



การแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์

Customers Segmentation using K-means Clustering Technique

for Customer Relationship Management

วนิษา แผลงรักษา* และ นิเวศ จิระวิชิตชัย

หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม 2410/2 ถ.พหลโยธิน เขตจตุจักร กรุงเทพฯ 10900

Submitted 16/7/2019 ; Revised 30/8/2019 ; Accepted 6/9/2019

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อนำเสนอการแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์โดยทดสอบกับระบบฐานข้อมูลเกี่ยวกับลักษณะลูกค้าของผู้จำหน่ายสินค้าออนไลน์แห่งหนึ่ง โดยพิจารณาจากคำสั่งซื้อสินค้าของลูกค้าทั้งหมดจำนวน 1,000 คน เพื่อนำมาวิเคราะห์การแบ่งกลุ่มคลัสเตอร์ด้วยเคมีน อาทิ ข้อมูลสถานภาพ ข้อมูลเพศ ข้อมูลรายได้ต่อเดือน จำนวนบุตร การสำเร็จการศึกษา อาชีพ ระยะทางระหว่างที่พัก-ที่ทำงาน ภูมิลำเนาและจำนวนรถที่เป็นเจ้าของนำวิเคราะห์ลักษณะทางประชากรศาสตร์ ผลจากการทดลองโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน พบว่าสามารถแบ่งกลุ่มลูกค้าที่มีคุณลักษณะเหมือนกันไว้ในกลุ่มเดียวกันได้อย่างมีประสิทธิภาพและสามารถนำไปใช้เพื่อการส่งเสริมการขายได้ตรงกลุ่มเป้าหมายได้เป็นอย่างดี

คำสำคัญ: เหมืองข้อมูล การบริหารลูกค้าสัมพันธ์ การทำคลัสเตอร์แบบเคมีน

Abstract

This objective of this research presents a customer segmentation using K-mean clustering techniques for customer relationship management by testing with a database on the characteristics of customers online distributor. The experiment considers the purchase order of the 1,000 customers to analyze the segmentation using K-means technique such as status, gender, monthly income, number of children, graduation, occupation, distance, domicile and number of vehicles owned to analyze demographic characteristics. The experiment results of the analysis using K-mean clustering techniques show that the group of customers with similar characteristics can be divided into the same group efficiently and can be used for sales promotion to the target group effectively.

Keyword: Data Mining, Customer Relationship Management, K-means Clustering

***ผู้ประสานงาน (Corresponding Author)**

E-mail: way.go_wan@hotmail.com



1. บทนำ

ในปัจจุบันเทคโนโลยีมีความเจริญก้าวหน้าเป็นอย่างมากและถือว่าเป็นอันดับต้น ๆ ซึ่งกล่าวได้ว่าเป็นส่วนหนึ่งของชีวิตประจำวัน ซึ่งการติดต่อสื่อสารนั้น จะรับรู้ได้อย่างรวดเร็วภายในระยะเวลาเสี้ยววินาที ซึ่งส่งผลให้การธุรกรรมต่าง ๆ รวมถึงมีการนำเอาเทคโนโลยีมาใช้กันอย่างมากขึ้น จากสภาวะการแข่งขันเพื่อให้ได้ชัยชนะทางธุรกิจจำเป็นต้องมีกลยุทธ์หรือยุทธวิธี ที่เชื่อมั่นได้ว่าจะลดความเสี่ยงขององค์กรลงได้ กลยุทธ์วิธีการต่าง ๆ จำเป็นต้องมีฐานความรู้ เพื่อใช้ในการสร้างกรอบในการทำงาน (Framework) ที่สนองตอบกับกลยุทธ์ทางธุรกิจ ส่งผลให้การแข่งขันทางการค้ามีปริมาณสูงขึ้น ทำให้บริษัทต่าง ๆ ตื่นตัวกับการนำเทคโนโลยีสารสนเทศและเทคโนโลยีการส่งเสริมมาใช้ในการสนับสนุนการทำงานประชาสัมพันธ์ รวมไปถึงกลยุทธ์การส่งเสริมการตลาดจะต้องประสานกับแผนการตลาดโดยรวมและควรกำหนดแผนการส่งเสริมการตลาดที่เฉพาะเจาะจงให้เกิดผลสูงสุดเพื่อกระตุ้นให้เกิดการซื้อขาย [1] และชักจูงใจให้ผู้บริโภคมีแรงจูงใจในการสั่งซื้อสินค้ามากมายหลายรูปแบบและหนึ่งในเทคนิควิธีที่จะช่วยให้ธุรกิจมีความแตกต่างและเป็นที่น่าสนใจของผู้บริโภค คือ การให้ความสนใจกับการสร้างความพึงพอใจให้กับลูกค้า การตอบสนองต่อความต้องการของลูกค้าที่ถูกต้องและรวดเร็วจะนำมาซึ่งการเติบโตทางธุรกิจ [2]

โดยที่วิธีการที่จะสามารถสร้างความพึงพอใจให้กับกลุ่มลูกค้าได้ดีและเหมาะสมที่สุด ก็คือการนำเอาฐานข้อมูลของพฤติกรรมผู้บริโภคหรือการสั่งซื้อสินค้าของลูกค้าที่มีอยู่มาทำการศึกษาวิเคราะห์เชิงธุรกิจ เพื่อวิเคราะห์พฤติกรรมผู้บริโภคสินค้าของลูกค้าเพื่อให้ทราบทิศทางในการนำเสนอรายการส่งเสริมการขายได้ตรงกับกลุ่มเป้าหมายและสอดคล้องกับพฤติกรรมในการสั่งซื้อสินค้า ซึ่งจะนำไปสู่กลยุทธ์ทางการตลาดที่มีประสิทธิภาพและเหมาะกับกลุ่มผู้บริโภค

การที่จะได้มาซึ่งฐานความรู้และกรอบการทำงานที่มีประโยชน์ จำเป็นต้องมีเทคโนโลยีสารสนเทศที่สามารถวิเคราะห์ข้อมูลทางธุรกิจ

เทคโนโลยีสารสนเทศที่สามารถกลั่นกรองข้อมูลทางธุรกิจที่มีปริมาณมหาศาล เพื่อให้ได้ข้อมูลที่มีประโยชน์ในการคิดกลยุทธ์ ดังนั้นในขณะนี้ เหมืองข้อมูลจึงเป็นเทคโนโลยีสารสนเทศที่ได้รับการกล่าวถึงมากที่สุด เหตุผลที่สำคัญสำหรับทำไมถึงมีเหมืองข้อมูลและทำไมถึงต้องทำเหมืองข้อมูลนั้นก็เพราะว่าเหมืองข้อมูล [3] เป็นเทคโนโลยีสารสนเทศที่สามารถกลั่นกรอง วิเคราะห์ ข้อมูลที่มีปริมาณมหาศาลเพื่อให้ได้ข้อมูลที่มีประโยชน์หรือได้ข้อมูลที่ซ่อนเร้นอยู่ในข้อมูลที่มีปริมาณมหาศาล และนำข้อมูลที่มีประโยชน์มาใช้เป็นฐานความรู้เพื่อช่วยในการบริหารงาน เช่น การบริหารลูกค้าสัมพันธ์ หรือการบริหารความสัมพันธ์กับลูกค้า (Customer Relationship Management) ด้วยการใช้เทคโนโลยีมีความเจริญก้าวหน้าเป็นอย่างมาก ส่งผลให้ธุรกิจหลาย ๆ ด้าน เป็นการบูรณาการเครื่องมือทางธุรกิจอิเล็กทรอนิกส์และกระบวนการให้บริการลูกค้าในการขาย การตลาดทางตรง การจัดการทางบัญชีและกระบวนการสั่งซื้อ และการสนับสนุนการให้บริการลูกค้าได้อย่างสมบูรณ์แบบ [4]

งานวิจัยนี้จึงนำเสนอทางเลือกเพื่อการวิเคราะห์ข้อมูลในเชิงธุรกิจ โดยจะใช้วิธีการสืบค้นข้อมูลจากฐานข้อมูลหรือที่เรียกว่าการค้นหาข้อมูลโดยเป็นสารสนเทศที่มีเหตุผลและสามารถนำไปใช้ได้ และช่วยในการแก้ปัญหาการกลั่นกรองและวิเคราะห์ ซึ่งเป็นสิ่งสำคัญในการที่จะช่วยการตัดสินใจในการทำธุรกิจ เหมืองข้อมูลซึ่งจะทำตามพฤติกรรมในการสั่งซื้อสินค้า ซึ่งมีการประมาณการจากข้อมูลประชากรตามกลุ่มเป้าหมายและข้อมูลจำเพาะทางกายภาพต่าง ๆ เพื่อตอบสนองความต้องการที่ถูกต้องและเอื้อประโยชน์ต่อการเสนอรายการส่งเสริม การขายได้ตรงกลุ่มเป้าหมายและประเภทสินค้าเทคนิคการทำคลัสเตอร์แบบเคมีน การแบ่งกลุ่มลูกค้าที่มีลักษณะเหมือนกันไว้อยู่ในกลุ่มเดียวกัน ทำให้เราเข้าใจคุณลักษณะของลูกค้ามากขึ้น และทราบว่าลูกค้าสามารถจำแนกออกมาได้กี่กลุ่มมีลักษณะทางพฤติกรรมในรูปแบบใด ซึ่งจะช่วยให้สามารถกำหนดกลยุทธ์ทางการตลาดเพื่อจับกลุ่มเป้าหมายได้ถูกต้องแม่นยำมากขึ้น [5] และเพื่อสนับสนุนการวิเคราะห์



ข้อมูล โดยนำข้อมูลจากระบบสารสนเทศมาสังเคราะห์ในการบริการลูกค้า [6]

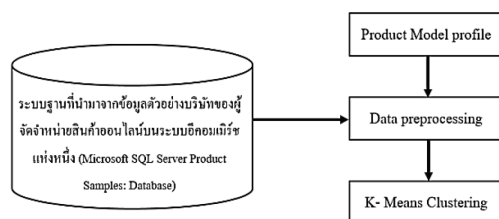
ดังนั้น คำถามวิจัยของงานวิจัยนี้คือ การใช้เทคนิคการทำคลัสเตอร์แบบเคมีนที่เหมาะสมกับการแบ่งกลุ่มลูกค้าที่มีประสิทธิภาพจริงหรือไม่ และการใช้เทคนิคการทำคลัสเตอร์แบบเคมีนในการทำคลัสเตอร์สำหรับการบริหารลูกค้าสัมพันธ์ต้องอยู่ในระดับใดจึงจะมีประสิทธิภาพ

2. วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาเทคนิคการแบ่งกลุ่มลูกค้าและคุณลักษณะเด่น โดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน
2. เพื่อศึกษาคุณลักษณะทางด้านประชากรศาสตร์ เพื่อใช้ในการทำคลัสเตอร์สำหรับการบริหารลูกค้าสัมพันธ์

3. วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเทคนิคการแบ่งกลุ่มประชากรและคุณลักษณะเด่นจากฐานข้อมูลที่นำมาจากข้อมูลตัวอย่าง โดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์ เพื่อศึกษาประสิทธิภาพของผู้ใช้ระบบ โดยมีกรอบแนวคิดการดำเนินงาน ดังแสดงในภาพที่ 1

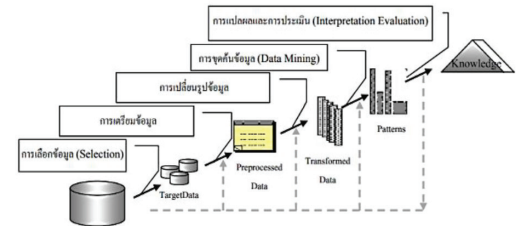


ภาพที่ 1 กรอบแนวคิดการดำเนินงาน เพื่อการแบ่งกลุ่มลูกค้า โดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์

3.1 ขั้นตอนในการดำเนินการวิจัย

โดยวิธีการทดลองใช้เทคนิควิธีการสืบค้นข้อมูลจากฐานข้อมูล สามารถสรุปวิธีการสืบค้นได้ดังนี้คือ การเตรียมข้อมูล (Pre-Processing) เป็นขั้นตอนการรวบรวมข้อมูลจากการขุดค้นข้อมูลเป็น

หนึ่งในขั้นตอนของกระบวนการค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery) ซึ่งประกอบไปด้วยขั้นตอน ดังภาพที่ 2

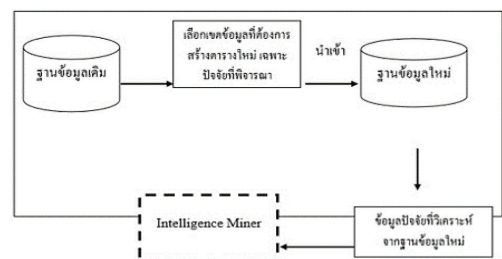


ภาพที่ 2 กระบวนการค้นหาความรู้ในการขุดค้นข้อมูล [3]

โดยมีข้อมูลชนิดของสินค้าและข้อมูลการสั่งซื้อสินค้า ประกอบด้วยขั้นตอนการดำเนินงาน 4 ขั้นตอน

1. การรวบรวมข้อมูล (Data Consolidation) จากแหล่งต่าง ๆ ที่เก็บไว้ให้มาอยู่รวมกันในข้อมูลระบบฐานที่นำมาจากข้อมูลตัวอย่างบริษัทของผู้จัดทำนายนิตยภัตออนไลน์บนระบบอีคอมเมิร์ซแห่งหนึ่ง (Microsoft SQL Server Product Samples: Database) โดยข้อมูลการสั่งซื้อสินค้าที่รวบรวมได้ทั้งหมดที่ใช้ในการทดลองจากฐานข้อมูลการสั่งซื้อสินค้าของลูกค้าทั้งหมดของรถจักรยาน จำนวน 1,000 คน

2. การคัดเลือกข้อมูล (Data Selection) จากฐานข้อมูลเดิมที่ต้องการจะนำไปใช้ประโยชน์เป็นข้อมูลที่เป็นปัจจุบัน และเป็นข้อมูลที่สนใจจะทำการเลือกเขตข้อมูลต่าง ๆ จากตาราง เฉพาะปัจจัยที่พิจารณานำมาแปลงให้อยู่ในรูปแบบ text file เพื่อทำการส่งข้อมูลออก (Export) ไปสู่ฐานข้อมูลใหม่ ดังภาพที่ 3



ภาพที่ 3 ภาพวิธีการคัดเลือกข้อมูล จากฐานข้อมูลเดิม สู่ฐานข้อมูลใหม่

3. การจัดรูปแบบข้อมูล หรือการเปลี่ยนแปลงรูปแบบข้อมูล (Data Transformation) เป็นการนำข้อมูลทั้งหมดมาจัดให้อยู่ในรูปแบบเดียวกัน เนื่องจากการวิจัยนี้ต้องใช้ข้อมูลจากหลายแหล่ง ได้แก่ ฐานข้อมูลส่วนตัวลูกค้า ข้อมูลสถานภาพ ข้อมูลเพศ ข้อมูลรายได้ต่อเดือน จำนวนบุตร การสำเร็จการศึกษา อาชีพ ระยะทางระหว่างที่พัก-ที่ทำงาน ภูมิลำเนาและจำนวนรถที่เป็นเจ้าของ ดังนั้นต้องมีการจัดรูปแบบข้อมูลให้อยู่ในรูปแบบเดียวกัน โดยสร้างโปรแกรมเพื่อใช้จัดการกับการทำงานในส่วนนี้จะสามารถนำไปใช้ในขั้นตอนต่อไปได้ และการแบ่งข้อมูลออกเป็นกลุ่มตามปัจจัยที่กล่าวไปข้างต้นเพื่อลดการกระจายข้อมูล [7]

4. ขั้นตอนการขุดค้นข้อมูล (Data Mining) ขั้นตอนนี้ผู้วิจัยใช้วิธีการผสมผสานอัลกอริทึมในการแบ่งกลุ่ม (Cluster) โดยใช้เทคนิคเคมีน [8]

ขั้นตอนที่ 1 ใช้เทคนิคคลัสเตอร์แบบ K-Means Clustering ในการแบ่งกลุ่มซึ่งผู้วิจัยทำการพัฒนาเครื่องมือสำหรับการทดลอง วิธีการแบ่งกลุ่มแบบ K-Means โดยใช้ SQL Server เหมือนข้อมูล Add-Ins for Office 365 เป็นเครื่องมือสร้างหลังจากนำข้อมูลไปทำการแบ่งกลุ่มด้วยวิธีการของ K-Means ในแต่ละลำดับขั้น (Level) ของการแบ่งกลุ่มนั้น จะมีการเก็บค่าส่วนเบี่ยงเบนมาตรฐานไว้ทุกครั้ง จากนั้นจะทำการเลือกลำดับขั้นของลำดับขั้นที่ให้ค่าส่วนเบี่ยงเบนมาตรฐานที่น้อยที่สุด นำค่าจำนวนกลุ่มของลำดับขั้นนั้นไปเป็นค่า K เพื่อส่งต่อเป็นค่า Input ให้ ผลที่ได้ก็คือ ข้อมูลลักษณะของลูกค้าตามตัวแปรที่พิจารณาโดยแบ่งกลุ่มตามลักษณะในการสั่งซื้อสินค้า และปัจจัยอื่น ๆ ตามที่กล่าวมาแล้วในเบื้องต้น โดยซอฟต์แวร์ที่ใช้ คือ SQL Server เหมือนข้อมูล Add-Ins for Office 365 ทำเหมือนข้อมูล

K-means หรือเรียกอีกอย่างหนึ่งว่าการวิเคราะห์ข้อมูลให้กลุ่มแบบไม่เป็นขั้นตอน (Nonhierarchical Cluster Analysis) หรือ การแบ่งส่วน (Partitioning)

- เป็นอัลกอริทึมเทคนิคการเรียนรู้โดยไม่มีผู้สอนที่ง่ายที่สุด เพราะเป็นการแก้ปัญหาการจัดกลุ่มที่รู้จักกันทั่วไป โดยอัลกอริทึม K-Means จะตัดแบ่ง

(Partition) วัตถุออกเป็น K กลุ่ม

- แทนค่าแต่ละกลุ่มด้วยค่าเฉลี่ยของกลุ่ม ซึ่งใช้เป็นจุดศูนย์กลาง (Centroid) ของกลุ่มในการวัดระยะห่างของข้อมูลในกลุ่มเดียวกัน [8]

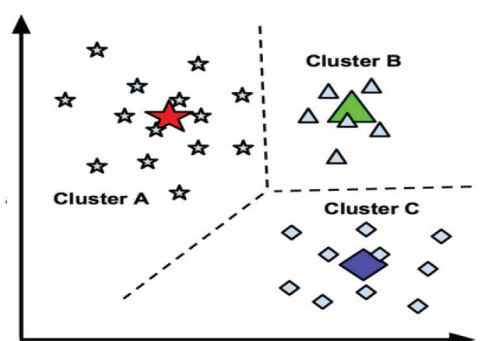
K-means Clustering Algorithm (Method)

1) กำหนดหรือสุ่มค่าเริ่มต้น จำนวน k ค่า (กลุ่ม) และกำหนดจุดศูนย์กลางเริ่มต้น k จุด เรียกว่า cluster center หรือ (centroid)

2) นำวัตถุทั้งหมดจัดเข้ากลุ่ม โดยทำการหาค่าระยะห่างระหว่างข้อมูลกับจุดศูนย์กลาง หากข้อมูลไหนใกล้ค่าจุดศูนย์กลางตัวไหนที่สุดอยู่กลุ่มนั้น

3) หาค่าเฉลี่ย (Mean) แต่ละกลุ่มให้เป็นค่าจุดศูนย์กลางใหม่

4) ทำซ้ำข้อ 2) จนกระทั่งค่าเฉลี่ยหรือจุดศูนย์กลางในแต่ละกลุ่มจะไม่เปลี่ยนแปลง [9] [10]



ภาพที่ 4 K-means Clustering Algorithm [9]

จากภาพที่ 4 เป็นการจัดกลุ่มในขั้นตอนแรกโดยที่กำหนดจำนวนกลุ่ม 3 กลุ่ม และกำหนดจุดศูนย์กลางเริ่มต้น ซึ่งใช้สัญลักษณ์ต่าง ๆ แทนจุดศูนย์กลางของแต่ละกลุ่มทั้ง 3 กลุ่ม จากนั้นวัตถุจะถูกกำหนดให้เข้ากลุ่มที่มีจุดศูนย์กลางอยู่ใกล้วัตถุมากที่สุด ทำซ้ำไปเรื่อย ๆ จนกระทั่งจุดศูนย์กลางไม่เปลี่ยนแปลงจึงจะได้ผลลัพธ์สุดท้าย

3.2 ข้อมูลสำหรับการทดลอง (Data set)

ข้อมูลตัวแปรที่ทำการทดลองจากฐานข้อมูลเดิมที่ต้องการจะนำไปใช้ประโยชน์ เป็นข้อมูลที่เป็นปัจจุบัน และเป็นข้อมูลที่น่าสนใจ ที่ใช้ในงานศึกษานี้ มีทั้งหมด 12 ตัวแปร ดังภาพที่ 5



ID	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner	Cars	Commute Distance	Region	Age	Purchased Bike
12496	Married	Female	40000	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	42	No
24107	Married	Male	30000	3	Partial College	Clerical	Yes	1	0-1 Miles	Europe	43	No
14177	Married	Male	80000	5	Partial College	Professional	No	2	2-5 Miles	Europe	60	No
24381	Single	Male	70000	0	Bachelors	Professional	Yes	1	5-10 Miles	Pacific	41	Yes
25597	Single	Male	30000	0	Bachelors	Clerical	No	0	0-1 Miles	Europe	36	Yes
13507	Married	Female	10000	2	Partial College	Manual	Yes	0	1-2 Miles	Europe	50	No
27974	Single	Male	160000	2	High School	Management	Yes	4	0-1 Miles	Pacific	33	Yes
19364	Married	Male	40000	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	43	Yes
22155	Married	Male	20000	2	Partial High School	Clerical	Yes	2	5-10 Miles	Pacific	58	No
19280	Married	Male	20000	2	Partial College	Manual	Yes	1	0-1 Miles	Europe	48	Yes
22173	Married	Female	30000	3	High School	Skilled Manual	No	2	1-2 Miles	Pacific	54	Yes
12697	Single	Female	90000	0	Bachelors	Professional	No	4	10+ Miles	Pacific	36	No
11434	Married	Male	170000	5	Partial College	Professional	Yes	4	0-1 Miles	Europe	55	No
25323	Married	Male	40000	2	Partial College	Clerical	Yes	1	1-2 Miles	Europe	35	Yes
23542	Single	Male	60000	1	Partial College	Skilled Manual	No	1	0-1 Miles	Pacific	45	Yes
20870	Single	Female	10000	2	High School	Manual	Yes	1	0-1 Miles	Europe	38	Yes
23316	Single	Male	30000	3	Partial College	Clerical	No	2	1-2 Miles	Pacific	59	Yes
12610	Married	Female	30000	1	Bachelors	Clerical	Yes	0	0-1 Miles	Europe	47	No
27183	Single	Male	40000	2	Partial College	Clerical	Yes	1	1-2 Miles	Europe	35	Yes
25940	Single	Male	20000	2	Partial High School	Clerical	Yes	2	5-10 Miles	Pacific	55	Yes

ภาพที่ 5 ฐานข้อมูลที่ใช้สำหรับวิเคราะห์ การแบ่งกลุ่มสำหรับการบริหารลูกค้าสัมพันธ์ [11]

ข้อมูลสำหรับวิเคราะห์ฐานข้อมูลส่วนตัวลูกค้า ระบบฐานที่นำมาจากข้อมูลตัวอย่างบริษัทของผู้จัดจำหน่ายสินค้าออนไลน์บนระบบอีคอมเมิร์ซแห่งหนึ่ง (Microsoft SQL Server Product Samples: Database) [11] ที่ใช้ในการทดลองจากฐานข้อมูลการของลูกค้าทั้งหมด ประเภทรถจักรยานจำนวน 1,000 คน นำเอาข้อมูลเกี่ยวกับกลุ่มตัวอย่างที่นำมาวิเคราะห์มาทำการแบ่งกลุ่ม โดยกำหนดค่าต่าง ๆ และปรับค่าตัวแปรต่าง ๆ เพื่อกำจัดความแตกต่างของข้อมูล [10] ทั้ง 12 ตัวแปร อาทิ ข้อมูลสถานภาพ ข้อมูลเพศ ข้อมูลรายได้ต่อเดือน จำนวนบุตร การสำเร็จการศึกษา อาชีพ ระยะทางระหว่างที่พัก-ที่ทำงาน ภูมิลำเนาและจำนวนรถที่เป็นเจ้าของ เป็นต้น

4. ผลการวิจัย

จากการวิเคราะห์กลุ่มลูกค้าเพื่อค้นหารูปแบบเฉพาะของปัจจัยการส่งเสริมการขายเพื่อให้ทราบทิศทางในการนำเสนอได้ตรงกับกลุ่มเป้าหมายและสอดคล้องกับพฤติกรรมลูกค้า ซึ่งจะนำไปสู่กลยุทธ์ทางการตลาดที่มีประสิทธิภาพสำหรับการบริหารลูกค้าสัมพันธ์และเหมาะกับกลุ่มผู้บริโภค ได้เป็น 7 กลุ่ม และทดสอบด้วยการแบ่งกลุ่มตัวแปรโดยใช้เทคนิค Cluster Analysis ในแบบเคมีน ปรากฏผลดังภาพที่ 6

Category	Solution	Value	Relative Importance
Category 2	Cars	0	
Category 2	Education	Graduate Degree	
Category 2	Income	Low:39050 - 71062	
Category 2	Commute Distance	0-1 Miles	
Category 2	Purchased Bike	Yes	
Category 2	Age	Low:37 - 46	
Category 2	Age	Very Low: 37	
Category 2	Occupation	Skilled Manual	
Category 2	Children	1	
Category 2	Home Owner	Yes	
Category 2	Marital Status	Married	

ภาพที่ 6 การประมวลผลและผลลัพธ์จาก Analysis server จากข้อมูล 1,000 แถว

ข้อมูลที่จะใช้สำหรับงานเชื่อมโยงของโคลเอนต์เหมือนข้อมูล หรือเครื่องมือวิเคราะห์การสั่งซื้อสินค้า คุณสามารถสร้างแบบจำลองเพื่อทำนายความสัมพันธ์ระหว่างหมวดหมู่หรือระหว่างผลิตภัณฑ์ ออกเป็นกลุ่ม ๆ ที่มีลักษณะคล้ายกันมาทำการ Cluster analysis ส่งค่าการประมวลผล ผ่านการ Auto detect จากระบบฐานที่นำมาจากข้อมูลตัวอย่างบริษัทของผู้จัดจำหน่ายสินค้าออนไลน์บนระบบอีคอมเมิร์ซแห่งหนึ่งที่ใช้ในการทดลองจากฐานข้อมูลการสั่งซื้อสินค้าของลูกค้าทั้งหมดของรถจักรยาน จำนวน 1,000 คน พบว่า มีการแบ่งกลุ่มคุณลักษณะตัวแปรที่เหมือนกันไว้อยู่ในกลุ่มเดียวกัน ซึ่งได้การแบ่งกลุ่มเป็น 7 กลุ่ม โดยมีการยกตัวอย่างมา 3 กลุ่ม ตามคุณสมบัติเด่น ๆ ของแต่ละกลุ่ม ดังภาพที่ 7 – 9

กลุ่มที่ 1

Category Name	Row Count
Category 1	189
Category 2	141
Category 3	158
Category 4	149
Category 5	126
Category 6	129
Category 7	108

ภาพที่ 7 การแจกแจงของค่าคุณสมบัติเด่น ๆ กลุ่มที่ 1

โดยสามารถบอกคุณสมบัติเด่น ๆ แต่ละหมวดหมู่ อาทิ กลุ่มที่ 1 ตัวที่ใช้แยกคุณสมบัติความสำคัญแรก คือตัวรายได้ ต่ำกว่า 39,050 เหรียญ และ รองลงมาคือ อยู่ในภูมิภาคยุโรป

กลุ่มที่ 2

Category	Column	Value	Relative Importance
Category 3	Age	Very Low< 37	
Category 3	Children	0	
Category 3	Education	Partial College	
Category 3	Marital Status	Single	
Category 3	Income	Very Low< 39050	
Category 3	Commute Distance	5-10 Miles	
Category 3	Home Owner	No	
Category 3	Cars	2	
Category 3	Occupation	Manual	
Category 3	Cars	1	
Category 3	Occupation	Skilled Manual	
Category 3	Education	Partial High School	
Category 3	Purchased Bike	No	
Category 3	Education	High School	
Category 3	Commute Distance	1-2 Miles	

ภาพที่ 8 การแจกแจงของค่าคุณสมบัติเด่น ๆ กลุ่มที่ 2

คุณสมบัติเด่น ๆ คือ เป็นกลุ่มคนที่ไม่มีรถ การศึกษาค่อนข้างดี และมีรายได้อยู่ในระดับกลาง คือ 39,050-71,062 เหรียญ ซึ่งมากกว่า กลุ่มที่ 1

กลุ่มที่ 3

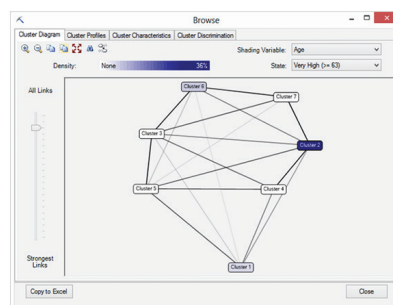
Category	Column	Value	Relative Importance
Category 1	Income	Very Low< 39050	
Category 1	Region	Europe	
Category 1	Occupation	Manual	
Category 1	Occupation	Clerical	
Category 1	Commute Distance	0-1 Miles	
Category 1	Cars	0	
Category 1	Children	2	
Category 1	Children	1	
Category 1	Education	Partial High School	
Category 1	Education	High School	
Category 1	Children	3	

ภาพที่ 9 การแจกแจงของค่าคุณสมบัติเด่น ๆ กลุ่มที่ 3

คุณสมบัติเด่น ๆ คือ กลุ่มคนที่มีช่วงอายุต่ำกว่า 37 ไม่มีบุตร การศึกษาอยู่ในระดับวิทยาลัยสถานะโสด และรายได้อยู่ในระดับต่ำกว่า 39,050 เหรียญ

4.1 การวิเคราะห์ข้อมูลจากกลุ่มประชากรทั้งหมด

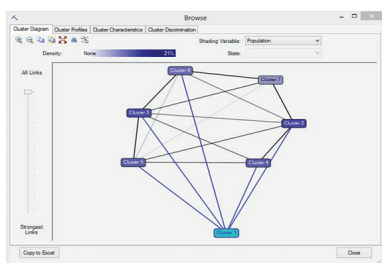
จากการวิเคราะห์กลุ่มประชากรทั้งหมดที่มีความสัมพันธ์มากเชื่อมโยงกันของทั้ง 7 กลุ่มดังภาพที่ 10



ภาพที่ 10 การแบ่งกลุ่มการวิเคราะห์ข้อมูลจากกลุ่มประชากร

จากการทดลอง การแบ่งกลุ่มข้อมูลลูกค้าที่สั่งซื้อสินค้าจากระบบ E-commerce ประเภทอุปกรณ์กีฬาจักรยาน จาก 1000 คน พบว่าแต่ละกลุ่มมีความเชื่อมโยงกันมากดังภาพ สามารถสร้างแบบจำลองเพื่อทำนายความสัมพันธ์ระหว่างกลุ่มหรือระหว่างผลิตภัณฑ์ จากชุดข้อมูลออกเป็นกลุ่ม ๆ ที่มีลักษณะคล้ายกัน มาทำการ K-mean Cluster analysis ส่งค่าการประมวลผลพบว่า มีการแบ่งกลุ่มได้ 7 กลุ่ม โดยกลุ่มที่มีความสัมพันธ์กันมากจะมีเส้นความเชื่อมโยงจะมีสีเข้ม หากกลุ่มมีความสัมพันธ์น้อยหรือแทบจะไม่มีความสัมพันธ์เส้นจะมีน้ำหนักเบา

จากการทดลองวิเคราะห์การเปลี่ยนแปลงกลุ่มตัวแปร ทั้ง 7 กลุ่ม โดยเลือกการแบ่งกลุ่มตามช่วงอายุ มีการเปลี่ยนแปลงความสัมพันธ์ตามเส้นความเชื่อมโยง ดังภาพที่ 11



ภาพที่ 11 การแบ่งกลุ่มข้อมูลตามช่วงอายุ

จากการทดลองวิเคราะห์การเปลี่ยนแปลงกลุ่มตัวแปร โดยเลือกการแบ่งกลุ่มตามช่วงอายุมากกว่า 63 ปี พบว่ากลุ่มคนที่มีอายุมากจะเป็นกลุ่มลูกค้าที่น้อย ซึ่งถ้ามีการเพิ่มความเข้มของแต่ละกลุ่มจะมีการกระจายค่าของการแบ่งกลุ่มมากขึ้น

5. อภิปรายผลการวิจัย

กลุ่มที่มีลักษณะคล้ายกันมาทำการ Cluster analysis ส่งค่าการประมวลผล จากชุดข้อมูลลูกค้า 1,000 คน แบ่งกลุ่มได้คุณลักษณะเด่นของตัวแปรได้ 12 ตัวแปร ที่มีคุณลักษณะเด่นที่สุด 2 ลำดับ จากการจัดกลุ่ม Category ดังนี้

กลุ่มที่ 7 อยู่ในช่วงอายุที่มีค่าเฉลี่ยสูงสุด = 64.2 ส่วนเบี่ยงเบนมาตรฐาน = 4.92 รองลงมาคือกลุ่มที่ 4 อยู่ในช่วงอายุ 50-65 ปี ค่าเฉลี่ย = 55.23 และส่วนเบี่ยงเบนมาตรฐาน = 5.03

กลุ่มที่ไม่ซื้อจักรยานมากที่สุดคิดเป็น กลุ่มที่ 6 = 95% รองลงมาคือกลุ่มที่ 3 = 94%

ส่วนใหญ่มีรถจักรยาน 2 คัน 2 คือ กลุ่มที่ 4 = 79% และมีรถ 1 คัน กลุ่มที่ 5 = 42%

ส่วนใหญ่ไม่มีบุตร กลุ่มที่ 6 คิดเป็น 100% รองลงมาคือ กลุ่มที่ 4 = 39%

ที่อยู่อาศัยมีระยะทางห่างจากชุมชน 0-1 ไมล์ กลุ่มที่ 3 = 57% และระยะห่าง 5-10 ไมล์ กลุ่มที่ 6 = 60% การศึกษาส่วนใหญ่จบปริญญาตรี กลุ่มที่ 5 = 55% รองลงมาจบปริญญาโท กลุ่มที่ 5 = 32%

เพศชาย ผลที่ได้จากการแบ่งกลุ่มมากที่สุดคือ กลุ่มที่ 3 = 53% เพศหญิง กลุ่มที่ 5 = 52%

ส่วนใหญ่มีบ้านของตัวเองจากการแบ่งกลุ่มมากที่สุดคือ กลุ่มที่ 7 = 83% ไม่มีบ้านเป็นของตัวเอง

อาทิ เข้าบ้าน หอพัก ฯลฯ คือกลุ่มที่ 3 = 48%

ประชากรส่วนใหญ่มีสถานะแต่งงาน ซึ่งส่วนใหญ่เป็นกลุ่มที่ 7 = 68% มีสถานะโสด ส่วนใหญ่เป็นกลุ่มที่ 6 = 63 %

อาชีพส่วนใหญ่เป็นผู้บริหารมีอาชีพมากที่สุดคือ กลุ่มที่ 1 = 52% รองลงมาคืออาชีพช่างเทคนิค กลุ่มที่ 6 = 83% ประชากรส่วนใหญ่อาศัยอยู่ในเขตอเมริกามากที่สุด กลุ่มที่ 6 = 100% รองลงมาอาศัยอยู่ในเขตยุโรป มากที่สุดในกลุ่มที่ 3 = 88%

ประชากรส่วนใหญ่มีรายได้มากที่สุดในกลุ่มที่ 2 มีค่าเฉลี่ยสูงสุด = 103,273.08 และส่วนเบี่ยงเบนมาตรฐาน = 26,483.70 รองลงมาคือกลุ่มที่ 7 มีค่าเฉลี่ยสูงสุด = 65,877.19 และส่วนเบี่ยงเบนมาตรฐาน = 15,325.20

5.1 ภาพรวมการวิเคราะห์ลักษณะกลุ่มที่แบ่งประเภทได้ 12 ตัวแปร

จากการวิเคราะห์ลักษณะกลุ่มที่แบ่งประเภทได้ 12 ตัวแปร โดยแต่ละตัวแปรมีความแตกต่างกันในแต่ละ Category ดังภาพรวมการวิเคราะห์ลักษณะการแบ่งกลุ่มตามตัวแปร ดังภาพที่ 12

จากการวิเคราะห์ลักษณะกลุ่มที่แบ่งประเภทได้ 12 ตัวแปร โดยแต่ละตัวแปรมีความแตกต่างกันในแต่ละ Category การกราฟแท่ง สังเกตตัวแปรอาชีพสามารถบอกการเปลี่ยนแปลงและค่าที่มีความสำคัญสูงสุดจะถูกเลือก ที่ไม่สำคัญจะมีการเก็บไว้ในอื่น ๆ ตัวแปรอายุจะมีผลลัพธ์ที่แตกต่างจากตัวแปรอื่น ๆ เพื่อให้การแสดงค่าเป็นจำนวนกลุ่มช่วงอายุ ทำให้มีการกระจายตัวตามค่า ซึ่งการสร้างกราฟแผนภูมิจะปรากฏเฉพาะข้อมูลที่เป็นตัวเลขซึ่งสามารถนำไปใช้ประโยชน์ในการหาค่าทางสถิติ อย่างค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานสำหรับค่าแต่ละการแบ่งกลุ่ม [12]



Variables	States	Population Size 101	Cluster 1 Size 134	Cluster 2 Size 127	Cluster 3 Size 117	Cluster 4 Size 97	Cluster 5 Size 96	Cluster 6 Size 96	Cluster 7 Size 96
Age	77.59 44.00 25.00								
Care	0 1 2 Other								
Children	0 1 2 Other								
Commute Distance	0-1 Miles 1-2 Miles 2-3 Miles 3-4 Miles Other								
Education	Elementary Partial College Graduate Degree High School Other								
Gender	Male Female missing								
Home Owner	Yes No missing								
Income	151,919.43 57,400.00 10,000.00								
Marital Status	Married Single missing								
Occupation	Not Specified Self-employed Unemployed Other								
Purchased Bike	No Yes missing								
Region	North America Europe Pacific missing								

ภาพที่ 12 ภาพรวมการวิเคราะห์ลักษณะการแบ่งกลุ่ม

6. สรุปผลการวิจัย

วิจัยนี้เสนอทางเลือกเพื่อการวิเคราะห์ข้อมูลในเชิงธุรกิจ โดยจะใช้การแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์ โดยใช้ข้อมูลระบบฐานข้อมูลเกี่ยวกับลักษณะลูกค้าที่ใช้ในการทดลองจากฐานข้อมูลการสั่งซื้อสินค้าของลูกค้า ประเภทรถจักรยานจำนวน 1,000 คน นำมาวิเคราะห์หาทำการแบ่งกลุ่ม อาทิ ข้อมูลสถานภาพ ข้อมูลเพศ ข้อมูล รายได้ต่อเดือน จำนวนบุตร การสำเร็จการศึกษา อาชีพ ระยะทางระหว่างที่พัก-ที่ทำงาน ภูมิลำเนาและจำนวนรถที่เป็นเจ้าของ ซึ่งปัจจัยคุณค่าตราสินค้าและบริการมีอิทธิพลต่อการตัดสินใจซื้อรถจักรยานของผู้บริโภค คือ พฤติกรรมการแบ่งปันความรู้ด้านคุณภาพผลิตภัณฑ์ และนำเสนอรายการส่งเสริมการขายได้ตรงกลุ่มเป้าหมายและประเภทสินค้า [13] ด้านความสัมพันธ์ของคุณภาพระบบ ด้านคุณภาพบริการ การใช้งาน ความพึงพอใจของผู้ใช้ และประสิทธิผลด้านการให้บริการของระบบการบริหารลูกค้าสัมพันธ์ เพื่อการวิเคราะห์ข้อมูลลูกค้าสำหรับการบริหารลูกค้าสัมพันธ์เพื่อสร้างความพึงพอใจให้กับผู้ใช้บริการและสร้างความได้เปรียบทางการแข่งขันได้อย่างมีประสิทธิภาพ [14] จากการวิเคราะห์โดยใช้โดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน พบว่าสามารถสังเคราะห์และจัดกลุ่มตัวอย่างลูกค้าได้ทั้งสิ้น 7 กลุ่ม

ตามประสิทธิภาพการแบ่งกลุ่มลูกค้า โดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน สำหรับการบริหารลูกค้าสัมพันธ์และปัจจัยที่มีผลต่อการตัดสินใจซื้อรถจักรยานของผู้บริโภคในแต่ละภูมิลำเนา ซึ่งประกอบด้วย

กลุ่มที่ 1 ลักษณะคุณสมบัติเด่นที่ใช้ในการแบ่งกลุ่มที่มีความสำคัญสูงอันดับแรก คือตัวรายได้ต่ำกว่า 39,050 เหรียญ และคุณสมบัติเด่นรองลงมาคือ อยู่ในภูมิลำเนายุโรป

กลุ่มที่ 2 มีลักษณะคุณสมบัติเด่น ๆ คือ เป็นกลุ่มคนที่ไม่มียศ การศึกษาค่อนข้างดี และมีรายได้อยู่ในระดับกลาง คือ 39,050-71,062 เหรียญ ซึ่งมากกว่ากลุ่มที่ 1

กลุ่มที่ 3 มีลักษณะ อายุเป็นปัจจัยสำคัญ คือกลุ่มคนที่มีช่วงอายุต่ำกว่า 37 และไม่มีบุตร จบการศึกษาระดับวิทยาลัย

กลุ่มที่ 4 มีลักษณะคุณสมบัติเด่น ๆ คือ กลุ่มคนที่มีรายได้อยู่ในระดับกลาง คือ 39,050-71,062 เหรียญ ภูมิลำเนาอเมริกาเหนือโดยมีระยะทางในการเดินทาง 10 ไมล์ มีรถอยู่แล้วและอยู่ในช่วงวัย 55-65 ปี

กลุ่มที่ 5 มีลักษณะคุณสมบัติเด่น ๆ คือ กลุ่มคนที่มืออาชีพผู้บริหารมืออาชีพ ผู้เชี่ยวชาญด้านต่าง ๆ กลุ่มคนที่มีมียศ มีรายได้อยู่ในระดับดีถึงดีมาก คือ 97,111-127,371 เหรียญ โดยมีระยะทางในการเดินทาง 10 ไมล์ และมีภูมิลำเนาแถวแปซิฟิก

กลุ่มที่ 6 มีลักษณะคุณสมบัติเด่น ๆ คือ กลุ่มคนที่อยู่ในอาชีพช่างเทคนิค อยู่ในช่วงวัย 46-55 ปี เป็นเจ้าของรถ 1 คัน บุตร 2 คน และมีภูมิลำเนาแถวอเมริกาเหนือ

กลุ่มที่ 7 มีลักษณะคุณสมบัติเด่น ๆ คือ กลุ่มคนที่อยู่ในทำงานในตำแหน่งบริหาร มีรายได้อยู่ในระดับดี-ดีมาก คือ 97,111-127,371 เหรียญ และเป็นเจ้าของรถถึง 3-4 คัน

จากผลการทดลองทางด้านประชากรศาสตร์ และปัจจัยประสมทางการตลาดที่มีอิทธิพลต่อพฤติกรรมการซื้อจักรยานโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน การแบ่งกลุ่มลูกค้าที่มีลักษณะเหมือนกันไว้อยู่ในกลุ่มเดียวกัน ทำให้เราเข้าใจ



คุณลักษณะของลูกค้ามากขึ้น และทราบว่าลูกค้าสามารถสังเคราะห์และจำแนกออกมาได้ตามลักษณะอิทธิพลต่อทางพฤติกรรมของกลุ่มลูกค้า [15] ว่ามีประสิทธิภาพดีในการใช้เทคนิคการทำคลัสเตอร์แบบเคมีนในระดับ 7 กลุ่ม ตามลักษณะคุณสมบัติเด่น ๆ ในการกำหนดแผนการส่งเสริมการตลาดที่เฉพาะเจาะจงให้เกิดผลสูงสุดเพื่อกระตุ้นให้เกิดการซื้อขายและชักจูงใจให้ผู้บริโภคมีแรงจูงใจในการสั่งซื้อสินค้าและบริการ สำหรับการบริหารลูกค้าสัมพันธ์ ซึ่งจะทำให้สามารถกำหนดกลยุทธ์ทางการตลาดเพื่อจับกลุ่มเป้าหมายได้ถูกต้องแม่นยำและมีประสิทธิภาพมากขึ้น

7. เอกสารอ้างอิง

- [1] พลอย แยมเสนาะ. (2552). *กลยุทธ์การตลาด*. กรุงเทพฯ: ภาควิชาการออกแบบผลิตภัณฑ์ คณะศิลปกรรมศาสตร์ มหาวิทยาลัยบูรพา
- [2] โกศล พรประสิทธิ์เวช. (2552). *สร้างยอดขายอันดับ 1 ด้วย Prosoft CRM*. กรุงเทพฯ: บริษัท พิมพ์ดี จำกัด.
- [3] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Framework (pp. 82-88). In *KDD-96 Proceedings*.
- [4] วีระพงศ์ จันทน์สนาม. (2551). *ผลกระทบของการบริหารลูกค้าสัมพันธ์ต่อความจงรักภักดีของลูกค้าของธุรกิจพาณิชย์อิเล็กทรอนิกส์ในประเทศไทย* (วิทยานิพนธ์ ปริญญาบริหารธุรกิจมหาบัณฑิต). มหาสารคาม: มหาวิทยาลัยมหาสารคาม.
- [5] Zeithaml, V. A., & Bitner, M. J. (2000). *Services marketing: integrating customer focus across the firm* (2nd ed). Boston: Irwin/McGraw-Hill.
- [6] ปัทมา เทียงสมบุญ, และนิเวศ จิระวิจิตชัย. (2561). การพัฒนาระบบธุรกิจอัจฉริยะเพื่อสนับสนุนการพยากรณ์และการตัดสินใจของผู้บริหาร กรณีศึกษากลุ่มโรงพยาบาล. *Veridian E-Journal, Science and Technology Silpakorn University*, 5(4), 16-30.
- [7] ัญญาพร ชื่นมัจฉา, และนิเวศ จิระวิจิตชัย. (2559). การหาความสัมพันธ์จากฐานข้อมูลการซื้อขายของลูกค้าโดยใช้เทคนิค เอฟพี-โกรธ. *วารสารวิทยาศาสตร์และเทคโนโลยี มทร.ธัญบุรี*, 6(1), 122-131.
- [8] Berson A., Smith, S., & Thearling, K. (1999). *Building data mining applications for CRM*. USA: McGraw-Hill.
- [9] วิภาวรรณ บัวทอง. (2559). *การทำเหมืองข้อมูล*. ภูเก็ต: มหาวิทยาลัยราชภัฏภูเก็ต.
- [10] MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations (pp. 281-297). In *5th Berkeley Symposium on Mathematical Statistics and Probability*. California: University of California Press.
- [11] AdventureWorks sampledatabases. (2017). [Online], Retrieved from <https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks>. (15 February 2014).
- [12] กัลยา วานิชย์บัญชา. (2552). *การวิเคราะห์ข้อมูลหลายตัวแปร*, (พิมพ์ครั้งที่ 4). กรุงเทพฯ: ธรรมสาร.
- [13] อีสริย ตรีประเสริฐ. (2559). *ปัจจัยคุณค่าตราสินค้า พฤติกรรมการแบ่งปันความรู้ และนวัตกรรมผลิตภัณฑ์ที่มีอิทธิพลต่อการตัดสินใจซื้อออร์จักรยานของผู้บริโภคในเขตกรุงเทพมหานคร* (การค้นคว้าอิสระ). กรุงเทพฯ: มหาวิทยาลัยกรุงเทพ.
- [14] ศศินิภา ทิวาลัย. (2555). การวัดประสิทธิผลด้านการให้บริการของระบบการบริหารลูกค้าสัมพันธ์ออนไลน์ของธุรกิจโรงแรมโดยการใช้แบบจำลองของดีลอนแอนด์แม็คคีนและอิมเมตริก. *วารสารบริหารธุรกิจ เศรษฐศาสตร์และการสื่อสาร*, 7(1), 89-101.



- [15] อธิษฐ์ รุ่งเรือง. (2558). ปัจจัยส่วนประสมทางการตลาดที่มีอิทธิพลต่อพฤติกรรมการซื้อจักรยานปั่นเพื่อสุขภาพในเขตกรุงเทพมหานครและปริมณฑล (การค้นคว้าอิสระ). กรุงเทพมหานคร: มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี