การวิเคราะห์กฎความสัมพันธ์ของการขายผลไม้ : กรณีศึกษาบริษัท อินสตาการ์ท จำกัด

ศุภพงศ์ คงสวี[1] ,นุวีย์ วิวัฒนวัฒนา[2]

บทคัดย่อ

การวิเคราะห์กฎความสัมพันธ์ในการขายสินค้า เป็นสิ่งที่จะช่วยให้ฝ่ายการตลาดสามารถนำไปวางแผนธุรกิจ หรือ การจัดโปรโมชั่นได้ ซึ่งกฎความสัมพันธ์ (Association Rules) จะวิเคราะห์มาจากข้อมูลของการทำธุรกรรมในการซื้อสินค้าของผู้บริโภค ซึ่งในแต่ละช่วงเวลาจะมีความแตกต่างกัน เป็นผลมาจากกระแสนิยม และ สภาพแวดล้อมในช่วงเวลานั้น การวิเคราะห์กฎความสัมพันธ์จึงมีความจำเป็นในการวางแผนกลยุทธ์เพื่อการขายสินค้า เช่น การสร้างสินค้าที่ซื้อคู่กัน (Bundled Discounts) หรือเป็นการจัดชั้นวางสินค้า ที่จะสามารถวิเคราะห์จากการซื้อสินค้าในแต่ละกฎที่ซ้ำกัน โดยจะทำการวิเคราะห์จากค่าตัวแปรของการวิเคราะห์กฎความสัมพันธ์ ได้แก่ Support,Confident,Lift,Leverage,Conviction โดยผู้วิจัยได้เลือกการขายผลไม้ เนื่องจากกล้วย ในชั้นวางผลไม้ เป็นสินค้าที่มีจำนวนการสั่งซื้อมากที่สุด ซึ่งไม่สอดคล้องกับ จำนวนรายชื่อชนิดสินค้ากลุ่มผลไม้ในชั้นวาง ที่มีน้อยกว่าแผนกอื่น ผู้วิจัยจึงต้องการทราบรูปแบบจากกฎความสัมพันธ์เพื่อแนะนำให้ฝ่ายการตลาด โดยวิเคราะห์กฎความสัมพันธ์ของกลุ่มสินค้าผลไม้ เพื่อวิเคราะห์ความต้องการในการซื้อผลไม้คู่กัน หรือ คู่กับสินค้าประเภทอื่นที่อยู่ในแผนกเดียวกัน รวมทั้ง การจัดชั้นวางให้สินค้าที่มีความสัมพันธ์กันอยู่ใกล้กัน เพื่อความสะดวกในการหยิบสินค้า และมีโอกาสที่จะทำให้ยอดขายของกลุ่มผลไม้เพิ่มมากขึ้น ส่งผลให้สามารถเพิ่มรายชื่อชนิดสินค้าในชั้นวางกลุ่มผลไม้ได้ โดยผู้วิจัยกำหนดค่าขั้นต่ำของ Support เป็น 0.01 เนื่องจากจะได้กฎที่เยอะที่สุด และค่า Lift จะต้องมากกว่าหรือเท่ากับ 1 เนื่องจากเป็นมาตรฐานในการเลือกกฎ หากต่ำกว่า 1 กฎนั้นจะไม่น่าเชื่อถือ ผู้วิจัยพบว่ามีกฎที่น่าสนใจ จำนวน 8 กฎ โดยกฎที่แนะนำให้กับฝ่ายการตลาด จำนวน 2 กฎ โดยกฎที่ดีที่สุด คือ Organic Strawberries to Bag of Organic Bananas เนื่องจากมีค่า Confident มากถึง 0.28 และ ค่า Lift เป็น 1.77 รวมถึงค่า Support เป็น 0.03 ซึ่งกฎนี้มากกว่าขั้นต่ำที่กำหนด และมีค่าเชื่อมั่นของกฎในระดับที่ดี ทำให้มีโอกาสผู้บริโภคจะซื้อคู่กันมากขึ้น หากนำไปจัดโปรโมชั่น และ การจัดชั้นวางสินค้า จะเลือกจากกฎทั้ง 8 กฎที่มีการซื้อสินค้าซ้ำกันในแต่ละกฎ เพื่อนำสินค้าที่ซ้ำกันในแต่ละกฎมาอยู่ในชั้นวางเดียวกัน โดยการจัดชั้นวางที่ดีที่สุด คือ Limes และ Large Lemon เนื่องจากค่า Lift มากถึง 3.15 แสดงให้เห็นว่า ความน่าจะเป็นที่ลูกค้าจะหยิบ Limes แล้วหยิบ Large สูงกว่าเกณฑ์ขั้นต่ำที่กำหนด มากเป็น 3 เท่า ซึ่งถือว่าข้อมูลจากการวิเคราะห์กฎความสัมพันธ์มีประโยชน์กับฝ่ายการตลาด

คำสำคัญ : การวิเคราะห์กฎความสัมพันธ์,การซื้อสินค้า,สินค้าที่ซื้อคู่กัน,การจัดชั้นวาง

---

[1] หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

[2] คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author:Tel.: 092-8682971 E-mail address: suppapong.kongsawee@g.swu.ac.th

Association rules analysis of fresh fruits : A case study of Instacart company

Supapong Kongsawee[1], Nuwee Wiwatwattana[2]

Abstract

Analyzing the association rules in product sales is crucial for marketing teams to plan their business strategies and promotions. Association rules are derived from transactional data of consumer product purchases, which vary over time due to trends and environmental factors. Analyzing these rules is essential for sales strategy planning, such as creating bundled discounts or organizing product placement based on the repeated patterns of purchases. This analysis involves variables such as support, confidence, lift, leverage, and conviction. In this study, the researcher chose to focus on fruit sales, specifically bananas, as they are the most frequently purchased product in the fruit section, despite having fewer varieties compared to other sections. The purpose of this research was to identify the patterns in association rules and provide recommendations to the marketing team. By analyzing the association rules of the fruit section, the study aimed to determine the demand for buying pairs of fruits or combining fruits with other products within the same section.The study aimed to analyze the relationship between the sales of fruits in a supermarket and the organization of their display shelves, with the objective of improving the sales of the fruits. The minimum support threshold was set at 0.01 to obtain the highest number of rules, and the minimum lift threshold was set at 1 as a standard for rule selection. The study found 8 interesting rules, with the two best rules being "Organic Strawberries to Bag of Organic Bananas" with a confidence value of 0.28 and a lift value of 1.77, and "Limes and Large Lemon" with a lift value of 3.15. These rules suggest a higher probability of customers buying the paired products together. The study recommended using the 8 rules to determine which products to place on the same display shelf, based on the commonality of product purchases within each rule. The study found that the display of Limes and Large Lemon together resulted in a higher purchase rate than the minimum threshold by a factor of 3. Overall, the study's findings are beneficial to both the marketing

Keywords : association rule analysis, purchasing behavior, bundled discounts, product placement.

[1] Data Science, Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

[2] Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

* Corresponding author:Tel.: 092-8682971 E-mail address: suppapong.kongsawee@g.swu.ac.th

บทนำ

แอปพลิเคชันการซื้อขายสินค้าออนไลน์ กำลังเติบโตอย่างมากในปัจจุบัน และสามารถเข้าถึงคนได้ทุกพื้นที่ ไม่จำเป็นต้องมีที่พักอาศัยอยู่ใกล้เคียงกับร้านค้า เพียงแค่กดสั่งสินค้าออนไลน์ ก็สามารถได้รับสินค้าได้ตามที่ต้องการ ซึ่งในปัจจุบันฝ่ายการตลาดของร้านค้า จำเป็นต้องทราบว่า การกำหนดกลยุทธ์ทางการตลาด จากข้อมูลธุรกรรมการซื้อสินค้าของลูกค้า เพราะว่าการเลือกซื้อสินค้าผ่านแอปพลิชั่นการซื้อขายสินค้าออนไลน์เพิ่มขึ้น ทำให้คู่แข่งทางการตลาดเพิ่มขึ้นด้วย การวิเคราะห์กฎความสัมพันธ์ในการขายสินค้า เป็นสิ่งที่จะช่วยให้ฝ่ายการตลาดสามารถนำไปวางแผนธุรกิจ หรือ การจัดโปรโมชั่นได้ ซึ่งกฎความสัมพันธ์ (Association Rules) สามารถนำไปใช้ประโยชน์ได้หลากหลาย เช่น การสร้างสินค้าที่ซื้อคู่กัน (Bundled Discounts) ,การจัดชั้นวางสินค้า หรือวิเคราะห์ความต้องการของสินค้าตามหลักภูมิศาสตร์ โดยจะใช้ข้อมูลของผู้ซื้อและข้อมูลการซื้อสินค้าของผู้บริโภค โดยจะทำการวิเคราะห์จากค่าตัวแปรของการวิเคราะห์กฎความสัมพันธ์ ได้แก่ Support,Confident,Lift,Leverage,Conviction โดยในงานวิจัยนี้ผู้วิจัยได้เลือกการขายผลไม้ และวิเคราะห์ความต้องการในการซื้อผลไม้คู่กัน หรือ คู่กับสินค้าประเภทอื่น รวมทั้ง การจัดชั้นวางให้สินค้าที่มีความสัมพันธ์กันอยู่ใกล้กัน โดยใช้ข้อมูล บริษัท อินสตาการ์ด จำกัด (Instacart Company) เป็นแอปพลิเคชันการซื้อขายสินค้าออนไลน์ โดยจะเลือกสินค้าจากร้านของชำในประเทศสหรัฐอเมริกา ก่อตั้งเมื่อ ค.ศ. 2012 ซึ่งเมื่อผู้ซื้อสินค้ากดสั่งสินค้า เจ้าหน้าที่ของบริษัทจะไปรับสินค้าตามคำสั่งซื้อที่ร้านที่ลูกค้าเลือก ในระยะเวลาที่ผู้ซื้อกำหนด โดยเน้นความไวการจัดส่ง โดยแอปพลิเคชันจะเน้นสินค้าที่เป็นของสด และ สินค้าที่จำเป็นในครัวเรือน ซึ่งในแต่ละวันจะมีการทำธุรกรรมการซื้อขายจำนวนมาก หากใช้มนุษย์ในการวิเคราะห์ข้อมูลก็จะล่าช้า และ อาจจะมีความไม่แม่นยำได้ ผู้วิจัยจึงเลือกใช้ข้อมูลธุรกรรมการซื้อขายสินค้า วิเคราะห์กฎความสัมพันธ์ด้วย Apriori Algorithm ในการสร้างกฎความสัมพันธ์และวัดค่าตัวแปรของแต่ละกฎ เพื่อนำมาวิเคราะห์โปรโมชั่นและการจัดชั้นวางได้อย่างรวดเร็ว และ แม่นยำ เพื่อให้สอดคล้องกับกระแสนิยม หรือ พฤติกรรมผู้บริโภคที่เปลี่ยนไป หากทีมการตลาดสามารถวางแผนกลยุทธ์ได้ตรงกลุ่มเป้าหมาย การซื้อขายสินค้าก็มีโอกาสเพิ่มมากขึ้น โดยเฉพาะกลุ่มผลไม้ ที่มีระยะเวลาการเก็บได้ไม่นาน เสี่ยงต่อการเน่าเสียได้ง่าย หากกำหนดโปรโมชั่นเพิ่มยอดขายให้สินค้ากลุ่มผลไม้ จะสามารถเพิ่มกำไรให้กับร้านค้า เพิ่มการเข้าถึงแอปพลิเคชั่นให้กับบริษัท ร้านค้าอื่นก็อาจมีความสนใจเข้ามาเป็นพันธมิตรในการขายสินค้าในแอปพลิเคชั่นได้

งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ทำการศึกษางานวิจัยที่เกี่ยวข้องกับการวิเคราะห์กฎความสัมพันธ์ของธุรกรรมซื้อขายสินค้า งานวิจัย [1] ได้ศึกษารูปแบบกฎความสัมพันธ์ของการซื้อของลูกค้าก่อนหน้า ที่สามารถช่วยเพิ่มกลยุทธ์ทางการตลาดได้ เพราะจะทำความเข้าใจว่าสินค้าที่ผู้บริโภคมักซื้อมีรูปแบบใด โดยข้อมูลของยอดขายที่ดีที่สุดของข้อมูลธุรกรรมจะไม่ถูกนำกลับมาใช้ใหม่ จัดเก็บเป็นไฟล์เก็บถาวรและใช้เพื่อจัดทำรายงานการขาย วัตถุประสงค์ทางธุรกิจในบริษัทค้าปลีก เช่น มินิมาร์ท มินิมาร์เก็ตสามารถวิเคราะห์พฤติกรรมการจับจ่ายของลูกค้าเพื่อค้นหาความสัมพันธ์และความสัมพันธ์ระหว่างสินค้าในตะกร้าสินค้า โดยใช้เทคนิคการวิเคราะห์กฎความสัมพันธ์ ที่คาดว่าจะให้ประโยชน์แก่บริษัทในการปรับปรุงกลยุทธ์ทางธุรกิจ โดยศึกษารูปแบบพฤติกรรมผู้บริโภค เพื่อควบคุมการจัดวางสินค้าหรือการจัดวางสินค้า โดยมีการกำหนด Minimum Support และ Confident คือ 0.1,0.65 ตามลำดับ โดยได้กฎความสัมพันธ์ 10 กฎ โดย มี 4 กฎที่มี Support 0.01 และ Confident เป็น 1 เช่น ผู้บริโภคที่ซื้อเนสกาแฟคอฟฟี่ครีม 200 มล. มักซื้อคู่กับ ผู้บริโภคจะซื้อน้ำแร่พรีมา 600 มล. เป็นต้น

งานวิจัย [2] มีวัตถุประสงค์เพื่อวิเคราะห์กฎความสัมพันธ์ของข้อมูลตะกร้าสินค้าเพื่อช่วยให้ผู้ค้าปลีกในพื้นที่เข้าใจผู้บริโภค โดยพฤติกรรมการซื้อโดยการหารูปแบบ 3 ประเภทของฤดูกาลในอุตสาหกรรมค้าปลีก ได้แก่ ช่วงพีค ฤดูกาลปกติ และฤดูกาลเฉื่อย โดยเก็บข้อมูลในประเทศอินโดนีเซีย ปี ค.ศ.2017 ช่วงพีค หมายถึง “เทศกาลรอมฎอน” คือ พฤษภาคม มิถุนายน พฤศจิกายน และ ธันวาคม และ ฤดูกาลปกติ หมายถึง “เทศกาลคริสต์มาส. คือ มิถุนายน และ ฤดูการหย่อน หมายถึง “เดือนที่ผู้บริโภคลดลง” คือ มกราคม กุมภาพันธ์ กรกฎาคม และ สิงหาคม ผู้วิจัยได้ใช้อัลกอริทึม Apriori ในการวิเคราะห์กฎความสัมพันธ์ โดยจะประมวลผลข้อมูลเพื่อค้นหาคู่กลุ่มที่ขายร่วมกัน เป็นคู่และสามรายการ โดยกำหนด Min Confident 0.8 และ Min Support 0.22 การวิเคราะห์กฎความสัมพันธ์ได้ผลลัพธ์ของฤดูกาลเฉื่อย 34 กฎ , ช่วงพีค 19 กฎ และ ฤดูกาลปกติ 32 กฎ โดยมีการเลือกพิจารณากฎจากค่า Lift ซึ่งงานวิจัยได้ค้นพบ 6 รูปแบบที่ไม่ซ้ำกันซึ่งพบได้ในแต่ละฤดูกาล โดยรูปแบบความสัมพันธ์ที่ค้นพบในแต่ละฤดูกาล

ผู้วิจัยได้ทำการศึกษางานวิจัยที่เกี่ยวข้องแล้วสังเคราะห์ได้ว่า การวิเคราะห์กฎความสัมพันธ์นั้นสามารถช่วยต่อยอดแผนกลยุทธ์ด้านการขายสินค้าให้กับร้านค้าได้ รวมถึงเป็นกลบยุทธ์ที่สามารถปรับเปลี่ยนได้ตามฤดูกาล ช่วงเวลา หรือ ภูมิศาสตร์ที่เปลี่ยนไปตลอดเวลา การปรับปรุงกลยุทธ์ทางธุรกิจด้วยการวิเคราะห์กฎความสัมพันธ์ จะมีความแม่นยำและความรวดเร็วมากกว่ามนุษย์คำนวณ ทำให้สามารถเข้าถึงความต้องการของผู้บริโภคได้อย่างทันท่วงที งานวิจัยส่วนใหญ่มักเลือกใช้ Apriori Algorithm ในการวิเคราะห์กฎความสัมพันธ์ และมีการกำหนดค่า Support,Confident,Lift เป็นเบื้องต้น โดยเลือกกฎที่น่าสนใจเพื่อนำเสนอให้แก่ฝ่ายการตลาดต่อไป ผู้วิจัยจึงเล็งเห็นความสำคัญของการวิเคราะห์กฎความสัมพันธ์ โดยใช้ Apriori Algorithm ในการวิเคราะห์กลุ่มสินค้าที่เน่าเสียได้ง่าย โดยเฉพาะกลุ่มของสด ซึ่งสอดคล้องกับวัตถุประสงค์ของบริษัท ที่เน้นกลุ่มสินค้าประเภทของสด โดยผู้วิจัยได้เลือก ผลไม้ เนื่องจากจากการสำรวจข้อมูลพบว่า ปริมาณการซื้อสินค้ามากที่สุด แต่จำนวนประเภทสินค้าอยู่ในลำดับที่ 10 เรียงจากมากไปน้อย หากสามารถทราบกฎความสัมพันธ์การซื้อสินค้าคู่กัน อาจจะสามารถปรับเปลี่ยนการจัดชั้นวางและการเติมสต็อกสินค้า เพิ่มประเภทสินค้า เพื่อสามารถนำไปสู่ยอดขายที่เพิ่มมากขึ้น

วิธีดำเนินการ

ขั้นตอนที่ 1 : การเก็บรวบรวมข้อมูล (Data Acquisition)

ในงานวิจัยนี้ได้ใช้ข้อมูลธุรกรรมการซื้อขายในแอปพลิเคชันขายสินค้าออนไลน์ Instracart โดยใช้ชุดข้อมูลปี 2017 ได้เผยแพร่ชุดข้อมูลที่มีคำสั่งซื้อของชำกว่า 3 ล้านรายการจากผู้ใช้ Instacart มากกว่า 200,000 ราย สำหรับผู้ใช้แต่ละรายจะมีคำสั่งซื้อระหว่าง 4 ถึง 100 รายการ โดยมีลำดับผลิตภัณฑ์ที่ซื้อในแต่ละคำสั่งซื้อ นอกจากนี้ยังระบุสัปดาห์และชั่วโมงของวันที่มีการสั่งซื้อและการเวลาที่เกี่ยวข้องระหว่างคำสั่งซื้อ โดยชุดข้อมูลจะถูกเผยแพร่ในเว็บไซต์สาธารณะจาก Kaggle.com (Pspark , 2017) โดยชุดข้อมูลประกอบด้วย 15 แอททริบิวต์ 6 ไฟล์ ได้แก่ ไฟล์ Aisles มีข้อมูลทั้งหมด 134 แถว ,ไฟล์ Departments มีข้อมูลทั้งหมด 21 แถว,ไฟล์ Order_products_test มีข้อมูลทั้งหมด 1,048,575 แถว,ไฟล์ Order_products_train มีข้อมูลทั้งหมด 1,384,617 แถว,ไฟล์ Orders มีข้อมูลทั้งหมด 3,421,083 แถว,ไฟล์ Products มีข้อมูลทั้งหมด 49,688 แถว โดยแสดงคุณลักษณะของตัวแปร ดังนี้

1	Aisle ID	แสดงข้อมูลระบุตัวตนของชั้นวาง
2	Aisle name	แสดงชื่อของชั้นวาง
3	Department ID	แสดงข้อมูลระบุตัวตนของแผนก
4	Department name	แสดงชื่อของแผนก
5	Order ID	แสดงข้อมูลระบุตัวตนของการสั่งซื้อ
6	Product ID	แสดงข้อมูลระบุตัวตนของสินค้า
7	Add to cart order	แสดงข้อมูลการสั่งซื้อในรถเข็น
8	Reordered	แสดงข้อมูลการสั่งซื้อซ้ำของสินค้า
9	User ID	แสดงข้อมูลระบุตัวตนของผู้ใช้
10	Order number	แสดงข้อมูลหมายเลขตามลำดับสินค้าในการสั่งซื้อ
11	Order_dow	แสดงข้อมูลการซื้อสินค้าวันจันทร์ ถึง วันอาทิตย์
12	Order hour of day	แสดงข้อมูลการสั่งซื้อสินค้าใน 24 ชั่วโมง
13	Days since prior order	แสดงข้อมูลวันแลละเวลาในการกลับมาซื้อสินค้าซ้ำของผู้ใช้
14	Product Name	แสดงชื่อของสินค้า

ตาราง 1 แสดงคุณลักษณะของตัวแปร

ขั้นตอนที่ 2 : การสำรวจข้อมูล (Exploratory Data Analysis)

ผู้วิจัยจะใช้เครื่องมือในการสำรวจข้อมูลธุรกรรมการซื้อสินค้า โดย Google Colab,Looker Studio และ Tableau เพื่อสามารถแสดงผลแบบภาพรวม เพื่อตั้งสมมติฐานในการหาข้อมูลเชิงลึก โดยการเลือกกลุ่มในการวิเคราะห์กฎความสัมพันธ์ โดยใน Google Colab จะเลือกใช้ไลบรารี่ Pandas และ Numpy ร่วมกับ Looker Studio เพื่อสำรวจข้อมูลที่อยู่ในแต่ละไฟล์ และรูปแบบที่น่าสนใจที่คาดว่าอาจมีประโยชน์ในการทำงานวิจัย ดังนี้

	product_name	Record C...	departm...	aisle_id
1.	#2 Coffee Filters	1	7	26
2.	#2 Cone White Coffee Filters	1	7	26
3.	#2 Mechanical Pencils	1	17	87
4.	#4 Natural Brown Coffee Filters	1	7	26
5.	& Go! Hazelnut Spread + Pretzel Sticks	1	13	88
6.	(70% Juice!) Mountain Raspberry Juice Squeeze	1	7	98
7.	+Energy Black Cherry Vegetable & Fruit Juice	1	7	31
8.	.5\" Waterproof Tape	1	11	118
9.	0 Calorie Acai Raspberry Water Beverage	1	7	64
10.	0 Calorie Fuji Apple Pear Water Beverage	1	7	64
11.	0 Calorie Strawberry Dragonfruit Water Beverage	1	7	64
12.	0% Fat Black Cherry Greek Yogurt y	1	16	120
13.	0% Fat Blueberry Greek Yogurt	1	16	120
14.	0% Fat Free Organic Milk	1	16	84
15.	0% Fat Greek Yogurt Black Cherry on the Bottom	1	16	120
16.	0% Fat Greek Yogurt Vanilla	1	16	120
17.	0% Fat Organic Greek Vanilla Yogurt	1	16	120
18.	0% Fat Peach Greek Yogurt	1	16	120
19.	0% Fat Strawberry Greek Yogurt	1	16	120

1 - 250 / 49688

ภาพประกอบ 1 แสดงข้อมูลในไฟล์ Product 19 แถวแรก

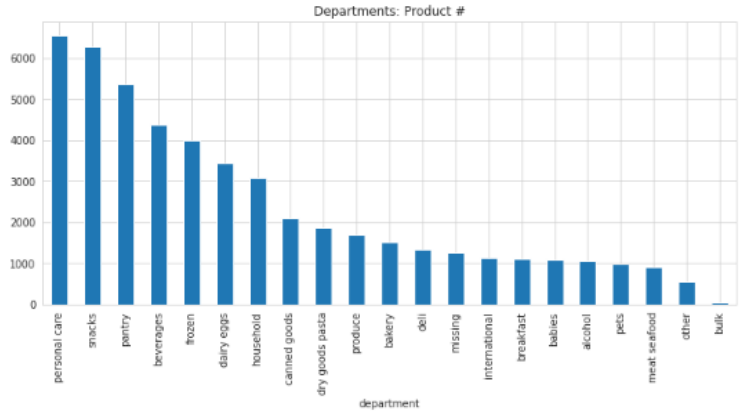
จากภาพประกอบ 1 ผู้วิจัยได้สำรวจรายชื่อสินค้า ไฟล์ ของ Product พบว่า Product Name ทั้งหมด 49,688 รายชื่อ ซึ่งแต่ละรายชื่อประกอบไปด้วย Department_id , Aisle_id แสดงให้เห็นว่าสินค้าแต่ละชนิดจะมีการจัดหมวดหมู่ไว้อย่างชัดเจน ผู้วิจัยจึงมีความสนใจในการสำรวจข้อมูล Aisle และ Department เพื่อเปรียบเทียบแต่ละหมวดหมู่สินค้าว่าสามารถค้นหารูปแบบที่น่าสนใจได้หรือไม่

Total aisles: 134

	aisle_id	aisle
0	1	prepared soups salads
1	2	specialty cheeses
2	3	energy granola bars
3	4	instant foods
4	5	marinades meat preparation

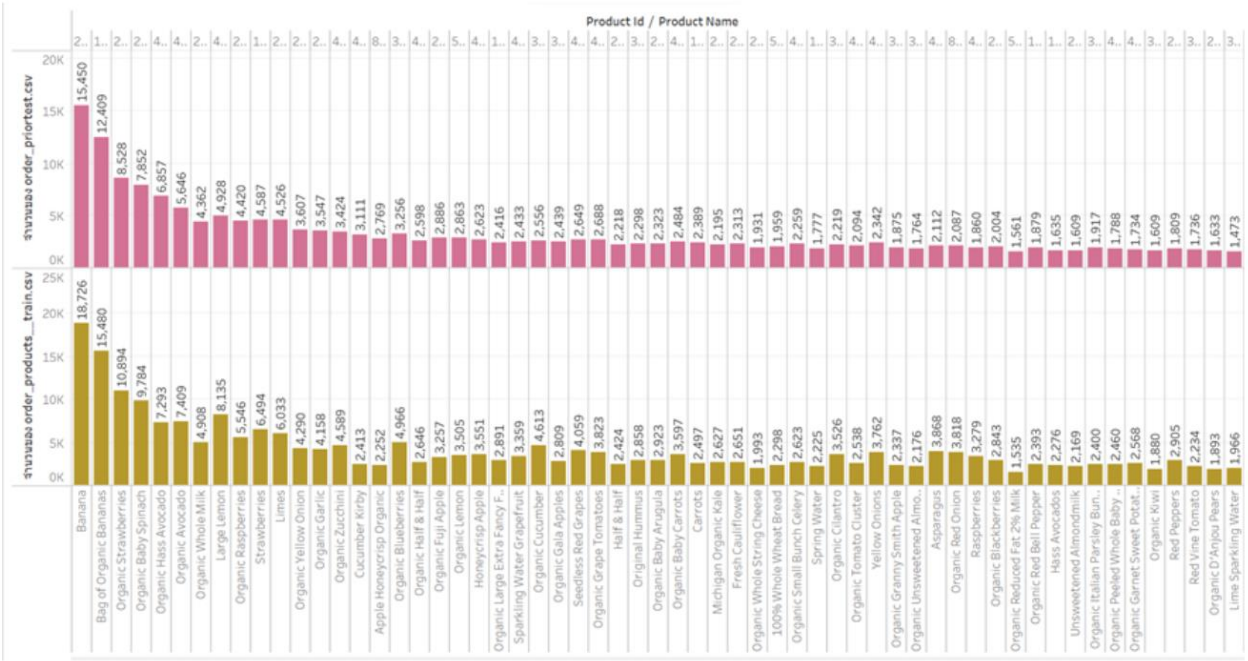
ภาพประกอบ 2 แสดงข้อมูลไฟล์ Aisle 5 แถวแรก

จากภาพประกอบ 2 ผู้วิจัยได้สำรวจข้อมูลชั้นวางสินค้า ว่าแต่ละชั้นวางมีสินค้าประเภทใดบ้าง ซึ่งพบว่าแต่ละชั้นวางมีการระบุชื่อของชั้นวาง และ มีการแสดงข้อมูลระบุตัวตนของชั้นวาง ซึ่งพบว่าประเภทสินค้าในแต่ละชั้นวาง มีลักษณะใกล้เคียงกันอยู่ แต่อยู่คนละชั้นวาง ซึ่งผู้วิจัยจึงมีความสนใจในการสำรวจข้อมูล Department ต่อไป เพื่อที่จะทราบว่าชั้นวางแต่ละประเภทอยู่ในแผนกใดบ้างและแต่ละแผนกมีได้แบ่งหมวดหมู่ไว้อย่างไร



ภาพประกอบ 3 แสดงข้อมูลไฟล์ Department

จากภาพประกอบ 3 ผู้วิจัยได้สำรวจข้อมูลแผนก เพื่อตรวจสอบหมวดหมู่ของสินค้า และปริมาณรายชื่อสินค้า พบว่าข้อมูลแผนก Personal Care มีรายชื่อสินค้ามากที่สุด ผู้วิจัยจึงตั้งสมมติฐานว่า หากมีรายชื่อสินค้ามากที่สุด คาดว่าจะมีการ Order สินค้าจากแผนก Personal Care มากที่สุด ผู้วิจัยจึงได้ตรวจสอบว่า ข้อมูลการ Order สินค้าทุกหมวดหมู่ เพื่อตรวจสอบว่ามีความสอดคล้องกับจำนวนรายชื่อสินค้ามากน้อยแค่ไหน



ภาพประกอบ 4 แสดงจำนวนของการ Order สินค้าไฟล์ Order Train และ Order Test

จากภาพประกอบ 4 ผู้วิจัยได้สำรวจข้อมูลการ Order สินค้าทุกหมวดหมู่ พบว่าอันดับที่ 1 กล้วย ซึ่งเป็นกลุ่มผลไม้ อยู่ในชั้นวาง Fresh Fruits แผนก Produce และ 3 อันดับแรกจะเป็นชั้นวาง Fresh Fruits เช่นเดียวกัน ส่วน อันดับที่ 4 เป็นผัก ซึ่งจากที่ผู้วิจัยได้ตั้งสมมติฐานไว้ว่า “หากมีรายชื่อสินค้ามากที่สุด หากมีรายชื่อสินค้ามากที่สุด คาดว่าจะมีการ Order สินค้าจากแผนก Personal Care มากที่สุด" นั้นไม่สอดคล้องกับข้อมูล แสดงให้เห็นว่า อาจมีรูปแบบความสัมพันธ์ที่ไม่เคยค้นพบมาก่อน อยู่ในชุด

ข้อมูล ซึ่งหากค้นหารูปแบบกฎความสัมพันธ์ของสินค้านั้น อาจจะสามารถสร้างเป็น Bundled Products ที่สามารถจัดโปรโมชั่นคู่กันได้ และ เพิ่มยอดขายได้ รวมถึงสามารถเพิ่มรายชื่อสินค้าในชั้นวางผลไม้ เพิ่มสต็อกสินค้าให้มากขึ้น ซึ่งผู้วิจัยเลือกชั้นวางผลไม้ในการศึกษา โดยจะทำการค้นหากฎความสัมพันธ์ของสินค้าที่มักซื้อคู่กับผลไม้แต่ละชนิด ซึ่งอาจจะเป็นสินค้าประเภทอื่นก็ได้ เช่น ผัก แต่จะต้องอยู่ในแผนก Produce เช่นเดียวกัน ดังนั้น ผู้วิจัยจะสามารถที่จะนำกฎความสัมพันธ์ที่ได้มา นำไปจัดชั้นวางใหม่ ให้ตอบโจทย์กับสิ่งที่ผู้ซื้อคาดหวังในการหยิบสินค้า เพื่อสามารถเพิ่มยอดขายให้กับกลุ่มผลไม้ รวมทั้งกลุ่มผลไม้เป็นกลุ่มที่มีการเก็บไว้ได้ไม่นาน เน่าเสียง่าย หากสามารถเพิ่มยอดขายได้เร็วขึ้น ก็จะลดความสิ้นเปลืองของทรัพยากรลง

ขั้นตอนที่ 3 : ค้นหารูปแบบและกฎความสัมพันธ์ (Association Rules Analysis)

การค้นหารูปแบบ คือ การมองหาในชุดข้อมูลว่ามีรูปแบบอะไรที่น่าสนใจบ้าง เช่น การค้นหากลุ่มของไอเท็มอะไรที่มักซื้อด้วยกัน จากข้อมูลธุรกรรมการซื้อขายสินค้า เพื่อค้นหารูปแบบ โดยที่ Itemset คือ เซตที่มีหนึ่งหรือมากกว่าไอเท็ม , K-Itemset คือ ไอเท็มเซตที่มี K ตัว หรือ เซตของไอเท็มที่เลือกซื้อในแต่ละตะกร้า ไอเท็มจะเกิดพร้อมกันเยอะ จะต้องมีการนับว่าเกิดเป็นเท่าไร การนับจะเรียกว่า Support , Absolute support (count) คือ การนับว่าไอเท็มเซ็ตหนึ่งเกิดพร้อมกันกี่ครั้ง เช่น หาก กล้วย,ส้ม มีการซื้อด้วยกัน จะเกิดขึ้นพร้อมกันในข้อมูลธุรกรรมกี่ครั้ง,Relative Support คือ การคิดสัดส่วนของข้อมูลธุรกรรมที่มีไอเท็มเซ็ตที่เกิดขึ้นพร้อมกัน ต่อ 1 กฎ คิดเป็นกี่เปอร์เซ็นของไอเท็มเซ็ตทั้งหมด เช่น กล้วย,ส้ม มีการเกิดขึ้นพร้อมกัน เป็น 60% ของไอเท็มเซ็ตทั้งหมด หากไอเท็มเซ็นที่เกิดขึ้นพร้อมกันบ่อย จะมีการกำหนด Minsup หรือ Minimum Support ขั้นต่ำของความถี่การเกิดขึ้นพร้อมกันตามที่กำหนด เช่น กำหนดให้ Minsup 50% ซึ่งสามารถนับได้ทั้งการซื้อ 1 Itemset , 2 Itemset หรือมากกว่านั้น แสดงให้เห็นว่าหาก กล้วย,ส้ม เกิดขึ้นพร้อมกัน 60% แสดงว่ากฎนี้ผ่าน Minsup สามารถนำไปวิเคราะห์ได้ การวิเคราะห์กฎความสัมพันธ์ จะพิจารณาจากค่า Support, Confidence, Lift, Leverage และ Conviction ดังนี้

จากไอเท็มเซ็ตที่เกิดบ่อยและผ่าน Minimum Support สามารถนำไปสร้างเป็นกฎความพันธ์ (Association Rules) โดยจะมีการกำหนด Antecedent คือ สิ่งที่เกิดขึ้นก่อน และ Consequent คือ สิ่งที่ตามมา
ดังนั้น **Support, Sup** : ความน่าจะเป็นที่ธุรกรรมจะมี X U Y

ภาพประกอบ 1 แสดงกฎความสัมพันธ์ระหว่างไอเท็มเซ็ตที่ผ่านค่า Minimum Support ที่มี Itemset = {กล้วย,ส้ม,มะเขือเทศ}

การจะเลือกกฎความสัมพันธ์ว่า Itemset ใดเหมาะสมกับ Antecedent หรือ Consequent จะต้องมีค่าความเชื่อมั่นของกฎ หรือ **Confidence, Conf** : ความน่าจะเป็นเงื่อนไขที่ธุรกรรมที่มี X จะมี Y ด้วย ; Confident = Sup(X U Y) / Sup(X)เช่น การซื้อกล้วย แล้วซื้อส้ม จะต้องมีเงื่อนไขของการซื้อกล้วยว่าเป็นเท่าไรด้วย ดังนั้นการหา Association Rule Mining จะทำการหากฎ X→Y ทั้งหมดที่ผ่าน Minimun support และ Minimum Confident

เมื่อได้กฎความสัมพันธ์แล้ว จะตรวจสอบความสัมพันธ์ของกฎจากค่าของการวัด Event การธุรกรรมการซื้อสินค้าในแต่ละครั้งมีความสัมพันธ์กันหรือไม่ เรียกว่า ค่า Lift โดย Lift(B,C) = $\frac{Conf(B \rightarrow C)}{Sup(C)} = \frac{Sup(B \cup C)}{Sup(B)x\ Sup(C)}$ ซึ่งค่า Lift เป็นการวัดความแตกต่างของ Support ของกฎทั้งหมด ว่ามีการขึ้นต่อกันหรือไม่ หรือ เป็นอิสระต่อกัน โดยมีการกำหนดช่วงค่าอยู่ใน [0, ∞] โดยสามารถใช้เป็นเกณฑ์ในการตัดสินใจได้ ค่า Lift(B,C) สามารถบอกได้ว่า B และ C เกี่ยวข้องกันอย่างไร โดยมีค่าดังนี้ Lift(B,C) = 1: B และ C เป็นอิสระต่อกัน (Independent) , Lift(B,C) > 1: B และ C เกี่ยวข้องกันในทิศทางบวก (Positively Correlated) และกฎที่สร้างขึ้นจาก B อาจมีประโยชน์ในการทำนาย C ในอนาคต , Lift(B,C) < 1: B และ C เกี่ยวข้องกันในทิศทางลบ (Negatively Correlated) และการมี B อาจส่งผลเสียต่อการเกิด C และกฎที่สร้างขึ้นจาก B อาจไม่มีประโยชน์ในการทำนาย C

หากค่า Confident,Lift ไม่ดี จะเลือกใช้ค่า Conviction ร่วมด้วย โดย Conv(X → Y) = $\frac{1-Sup(Y)}{1-Conf(X \rightarrow Y)}$ แสดงว่า Conviction จะเป็นตัววัดที่คำนวณจากการพิจารณา Support ของสิ่งที่เกิดก่อน (Antecedent) และสิ่งที่ตามมา (Consequent) ของกฎ โดย Conviction สามารถจำแนกทิศทางของกฎได้ (Sensitive To Rule Direction) โดยแตกต่างจาก Lift ที่ไม่สามารถจำแนกทิศทางของกฎได้ Conv(X→ Y) ≠ Conv(Y→X) ดังนั้น Conviction ที่บ่งบอกถึงความแตกต่างของ X และ Y โดยพิจารณาถึงทิศทางของการนำมาซึ่งกันและกัน โดยวัดการอุปนัยของการเชื่อมโยงของกฎว่าดีมากแค่ไหน โดย Conviction ไม่มีการจำกัดขอบเขตตัวแปร กำหนด threshold ของ Conviction จะแปรผลได้ดังนี้ 0.5 <= Conv < 1 → Negative Dependence ;conv = 1 Independence ; conv >1 → Positive Dependence ; conv → ∞ → Logical implication

ซึ่งจะมีค่าตัวแปรที่มีการพิจารณาหาความแตกต่างของสิ่งที่เกิดก่อน (Antecedent) และสิ่งที่ตามมา (Consequent) ที่มีการขึ้นต่อกัน หรือ อิสระต่อกัน เรียกว่า Leverage โดยค่า Leverage เท่ากับ 0 จะแสดงถึงความอิสระ ซึ่งเมื่อสร้างกฎขึ้นมาแล้วสามารถนำมาเป็นเครื่องมือในการเลือกกฎ หรือ เปรียบเทียบกฎที่มีการขึ้นต่อกันและอิสระต่อกัน ว่ากฎใดน่าสนใจมากกว่ากัน เพื่อนำไปใช้จริงในแต่ละแผนกได้ โดย Leverage(X → Y) = Sup(X ∪ Y) – (Sup(X)*Sup(Y))

โดยผู้วิจัยได้ทำการวิเคราะห์กฎความสัมพันธ์ของผลไม้ โดยกำหนดให้ สิ่งที่เกิดก่อน (Antecedent) และสิ่งที่ตามมา (consequent) ของกฎ ต้องเป็นผลไม้ อย่างใดอย่างหนึ่ง หรือ ทั้งสอง โดยเลือกพิจารณาจากค่า Support, Confidence, Lift, Leverage และ Conviction โดยกำหนดให้ Minimum Support คือ 0.01 , Minimum Confidence คือ 0.01 และ Minimum Lift คือ 1 โดยผู้วิจัยได้พิจารณาเลือกมา 8 กฎ ดังนี้

กฎ	Support	Confidence	Lift	Leverage	Conviction
กฎที่ 1 : Bag of Organic Bananas To Organic Hass Avocado	0.02	0.16	2.07	0.01	1.1
กฎที่ 2 : Organic Strawberries	0.03	0.28	1.77	0.01	1.17

To Bag of Organic Bananas					
กฎที่ 3 : Organic Baby Spinach To Bag of Organic Bananas	0.02	0.23	1.43	0.01	1.10
กฎที่ 4 : Limes To Large Lemon	0.02	0.26	3.15	0.01	1.25
กฎที่ 5 : Organic Large Extra Fancy Fuji Apple To Bag of Organic Bananas	0.01	0.34	2.11	0.01	1.27
กฎที่ 6 : Banana To Honeycrisp Apple	0.01	0.07	1.8	0.01	1.03
กฎที่ 7 : Organic Yellow Onion To Bag of Organic Bananas	0.01	0.23	1.46	0.003	1.1
กฎที่ 8 : Seedless Red Grapes	0.01	0.3	1.48	0.004	1.13

To Banana					

ตาราง 2 แสดงกฎความสัมพันธ์ที่เลือกจากของข้อมูล Train

กฎ	Support	Confidence	Lift	Leverage	Conviction
กฎที่ 1 : Bag of Organic Bananas To Organic Hass Avocado	0.03	0.16	1.82	0.01	1.09
กฎที่ 2 : Organic Strawberries To Bag of Organic Bananas	0.03	0.23	1.44	0.01	1.09
กฎที่ 3 : Organic Baby Spinach To Bag of Organic Bananas	0.02	0.21	1.33	0.01	1.07
กฎที่ 4 : Limes To Large Lemon	0.01	0.19	3.04	0.01	1.16
กฎที่ 5 :	0.01	0.32	1.97	0.01	1.23

Organic Large Extra Fancy Fuji Apple To Bag of Organic Bananas					
กฎที่ 6 : Banana To Honeycrisp Apple	0.01	0.06	1.82	0.01	1.03
กฎที่ 7 : Organic Yellow Onion To Bag of Organic Bananas	0.01	0.22	1.37	0.0028	1.08
กฎที่ 8 : Seedless Red Grapes To Banana	0.01	0.31	1.55	0.0037	1.16

ตาราง 3 แสดงกฎความสัมพันธ์ที่เลือกจากของข้อมูล Test

ขั้นตอนที่ 4 : การประเมินผลการทดลอง (Model Evaluation)

การประเมินผลการทดลอง จะเป็นการเปรียบเทียบผลการวิเคราะห์กฎความสัมพันธ์ในแต่ละค่าตัวแปร ได้แก่ Support, Confidence, Lift, Leverage และ Conviction โดยที่จะนำมาเปรียบเทียบระหว่างข้อมูลจาก Train set และ Test set โดยการใช้ Precision@k และ Recall@K ซึ่งเว็บไซต์ Insidelearningmachines.com/precisionk_and_recallk/ [3] กล่าวไว้ว่าค่าเมตริกจะถูกใช้ในการประเมินโมเดลการแนะนำ เพื่อวัดว่าโมเดลนั้นมีประสิทธิภาพในการให้ข้อเสนอแนะที่เกี่ยวข้องกับผู้ใช้งานหรือไม่ ค่าเหล่านี้มุ่งเน้นวัดความเป็นประโยชน์ของโมเดลในการแนะนำสิ่งที่เกี่ยวข้องแก่ผู้ใช้งานในทางปฏิบัติ จากการดูทั้งชุดข้อเสนอแนะที่เป็นไปได้ทั้งหมด ค่า Precision@k และ Recall@k จะใช้ในการพิจารณาเฉพาะชุดข้อเสนอแนะที่ดีที่สุดแค่ Top k เท่านั้น Recommended items คือ รายการสิ่งของที่โมเดลของเราแนะนำให้กับผู้ใช้งานเฉพาะ , Levant items คือ รายการสิ่งของที่ถูกเลือกจริงๆโดยผู้ใช้งานเฉพาะ , @k อ่านว่า "ที่ k" โดย k เป็นจำนวนเต็มของรายการที่เราพิจารณา**จาก**

$$\text{Precision@K} = \frac{\text{จำนวนของ recommended items ที่เป็น relevant @k}}{\text{จำนวนของ recommended items @k}}$$

$$\text{Recall@K} = \frac{\text{จำนวนของ recommended items ที่เป็น relevant @k}}{\text{จำนวนของ relevant items ทั้งหมด}}$$

ภาพประกอบ 2 แสดงสูตรการคำนวณ Precision@K และ Recall@K

ผลการวิจัยและอภิปรายผลการวิจัย

หลังจากได้ทำการวิเคราะห์กฎความสัมพันธ์ของการกลุ่มผลไม้พบว่าได้มาทั้งหมด 8 กฎ โดยผู้วิจัยได้ทำการวิเคราะห์กฎความสัมพันธ์ที่มีความน่าสนใจ โดยเลือกคู่สินค้าที่มักซื้อคู่กัน และ วิเคราะห์หาวิธีปรับเปลี่ยนชั้นวาง จากกฎความสัมพันธ์ที่มีการซื้อสินค้าเหมือนกันให้มาอยู่ใกล้กัน โดยผู้วิจัยได้ทำการวิเคราะห์กฎความสัมพันธ์ทั้ง 8 กฎไว้ดังนี้ โดยพิจารณาจาก Train set และ Test set ดังนี้

ผลการวิจัย

กฎที่ 1 ใน Test set จะมีค่า Support เป็น 0.03 ซึ่งสูงกว่า Min Support ที่กำหนด ดังนั้นการซื้อ Bag of Organic Bananas มักซื้อคู่กับ Organic Hass Avocado อยู่ 0.02 และมีค่า Confident หรือ ความเชื่อมั่นในการซื้อ Bag of Organic Bananas จะมี Organic Hass Avocado อยู่เป็น 0.16 ซึ่งเมื่อทำการวัดค่า Lift พบว่ามีค่าเป็น 1.82 ซึ่งกฎนี้จะมีทิศทางบวก (Positively Correlated) ดังนั้น กฎที่สร้างขึ้นจาก Bag of Organic Bananas อาจมีประโยชน์ในการทำนาย Organic Hass Avocado ในอนาคตได้ ซึ่งเมื่อวัดการอุปนัยของการเชื่อมโยงของกฎ โดย Conviction ของกฎเป็น 1.09 พบว่ากฎนี้มียังมีค่าการเชื่อมโยงเชิงบวก (Positive Dependence) และเมื่อวัดว่ากฎว่ามีการขึ้นต่อกัน หรือ เป็นอิสระต่อกัน โดยค่า Leverage พบว่า กฎนี้มีความเชื่อมโยงกันเป็น 0.01

กฎที่ 2 ใน Test set จะมีค่า Support เป็น 0.03 ซึ่งสูงกว่า Min Support ที่กำหนด ดังนั้นการซื้อ Organic Strawberries มักซื้อคู่กับ Bag of Organic Bananas อยู่ 0.03 และมีค่า Confident หรือ ความเชื่อมั่นในการซื้อ Organic Strawberries จะมี Bag of Organic Bananas อยู่เป็น 0.23 ซึ่งเมื่อทำการวัดค่า Lift พบว่ามีค่าเป็น 1.44 ซึ่งกฎนี้จะมีทิศทางบวก (Positively Correlated) ดังนั้น กฎที่สร้างขึ้นจาก Organic Strawberries อาจมีประโยชน์ในการทำนาย Bag of Organic Bananas ในอนาคตได้ ซึ่งเมื่อวัดการอุปนัยของการเชื่อมโยงของกฎ โดย Conviction ของกฎเป็น 1.09 พบว่ากฎนี้มียังมีค่าการเชื่อมโยงเชิงบวก

(Positive Dependence) และเมื่อวัดว่ากฎว่ามีการขึ้นต่อกัน หรือ เป็นอิสระต่อกัน โดยค่า Leverage พบว่า กฎนี้มีความเชื่อมโยงกันเป็น 0.01

กฎที่ 3 ใน Test set จะมีค่า Support เป็น 0.02 ซึ่งสูงกว่า Min Support ที่กำหนด ดังนั้นการซื้อ Organic Baby Spinachมักซื้อคู่กับ Bag of Organic Bananas อยู่ 0.02 และมีค่า Confident หรือ ความเชื่อมั่นในการซื้อ Organic Baby Spinach จะมี Bag of Organic Bananas อยู่เป็น 0.21 ซึ่งเมื่อทำการวัดค่า Lift พบว่ามีค่าเป็น 1.33 ซึ่งกฎนี้จะมีทิศทางบวก (Positively Correlated) ดังนั้น กฎที่สร้างขึ้นจาก Organic Baby Spinach อาจมีประโยชน์ในการทำนาย Bag of Organic Bananas ในอนาคตได้ ซึ่งเมื่อวัดการอุปนัยของการเชื่อมโยงของกฎ โดย Conviction ของกฎเป็น 1.07 พบว่ากฎนี้มียังมีค่าการเชื่อมโยงเชิงบวก (Positive Dependence) และเมื่อวัดว่ากฎว่ามีการขึ้นต่อกัน หรือ เป็นอิสระต่อกัน โดยค่า Leverage พบว่า กฎนี้มีความเชื่อมโยงกันเป็น 0.01

กฎที่ 4 ใน Test set จะมีค่า Support เป็น 0.01 ซึ่งสูงกว่า Min Support ที่กำหนด ดังนั้นการซื้อ Limes มักซื้อคู่กับ Lemon อยู่ 0.01 และมีค่า Confident หรือ ความเชื่อมั่นในการซื้อ Limes จะมี Lemon อยู่เป็น 0.19 ซึ่งเมื่อทำการวัดค่า Lift พบว่ามีค่าเป็น 3.04 ซึ่งกฎนี้จะมีทิศทางบวก (Positively Correlated) ดังนั้น กฎที่สร้างขึ้นจาก Limes อาจมีประโยชน์ในการทำนาย Large Lemon ในอนาคตได้ ซึ่งเมื่อวัดการอุปนัยของการเชื่อมโยงของกฎ โดย Conviction ของกฎเป็น 1.16 พบว่ากฎนี้มียังมีค่าการเชื่อมโยงเชิงบวก (Positive Dependence) และเมื่อวัดว่ากฎว่ามีการขึ้นต่อกัน หรือ เป็นอิสระต่อกัน โดยค่า Leverage พบว่า กฎนี้มีความเชื่อมโยงกันเป็น 0.01

กฎที่ 5 ใน Test set จะมีค่า Support เป็น 0.01 ซึ่งสูงกว่า Min Support ที่กำหนด ดังนั้นการ Organic Large Extra Fancy Fuji Apple มักซื้อคู่กับ Bag of Organic Bananas อยู่ 0.01 และมีค่า Confident หรือ ความเชื่อมั่นในการซื้อ Organic Large Extra Fancy Fuji Apple จะมี Bag of Organic Bananas อยู่เป็น 0.32 ซึ่งเมื่อทำการวัดค่า Lift พบว่ามีค่าเป็น 1.97 ซึ่งกฎนี้จะมีทิศทางบวก (Positively Correlated) ดังนั้น กฎที่สร้างขึ้นจาก Organic Large Extra Fancy Fuji Apple อาจมีประโยชน์ในการทำนาย Bag of Organic Bananas ในอนาคตได้ ซึ่งเมื่อวัดการอุปนัยของการเชื่อมโยงของกฎ โดย Conviction ของกฎเป็น 1.23 พบว่ากฎนี้มียังมีค่าการเชื่อมโยงเชิงบวก (Positive Dependence) และเมื่อวัดว่ากฎว่ามีการขึ้นต่อกัน หรือ เป็นอิสระต่อกัน โดยค่า Leverage พบว่า กฎนี้มีความเชื่อมโยงกันเป็น 0.01

กฎที่ 6 ใน Test set จะมีค่า Support เป็น 0.01 ซึ่งสูงกว่า Min Support ที่กำหนด ดังนั้นการซื้อ Banana มักซื้อคู่กับ Honeycrisp Apple อยู่ 0.01 และมีค่า Confident หรือ ความเชื่อมั่นในการซื้อ Banana จะมี Honeycrisp Apple อยู่เป็น 0.06 ซึ่งเมื่อทำการวัดค่า Lift พบว่ามีค่าเป็น 1.82 ซึ่งกฎนี้จะมีทิศทางบวก (Positively Correlated) ดังนั้น กฎที่สร้างขึ้นจาก Banana อาจมีประโยชน์ในการทำนาย Honeycrisp Apple ในอนาคตได้ ซึ่งเมื่อวัดการอุปนัยของการเชื่อมโยงของกฎ โดย Conviction ของกฎเป็น 1.03 พบว่ากฎนี้มียังมีค่าการเชื่อมโยงเชิงบวก (Positive Dependence) และเมื่อวัดว่ากฎว่ามีการขึ้นต่อกัน หรือ เป็นอิสระต่อกัน โดยค่า Leverage พบว่า กฎนี้มีความเชื่อมโยงกันเป็น 0.01

กฎที่ 7 ใน Test set จะมีค่า Support เป็น 0.01 ซึ่งสูงกว่า Min Support ที่กำหนด ดังนั้นการซื้อ Organic Yellow Onion มักซื้อคู่กับ Bag of Organic Bananas อยู่ 0.01 และมีค่า Confident หรือ ความเชื่อมั่นในการซื้อ Organic Yellow Onion จะมี Bag of Organic Bananas อยู่เป็น 0.22 ซึ่งเมื่อทำการวัดค่า Lift พบว่ามีค่าเป็น 1.37 ซึ่งกฎนี้จะมีทิศทางบวก (Positively Correlated) ดังนั้น กฎที่สร้างขึ้นจาก Organic Yellow Onion อาจมีประโยชน์ในการทำนาย Bag of Organic Bananas ในอนาคตได้ ซึ่งเมื่อวัดการอุปนัยของการเชื่อมโยงของกฎ โดย Conviction ของกฎเป็น 1.08 พบว่ากฎนี้มียังมีค่าการเชื่อมโยงเชิงบวก (Positive Dependence) และเมื่อวัดว่ากฎว่ามีการขึ้นต่อกัน หรือ เป็นอิสระต่อกัน โดยค่า Leverage พบว่า กฎนี้มีความเชื่อมโยงกันเป็น 0.0028

กฎที่ 8 ใน Test set จะมีค่า Support เป็น 0.01 ซึ่งสูงกว่า Min Support ที่กำหนด ดังนั้นการซื้อ Organic Yellow Onion มักซื้อคู่กับ Bag of Organic Bananas อยู่ 0.01 และมีค่า Confident หรือ ความเชื่อมั่นในการซื้อ Organic Yellow Onion จะมี Bag of Organic Bananas อยู่เป็น 0.31 ซึ่งเมื่อทำการวัดค่า Lift พบว่ามีค่าเป็น 1.55 ซึ่งกฎนี้จะมีทิศทางบวก (Positively Correlated) ดังนั้น กฎที่สร้างขึ้นจาก Organic Yellow Onion อาจมีประโยชน์ในการทำนาย Bag of Organic Bananas ในอนาคตได้ ซึ่งเมื่อวัดการอุปนัยของการเชื่อมโยงของกฎ โดย Conviction ของกฎเป็น 1.16 พบว่ากฎนี้มียังมีค่าการเชื่อมโยงเชิงบวก (Positive Dependence) และเมื่อวัดว่ากฎว่ามีการขึ้นต่อกัน หรือ เป็นอิสระต่อกัน โดยค่า Leverage พบว่า กฎนี้มีความเชื่อมโยงกันเป็น 0.0037

เมื่อทำการเปรียบเทียบกฎทั้ง8กฎในข้อมูล Test และ Train พบว่ามีความใกล้เคียงกัน และการวิเคราะห์เป็นไปในทิศทางเดียวกันทุกกฎ

อภิปรายผลการวิจัย

ผู้วิจัยได้นำผลการวิจัยจากการวิเคราะห์กฎความสัมพันธ์มาเลือกกฎที่มีความน่าสนใจ โดยผู้วิจัยได้เลือกกฎที่น่าสนใจสำหรับการจัดโปรโมชั่นสินค้าคู่กัน จากข้อมูลธุรกรรมการซื้อขายสินค้าที่มักซื้อสินค้าชนิดนี้แล้วจะซื้อคู่กับอีกชนิด ได้ทั้งหมด 2 กฎ ได้แก่ กฎที่ 2 Organic Strawberries to Bag of Organic Bananas เนื่องจากค่า Confident มีค่าที่สูง เป็น 0.28 และค่า Lift เป็น 1.77 ซึ่งค่า Support มากที่สุดในทุกกฎ เป็น 0.03 รวมทั้งมีค่าที่บ่งบอกถึงความเชื่อมโยงกันของกฎ เป็น 1.17 และ กฎที่ 4 Limes to Large Lemon เนื่องจากมีค่า Confident เป็น 0.26 และมีค่า lift มากที่สุดในทุกกฎ คือ 3.15 ดังนั้น กฎที่สร้างขึ้นจาก Limes อาจมีประโยชน์ในการทำนาย Large Lemon ในอนาคตได้ และมีโอกาสที่จะเป็นต้นแบบในการเพิ่มสต็อกสินค้ากลุ่ม Large lemon ได้ ทั้งนี้ผู้วิจัยได้พบว่า มีหลายกฎที่มีสินค้าซ้ำกัน เช่น Bag of Organic Bananas ที่จะอยู่ในกฎมากกว่า 2 กฎ ทำให้สามารถจัดชั้นวางใหม่ให้เป็นโอกาสที่ลูกค้าสามารถหยิบซื้อสินค้าได้ง่ายขึ้น ดังนี้การจัดชั้นวางที่ผู้วิจัยแนะนำ ได้แก่ การจัดชั้นวาง Bag of Organic Bananas ให้อยู่ใกล้กับ Organic Hass Avocado , Organic Baby Spinach , Organic Yellow Onion เนื่องจากมีความเกี่ยวข้องกันในแต่ละกฎ แต่สินค้าจะอยู่ชั้นวางที่ต่างกัน ดังนั้น Bag of Organic Bananas อาจจะมีการเพิ่มช่องวางเล็กๆให้ใกล้กับสินค้าประเภทผัก และอาจมีเมนูที่แนะนำที่ทำจากสินค้าที่มีความสัมพันธ์กัน เพื่อเพิ่มยอดขายให้ลูกค้าตัดสินใจที่จะเลือกซื้อ

ได้มากขึ้น และ การจัดชั้นวางให้ Limes และ Large Lemon อยู่ด้วยกัน เนื่องจากลูกค้ามักซื้อสินค้าร่วมกัน โดยอาจจะมีการจัดโปรโมชั่นส่วนลดสินค้าเมื่อซื้อคู่กันและนำสินค้าทั้ง2ชนิดนี้ อยู่ในจุดที่มองเห็นได้มากขึ้น เพื่อเพิ่มโอกาสในการขายสินค้าได้ อีกทั้งสามารถนำไปสู่การวางแผนเพิ่มสต็อกสินค้าได้ เมื่อทราบว่าความต้องการของลูกค้าในแต่ละช่วงเวลาเป็นอย่างไร ซึ่งผู้วิจัยแนะนำให้ฝ่ายการตลาดทำการวิเคราะก์กฎความสัมพันธ์ในแต่ละแต่ละสัปดาห์หรือแต่ละช่วงเวลา เป็นอีกเครื่องมือที่นำมาวางแผนกลยุทธ์ได้ ซึ่งในอนาคตสามารถนำการวิเคราะห์กฎความสัมพันธ์ไปประยุกต์ใช้ในร้านค้าแต่ละพื้นที่ เพื่อทราบว่าในแต่ละพื้นที่มีความต้องการของสินค้าต่างกันหรือไม่ และอาจจะเชื่อมโยงกับภูมิศาสตร์ของพื้นที่นั้น เช่นหากความต้องการของลูกค้าในพื้นที่ต้องการข้าวโพดมากที่สุด ฝ่ายการตลาดก็จะทำการจัดสรรผู้ผลิตในพื้นที่ หรือ นอกพื้นที่ เพื่อให้ตรงตามความต้องการ หรือ เพิ่มโอกาสให้ชาวบ้านในชนบท ปลูกข้าวโพด และทางร้านค้ารับซื้อ เพื่อเพิ่มรายได้ให้กับชุมชน เพิ่มยอดขายให้กับร้านค้า เพื่อคุณภาพชีวิตที่ดี ความเป็นอยู่ที่ดี สอดคล้องความต้องการของลูกค้า เพิ่มความประทับใจของลูกค้าที่มีต่อร้านค้าได้

สรุปผลการวิจัย

งานวิจัยได้ทำการวิเคราะห์กฎความสัมพันธ์โดยใช้เครื่องมือ Apriori Algorithm ทดแทนการใช้มนุษย์ในการคำนวณ เพิ่มความรวดเร็ว และ แม่นยำในการวัดค่าตัวแปรต่างๆได้แก่ ค่า Support, Confidence, Lift, Leverage และ Conviction เพื่อนำไปสร้างกฎความสัมพันธ์ (Association rules) และเลือกกฎที่น่าสนใจไปสร้างโปรโมชั่นการซื้อสินค้าคู่กัน และ วางแผนการจัดชั้นวางใหม่ เพื่อเพิ่มยอดขายให้กับร้านค้าได้ ซึ่งผู้วิจัยได้ทำการประเมินประสิทธิภาพของกฎด้วย Precision@K และ Recall@K เพื่อให้กฎที่ออกมามีความน่าเชื่อถือมากขึ้น ทั้งนี้ผู้วิจัยได้ทำการแนะนำงานวิจัยในอนาคตในการสร้างกฎความสัมพันธ์ในแต่ละช่วงเวลา เนื่องจาก พฤติกรรมลูกค้ามีการเปลี่ยนแปลงไปตลอด รวมถึงการวิเคราะห์ความต้องการสินค้าของลูกค้าในแต่ละพื้นที่ เพื่อทราบว่าในแต่ละพื้นที่มีความต้องการของสินค้าต่างกันหรือไม่ และอาจจะเชื่อมโยงกับภูมิศาสตร์ของพื้นที่นั้นได้ ซึ่งมีโอกาสที่จะทำให้รายได้โดยรวมของประเทศสูงขึ้น ชาวบ้านในชนบทแต่ละพื้นที่มีโอกาสที่จะมีงานทำ ซึ่งเป็นประโยชน์และเพิ่มคุณภาพชีวิตที่ดีขึ้นได้

เอกสารอ้างอิง

[1] A R Efrat et al , “Consumer purchase patterns based on market basket analysis using apriori algorithms”,2020

[2] Sabrina Kusuma Ayu, Isti Surjandari, Zulkarnain ,”Mining Association Rules in Seasonal Transaction Data”,2018

[3] Inside learning machines, " Precision@k and Recall@k Made Easy with 1 Python Example" [Online]. สืบค้นจาก https://insidelearningmachines.com/precisionk_and_recallk/