศ. 2547 จำนวน 8,462 รายการ ถูกนำมาใช้ทดสอบความถูกต้องของ Model เนื่องจากไม่สามารถนำข้อมูลที่มีอยู่ก่อนเดือน ตุลาคม ปี พ.ศ. 2547 มาใช้ได้เพราะสินค้าในกลุ่มผลิตภัณฑ์ดังกล่าวที่นำมาใช้ในการวิเคราะห์ยังมีจำนวนน้อยอาจทำให้ได้กฎความสัมพันธ์ที่ไม่ถูกต้อง ซึ่งหลังจากวิเคราะห์ประมวลผลจะได้กฎที่สอดคล้องกันโดยอัตราส่วนระหว่าง Training set และ Test set อยู่ในช่วง 74% ต่อ 26% ซึ่งสอดคล้องกับกฎที่ได้จำนวน 6 กฎ โดยในบทความนี้จะกล่าวถึงเฉพาะ Training set เท่านั้นเพราะข้อมูลที่นำมาวิเคราะห์ครอบคลุมทั้ง ปี 2548 ซึ่งหลังจากการทดสอบ และวิเคราะห์กฎความสัมพันธ์ของการสั่งซื้อสินค้ากลุ่มผลิตภัณฑ์อาหารเสริม เครื่องสำอาง และกลุ่มอุปโภคบริโภค จะพบกฎความสัมพันธ์ทั้งสิ้นจำนวน 6 กฎ ดังนี้

กฎข้อที่ 1. เมื่อซื้อสินค้า ผงซักฟอก The right (0401007) ก็จะซื้อยาสีฟันสูตรสมุนไพร (0301001) ไปด้วย โดยมีความเชื่อมั่นที่ 61%

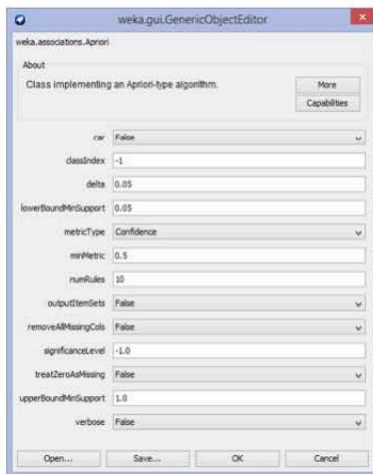
กฎข้อที่ 2. เมื่อซื้อสินค้า ครีมอาบน้ำไรท์ เนเจอร์ สปา (0302018) ก็จะซื้อยาสีฟันสูตรสมุนไพร (0301001) ไปด้วย โดยมีความเชื่อมั่นที่ 60%

กฎข้อที่ 3. เมื่อซื้อสินค้า สบู่ The Right 1แพ็ค (0302017) ก็จะซื้อยาสีฟันสูตรสมุนไพร (0301001) ไปด้วย โดยมีความเชื่อมั่นที่ 59%

กฎข้อที่ 4. เมื่อซื้อสินค้า ยาสีฟันไรท์ ไวท์เทนนิ่ง (0301005) ก็จะซื้อยาสีฟันสูตรสมุนไพร (0301001) ไปด้วย โดยมีความเชื่อมั่นที่ 58%

กฎข้อที่ 5. เมื่อซื้อสินค้า โยเกิร์ตโฟม (0302017) ก็จะซื้อยาสีฟันสูตรสมุนไพร (0301001) ไปด้วย โดยมีความเชื่อมั่นที่ 54%

กฎข้อที่ 6. เมื่อซื้อสินค้า โยเกิร์ตโฟม (0302017) ก็จะซื้อโฟมวิตามิน ซี (0201002) ไปด้วย โดยมีความเชื่อมั่นที่ 50%



รูปที่ 4. ส่วนการปรับค่า Parameter ของอัลกอริทึม Apriori แบบ Associate

จากรูปที่ 4. เป็นส่วนการปรับค่า Parameter ของอัลกอริทึม Apriori สำหรับหากฎความสัมพันธ์ในโปรแกรม WEKA 3.7.9 ซึ่งเป็นค่า parameter ที่เลือกใช้ในการสร้างกฎ ประกอบไปด้วยค่าต่าง ๆ ดังนี้

(1) lowerBoundMinSupport เป็นส่วนการตั้งค่า Minimum Support ที่ 0.05 หรือ 5%

(2) metricType เป็นส่วนการเลือกแบบค่าความเชื่อมั่น (Confidence)

(3) minMetric เป็นการกำหนดค่าความเชื่อมั่นที่ 0.50 หรือ 50%

(4) numRules เป็นจำนวนของกฎที่เลือก

(5) outputItemSets เป็นการแสดงจำนวน Item

รูปแบบของกฎความสัมพันธ์ที่ได้ มีการกำหนดค่า Minimum Support ที่ 0.05 ถ้ากำหนดค่ามากกว่านี้ จะไม่ได้กฎความสัมพันธ์ใดๆ นอกจากนั้นค่าความเชื่อมั่น minMetric กำหนดเท่ากับ 0.50 (50%) ถ้ากำหนดที่ 0.60 (60%) จะได้เพียง 2 กฎ และกำหนด 0.70 (70%) ขึ้นไปจะไม่ได้กฎใดๆ ดังนั้นจึงเลือกกำหนดค่าความเชื่อมั่นที่ 50% ซึ่งจะได้ทั้งหมด 6 กฎ ดังที่กล่าวมาแล้ว

เมื่อพิจารณาจาก ไอเท็มสินค้าที่วิเคราะห์ออกมาพบว่า จากสินค้าทั้งหมด 3 กลุ่มจะมีเพียงกลุ่มสินค้าอุปโภคบริโภคเท่านั้นที่พบกฎความสัมพันธ์ สินค้าในกลุ่มอาหารเสริม และเครื่องสำอางไม่พบกฎความสัมพันธ์ใด ๆ คาดว่ามาจากกลุ่มสินค้าอุปโภคบริโภคจะมีราคาค่อนข้างต่ำ และจำหน่ายต่อได้ง่ายกว่าสินค้าอีก 2 กลุ่มที่เหลือซึ่งมีราคาแพงกว่ามาก

## สรุปและอภิปรายผล

การนำรูปแบบวิธีการทำเหมืองข้อมูลมาใช้กับระบบงานธุรกิจนั้น จะต้องเข้าใจถึงประโยชน์แต่ละขั้นตอนวิธี และเลือกใช้ขั้นตอนวิธีที่เหมาะสมกับระบบงานมากที่สุด โดยต้องเข้าใจขั้นตอนกระบวนการ ตลอดจนความต้องการในระบบธุรกิจเพื่อช่วยให้ทราบความสัมพันธ์ของข้อมูลที่ต้องการ

อย่างชัดเจน การเลือกใช้รูปแบบของกฎใด ๆ จะต้องทำการทดสอบในหลาย ๆ กฎหลายวิธี เพื่อเลือกกฎที่ตรงกับความต้องการทำให้ได้รูปแบบของกฎที่เหมาะสมนำมาใช้ และพัฒนาระบบธุรกิจนั้น ๆ ต่อไป

## เอกสารอ้างอิง

กฤษฎาภกร กังอุบล. (2552). การใช้เกณฑ์ความต่างลำดับในการปรับปรุงกฎความสัมพันธ์จำแนกประเภทข้อมูล. ปรินญาณิพนธ์ วศ.ม. (วิทยาศาสตรกรรมคอมพิวเตอร์) มหาวิทยาลัยเกษตรศาสตร์.

ธีรวัฒน์ แสนปัญญา. (2554). การพยากรณ์การซื้อสินค้าประเภทกระเบื้องปูพื้นสำหรับร้านค้าปลีกโดยวิธีกฎความสัมพันธ์. ปรินญาณิพนธ์ วท.ม. (วิทยาการคอมพิวเตอร์และสารสนเทศ) มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

บุญเสริม กิจศิริกุล. (2545). รายงานวิจัยฉบับสมบูรณ์โครงการวิจัยร่วมภาครัฐและเอกชน ปีงบประมาณ 2545 โครงการย่อยที่ 7 อัลกอริทึมการทำเหมืองข้อมูล. ในรายงานวิจัยฉบับสมบูรณ์. ภาควิชา วิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

พนิดา ทรงรัมย์. (2555). การสร้างกฎความสัมพันธ์จากเซตรายการความถี่ และเซตรายการความถี่แบบปิด. วารสารเทคโนโลยีสารสนเทศ. 8(2): 22-28. มหาวิทยาลัยมหาสารคาม.

พีรภาวี เกียรติเฉลิมคุณ. (2553). การพัฒนาแบบจำลองสำหรับการพยากรณ์ยอดขายด้วยเทคนิคการทำเหมืองข้อมูล กรณีศึกษา: บริษัท ที. สไตล์โปรดักส์ จำกัด. ปรินญาณิพนธ์ วท.ม. (การบริหารเทคโนโลยี) วิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์.

เรวดี ศักดิ์ดุษฎยธรรม. (2552). การใช้เทคนิคดาต้าไมนิงในการสร้างฐานความรู้ เพื่อการทำนายสัมฤทธิ์ผลทางการเรียนของนักศึกษา วิทยาลัยราชพฤกษ์. วิทยาลัยราชพฤกษ์.

ศราวุธ สิริขจรเดชสกุล และกมล เกียรติเรืองกมล. (2555). การออกแบบและพัฒนางานขายสินค้าต่อเนื่องในพื้นที่กรุงเทพมหานครย่านธุรกิจ สำหรับร้านค้าปลีก โดยใช้เทคโนโลยีเหมืองข้อมูลกรณีศึกษา ร้านมิตรใจ. ปรินญาณิพนธ์ วท.ม. (บริหารเทคโนโลยี) วิทยาลัยนวัตกรรม มหาวิทยาลัยธรรมศาสตร์.