การศึกษาแนวโน้มพฤติกรรมการบริโภคอาหารในปัจจุบันของนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต โดยใช้วิธีการทำเหมืองข้อมูล

The Study of the Students Behavior from the Faculty of Science and Technology Phuket Rajabhat University in Consuming the Food and Beverage using Data Mining

ธิดารัตน์ ศิวะนาถนุสรณ์[1]*, อธิสิน ชิตรัตน์[2], ประคอง เอี่ยมพิกุล[3] , วีระวุฒิ ทองโคตร[4] , วิภาวรรณ บัวทอง[5]
Thidarat Siwanartnusorn[1]*, Atisin Chitaratana[2], Prakong Eiampikul[3], Weerawut Thongkhot[4], Wipawan Buathong[5]
สาขาวิชาเทคโนโลยีดิจิทัล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต
Digital Technology Program Faculty of Science and Technology Phuket Rajabhat University
Email: s6381423102@pkru.ac.th

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อที่จะหาอุปสงค์การบริโภคอาหารที่แท้จริงของนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต จากกลุ่มตัวอย่าง จำนวน 170 คน ด้วยกระบวนการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูลโดยใช้อัลกอริทึม เพื่อนบ้านใกล้เคียง ต้นไม้ตัดสินใจ เบย์เซียนแบบง่าย และการทำเหมืองข้อมูลแบบกฎความสัมพันธ์ โดยใช้อัลกอริทึม อะพริโอริ ผลลัพธ์ของโมเดลเพื่อหาคำตอบของ อุปสงค์ด้านความเชื่อมโยงของประเภทอาหาร การเลือกเครื่องดื่ม และช่วงเวลาที่นักศึกษาทานอาหารหรือเครื่องดื่ม ผ่านทางเครื่องมือ Orange 3.27.1 ผลลัพธ์จากการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล พบว่า อัลกอริทึมต้นไม้ตัดสินใจและวิธีการแบ่งข้อมูลแบบ Percentage Split 70/30 ให้ค่าประสิทธิภาพของโมเดลสูงสุด สำหรับการทำเหมืองข้อมูลแบบกฎความสัมพันธ์ โดยใช้อัลกอริทึมอะพริโอริ แสดงกฎความสัมพันธ์ทั้งหมด 30 กฎ เมื่อนำมาเปรียบเทียบกับกฎของต้นไม้ตัดสินใจ พบว่ามีจำนวน 16 กฎที่มีความเชื่อมโยงและไปในทิศทางเดียวกัน โดยพบความสอดคล้องกับโมเดลการเลือกประเภทอาหาร 5 กฎ สอดคล้องกับโมเดลการเลือกเครื่องดื่ม 6 กฎ สอดคล้องกับโมเดลการเลือกช่วงเวลา 5 กฎ

คำสำคัญ: เพื่อนบ้านใกล้เคียง, ต้นไม้ตัดสินใจ, เบย์เซียนแบบง่าย, อะพริโอริ

Abstract

The purpose of this study is to find the real demand from the students of Phuket Rajabhat University, the Faculty of Science and Technology by using the methodology of classification in data mining with K-nearest Neighbor, Decision Tree and Naive Bayes algorithm to construct the model from 170 consumers and use the association rules with Apriori algorithm to find out the result of relationship between food, beverage, and time range via Orange 3.27.1. Results show that Decision Tree algorithm combined with Percentage Split

70/30 model gives the highest accuracy rate, while the association rules with Apriori method give 30 associated rules. As a result, there are 16 rules that originated from Association rules and Decision Tree. 5 rules from the choosing food model, 6 rules from choosing the beverage models, and 5 rules from the consuming time range model.

Keywords: K-Nearest Neighbor, Decision Tree, Naive Bayes, Apriori

บทนำ

ประเทศไทยเป็นประเทศที่มีความหลากหลายทางอาหารและเป็นแหล่งผลิตอาหารสำคัญของโลก ในปัจจุบัน ธุรกิจด้านอาหารมีความสำคัญและเติบโตเสมอมาอย่างต่อเนื่อง เพราะจำนวนของประชาการเพิ่มขึ้น ความต้องการอาหารก็เพิ่มขึ้น ประเทศไทยเองเช่นกัน จำนวนประชากรได้เพิ่มขึ้นจาก 62.3 ล้านคนเป็น 69.63 ล้านคน จากปีพ.ศ. 2542 ถึง ปี พ.ศ 2562 (https://www.worldbank.org/th/country/thailand/overview) เมื่อธุรกิจอาหารจึงเป็นที่น่าสนใจ ในมหาวิทยาลัยราชภัฏภูเก็ต ก็มีร้านอาหารให้นักศึกษาเลือกที่จะบริโภค แต่ทั้งนี้ทั้งนั้นอาหารที่ขายอยู่อาจจะไม่ได้ตรงกับความต้องการของนักศึกษา ทำให้ผู้ประกอบการเสียโอกาสในการเข้าถึงอุปสงค์ของตลาด อีกทั้งอุปทานร้านอาหารต่าง ๆ นอกมหาวิทยาลัย ก็แย่งส่วนแบ่งตลาดในการบริโภคอาหารในมหาวิทยาลัยไปได้ในระดับหนึ่ง

ปัจจุบันเทคโนโลยีสารสนเทศมีส่วนเกี่ยวข้องกับชีวิตของมนุษย์ในแทบทุก ๆ ด้าน ทำให้เกิดการเก็บรวมรวมข้อมูลที่มากมายอยู่ในฐานข้อมูล และสามารถที่จะนำข้อมูลนั้นมาวิเคราะห์เพื่อที่จะทำการทำนายในเรื่องที่ต้องการศึกษาต่อไปได้ กระบวนการที่นำข้อมูลจำนวนมากมาทำการวิเคราะห์นั้นคือกระบวนการทำเหมืองข้อมูล

การทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการในการค้นหาชุดข้อมูลขนาดใหญ่ โดยใช้วิธีการของการเรียนรู้ของเครื่อง (Machine Learning) กระบวนการทางสถิติ ระบบฐานข้อมูล และคณิตศาสตร์ อีกทั้งใช้กระบวนการเหล่านี้ค้นหาความสัมพันธ์ในชุดข้อมูลที่มีอยู่เพื่อนำใช้พยากรณ์หรือทำนายในสิ่งต่าง ๆ ในหลากหลายแง่มุมครอบคลุมในทุก ๆ ด้านของชีวิตมนุษย์ (ศุภามณ จันทร์สกุล, 2561)

ผู้ทำวิจัยได้รวบรวมข้อมูลจากนักศึกษาระดับปริญญาตรี มหาวิทยาลัยราชภัฏภูเก็ต ชั้นปีที่ 1-4 จำนวน 170 คน โดยมีตัวแปรเป็นข้อมูลพฤติกรรมการบริโภคอาหารซึ่งผู้วิจัยใช้ตัวแบบการจำแนกประเภทข้อมูลในการความทำนายพฤติกรรมในการบริโภคอาหาร และใช้การทำเหมืองข้อมูลแบบกฎความสัมพันธ์ ในการหาความสัมพันธ์ในพฤติกรรมในการบริโภคอาหารและนำผลลัพธ์ที่ได้จากทั้งสองส่วนเทียบหาความสัมพันธ์กัน จากการศึกษายังพบว่ามีงานวิจัยหลายๆ งานวิจัย ได้ใช้เทคนิคเหล่านี้ในการสร้างต้นแบบทำนายและหาความสัมพันธ์เพื่อแก้ไขปัญหา เช่น การศึกษาปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพ ตัวแบบพยากรณ์การลาออกของพนักงาน กรณีศึกษา บริษัทประกันภัย เป็นการศึกษาเกี่ยวกับเทคนิคการทำเหมืองข้อมูล ในการหาความสัมพันธ์การตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพ ด้วยเทคนิคกฎความสัมพันธ์ ใช้หารูปแบบของความสัมพันธ์การลาออกของพนักงาน และ สร้างตัวแบบพยากรณ์การลาออกของพนักงาน ด้วยการวิเคราะห์ 5 อัลกอริทึมคือต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม เบย์เซียน

แบบง่ายและเพื่อนบ้านใกล้เคียง และเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์เหล่านั้น เพื่อใช้เป็นตัวแบบในการลดปัญหาอัตราการลาออกของพนักงานและลดค่าใช้จ่ายที่สูงขึ้นจากการสรรหาพนักงานใหม่ (วทัญญุตา นีลาภาตระกูล และ ชุติมา เบี้ยวไข่มุข, 2562) และงานวิจัย ศศิมา มณฑาสุวรรณ (2557) ได้ศึกษาถึงการพัฒนาระบบค้นคว้ารหัส ICD - 10 สำหรับงานเวชระเบียน โดยนำเทคนิคต่าง ๆ มาทดลองและเปรียบเทียบค่าความถูกต้อง และนำเทคนิคที่ได้ค่าความถูกต้องมากที่สุดคือ อัลกอริทึมเบย์เซียนแบบง่าย มาประยุกต์ใช้ร่วมกับอัลกอริทึมอะพริโอริ เพื่อให้ได้ผลลัพธ์การค้นหาที่ดีขึ้น รวมถึงยังได้มีการรวบรวมจากงานวิจัยเรื่อง ปัจจัยที่มีผลต่อพฤติกรรมการบริโภคอาหาร ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในเขตอำเภอเมืองชุมพร จังหวัดชุมพร มีการศึกษาเกี่ยวกับปัจจัยที่มีผลต่อการบริโภคอาหารของนักเรียน พฤติกรรมการบริโภคอาหารของนักเรียนและ ความสัมพันธ์ ระหว่างปัจจัยที่มีผลต่อการบริโภคอาหารกับพฤติกรรมการบริโภคอาหารของนักเรียนกลุ่มตัวอย่าง นำมาวิเคราะห์ข้อมูลโดยหาค่าร้อยละ ความถี่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการทดสอบความสัมพันธ์ด้วย ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน เพื่อที่จะนำไปพัฒนาด้านสุขลักษณะการบริโภคอาหารของนักเรียนชั้นมัธยม (ศักดิ์อนันต์ รัตนสาครชัย, 2558)

ด้วยปัญหาเรื่องความต้องการที่แท้จริงในการบริโภคอาหารของนักศึกษาคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต ผู้วิจัยจึงได้นำกระบวนการทำเหมืองข้อมูลมาเป็นเครื่องมือในการค้นหาอุปสงค์ในการบริโภคอาหารของนักศึกษาเพื่อให้ได้คำตอบที่แม่นยำ และนำไปใช้ประโยชน์ได้ตามวัตถุประสงค์ที่ต้องการศึกษา

วัตถุประสงค์การวิจัย

1. เพื่อทราบถึงความต้องการที่แท้จริงของการบริโภคอาหารและเครื่องดื่ม รวมถึงเวลาที่ต้องการบริโภค
2. เพื่อหาความสัมพันธ์ระหว่างประเภทของอาหาร เครื่องดื่ม และเวลาในการรับประทาน

ขอบเขตงานวิจัย

1. ประชากรและกลุ่มตัวอย่าง ประชากร ได้แก่ กลุ่มนักศึกษา คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต จำนวน 1,018 คน กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้คือ กลุ่มนักศึกษาซึ่งทางผู้วิจัยได้มีการเก็บแบบสอบถาม ได้ค่าขนาดกลุ่มตัวอย่าง คือ 170 คน ผู้วิจัยได้ใช้สูตรการคำนวณกลุ่มตัวอย่างของทาโร่ ยามาเน่ ที่ระดับความเชื่อมั่นร้อยละ 93 หรือมีค่าความคลาดเคลื่อน $\pm$7%

2. สถานที่ดำเนินการวิจัย คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต

3. ระยะเวลาทำงานวิจัย ดำเนินการในช่วงระยะเวลาระหว่างเดือน มกราคม - มิถุนายน 2564

วิธีการดำเนินการวิจัย

ตามวัตถุประสงค์ในการศึกษานั้น จะใช้วิธีการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล โดยเลือกอัลกอริทึม 3 ชนิดมาใช้ในการประมวลผลเพื่อหาค่าความแม่นยำและค่าความถูกต้องของการจำแนกประเภท

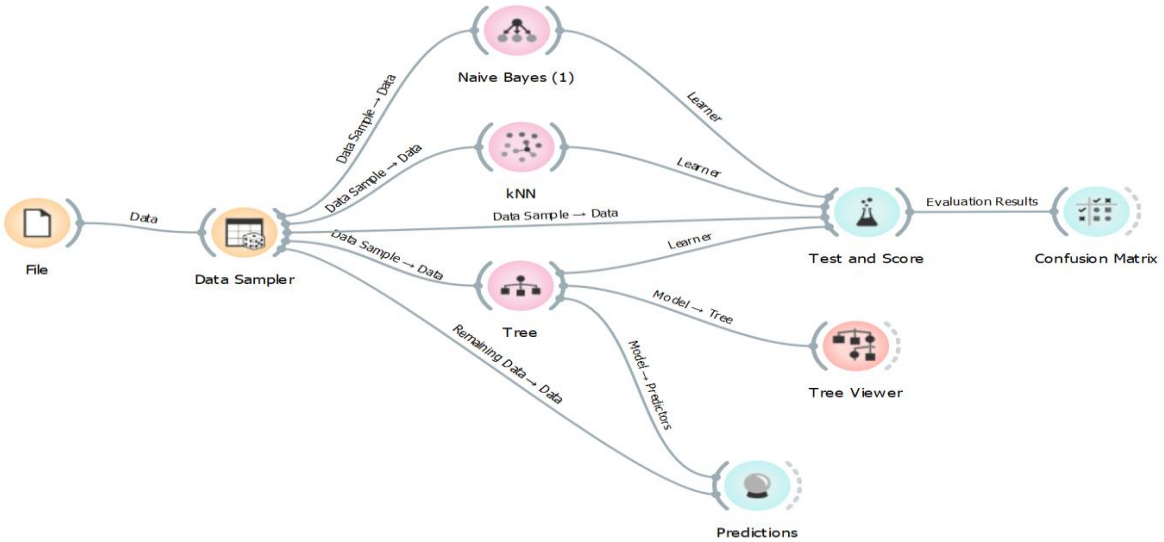
ข้อมูลที่สูงที่สุด และหาความสัมพันธ์ โดยวิธีการทำเหมืองข้อมูลแบบกฎความสัมพันธ์ โดยใช้อัลกอริทึมอะพริโอริ

คณะผู้วิจัยดำเนินการวิจัยโดย อ้างอิงขั้นตอนการทำเหมืองข้อมูลแบบ CRISP-DM (Cross-Industry Standard Process for Data Mining) ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

1. Business Understanding การทำความเข้าใจระบุปัญหา ผู้วิจัยมีความสนใจในเรื่องของอุปสงค์การบริโภคอาหารของนักศึกษา ภายในมหาวิทยาลัยราชภัฏภูเก็ต
2. Data Understanding การรวบรวมและทำความเข้าใจข้อมูล ในการวิจัยครั้งนี้ มีข้อมูลหลายประเภท เช่น ข้อมูล nominal, ordinal, interval, discrete เป็นต้น
3. Data Preparation การเตรียมข้อมูล จากชุดข้อมูลที่ได้จากแบบสอบถาม ก่อนจะที่นำมาเข้าสู่กระบวนการทำเหมืองข้อมูลจำเป็นต้องมีการจัดการ จำแนก หรือทำความสะอาดข้อมูลให้เหมาะสม และมีการหาค่าแอตทริบิวท์
4. Modeling การวิเคราะห์ข้อมูล เครื่องมือที่ใช้ในการดำเนินการทำเหมืองข้อมูลคือ โปรแกรม Orange เวอร์ชั่น 3.27.1 เพื่อหาความสัมพันธ์และความสอดคล้องของกฎที่ได้นำไปเปรียบเทียบและหาความสอดคล้องของผลลัพธ์ระหว่างการทำเหมืองข้อมูล 2 เทคนิค แบบจำแนกข้อมูล และ แบบกฎความสัมพันธ์
5. Evaluation การประเมินและวัดประสิทธิภาพของแบบจำลอง ได้ถูกนำมาเปรียบเทียบเพื่อยืนยันความสอดคล้องกับแบบจำลองของกฎความสัมพันธ์
6. Deployment การนำไปใช้งานจริง สามารถนำไปเป็นข้อมูลในการวางแผน จัดเตรียมวัตถุดิบในการปรุงอาหารตามเมนูที่สอดคล้องกับอุปสงค์ของนักศึกษาได้

ผลการวิจัย

เทคนิคการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูล โดยใช้อัลกอริทึม ต้นไม้ตัดสินใจ เบย์เซียนแบบง่าย และเพื่อนบ้านใกล้เคียง สามารถแสดงดังภาพที่ 1 โมเดลการจำแนกและหาความแม่นยำ มาใช้ในการหาคำตอบ โดยใช้ค่าความถูกต้อง (Accuracy) ค่าประสิทธิภาพโดยรวม (F1) ค่าความแม่นยำ (Precision) และค่าความครบถ้วน (Recall) มาเปรียบเทียบการประเมินที่ดีที่สุด ระหว่าง Percentage Split และ Cross Validation เพื่อเป็นเกณฑ์ในการเลือกอัลกอริทึมที่เหมาะสมที่สุด



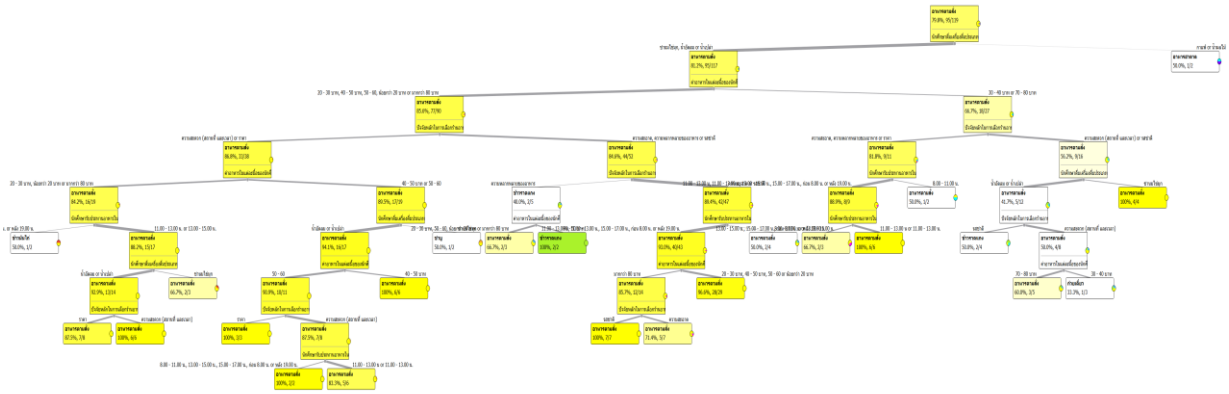
ภาพที่ 1 โมเดลการจำแนกและหาความแม่นยำ

จากการทดลองในการสร้างต้นแบบเพื่อหาความต้องการของประเภทอาหารพบว่า การใช้อัลกอริทึมต้นไม้ตัดสินใจ ให้ค่าประเมินความแม่นยำสูงกว่าอัลกอรึทึมอื่น และการแบ่งข้อมูลแบบ Percentage Split (70/30) ให้ค่าประเมินความแม่นยำสูงกว่า การแบ่งข้อมูลแบบ 5-Fold Validation ดังตารางที่ 1

ตารางที่ 1 เปรียบเทียบค่าการประเมินในการหาความต้องการของประเภทอาหาร

Technique and Evaluation	ค่าความถูกต้อง Accuracy	ค่าประสิทธิภาพโดยรวม F1	ค่าความแม่นยำ Precision	ค่าความครบถ้วน Recall
Decision Tree				
• Percentage Split (70/30)	**84.90**	**83.10**	**81.90**	**84.90**
• 5-Fold Cross Validation	60.30	60.90	61.60	60.30
K-Nearest Neighbor				
• Percentage Split (70/30)	79.80	70.90	63.70	79.80
• 5-Fold Cross Validation	77.20	67.30	59.60	77.20
Naïve Bayes				
• Percentage Split (70/30)	52.10	59.90	75.80	52.10
• 5-Fold Cross Validation	37.50	46.40	63.40	37.50

จากตารางที่ 1 แสดงค่าประเมินของการใช้อัลกอริทึมต้นไม้ตัดสินใจสูงที่สุด โดยมีค่าความถูกต้องร้อยละ 84.90 ค่าประเมินประสิทธิภาพร้อยละ 83.10 ค่าความแม่นยำร้อยละ 81.90 และค่าความครบถ้วน 84.90 ส่วน Naive Bayes ให้ค่าความถูกต้องและค่าประสิทธิภาพการประเมินต่ำที่สุด



ภาพที่ 2 แผนภูมิต้นไม้ตัดสินใจของการเลือกประเภทอาหาร

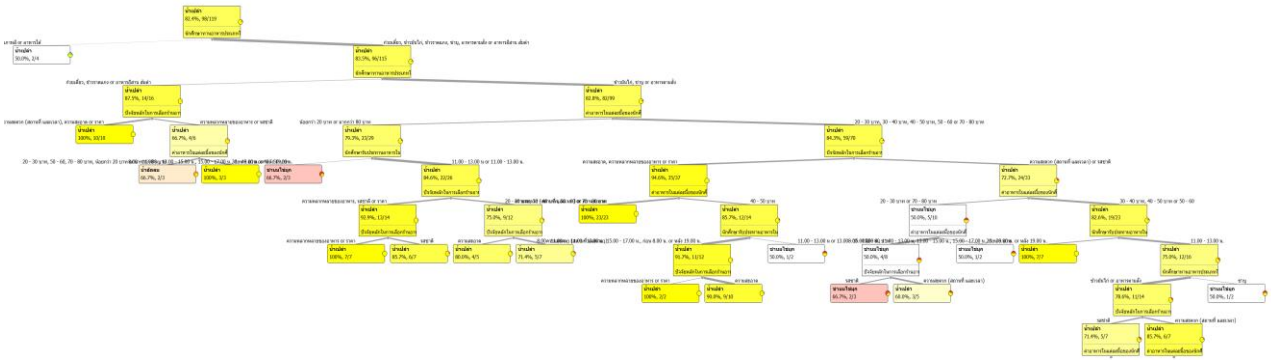
อัลกอริทึมที่มีประสิทธิภาพสูงสุดคือ ต้นไม้ตัดสินใจ โดยแสดงออกมาเป็นแผนภูมิ ดังภาพที่ 2 แผนภูมิต้นไม้ตัดสินใจของการเลือกประเภทอาหาร และมีค่าที่ถูกทำนายทั้งหมด 24 ค่า ซึ่งอาหารที่เป็นที่ต้องการสูงที่สุดของนักศึกษาคือ **อาหารตามสั่ง**

จากนั้นจึงทำการทดลองสร้างต้นแบบเพื่อหาความต้องการเครื่องดื่ม ให้ผลลัพธ์ไปในทิศทางเดียวกัน คือ การใช้อัลกอริทึมต้นไม้ตัดสินใจ ให้ค่าประเมินความแม่นยำสูงกว่าอัลกอรึทึมอื่น และการแบ่งข้อมูลแบบ Percentage Split (70/30) ให้ค่าความแม่นยำสูงกว่า การแบ่งข้อมูลแบบ 5-Fold Validation ดังตารางที่ 2

ตารางที่ 2 เปรียบเทียบค่าการประเมินในการหาความต้องการของเครื่องดื่ม

Technique and Evaluation	ค่าความถูกต้อง Accuracy	ค่าประสิทธิภาพโดยรวม F1	ค่าความแม่นยำ Precision	ค่าความครบถ้วน Recall
Decision Tree				
• Percentage Split (70/30)	**88.20**	**87.60**	**86.90**	**88.20**
• 5-Fold Cross Validation	66.20	68.50	71.40	66.20
K-Nearest Neighbor				
• Percentage Split (70/30)	82.40	74.40	67.80	82.40
• 5-Fold Cross Validation	81.60	74.00	67.70	81.60
Naïve Bayes				
• Percentage Split (70/30)	54.60	61.60	76.10	54.60
• 5-Fold Cross Validation	41.20	53.00	74.40	41.20

จากตารางที่ 2 แสดงค่าประเมินของการใช้อัลกอริทึมต้นไม้ตัดสินใจสูงที่สุด โดยมีค่าความถูกต้องร้อยละ 88.20 ค่าประเมินประสิทธิภาพร้อยละ 87.60 ค่าความแม่นยำร้อยละ 86.90 และค่าความครบถ้วน 88.20 ส่วน Naive Bayes ให้ค่าความถูกต้องและค่าประสิทธิภาพการประเมินต่ำที่สุด



ภาพที่ 3 แผนภูมิต้นไม้ตัดสินใจประเภทเครื่องดื่ม

อัลกอริทึมที่มีประสิทธิภาพสูงสุดคือ ต้นไม้ตัดสินใจ โดยแสดงออกมาเป็นแผนภูมิ ดังภาพที่ 3 แผนภูมิต้นไม้ตัดสินใจของการเลือกประเภทเครื่องดื่ม และมีค่าที่ถูกทำนายทั้งหมด 22 ค่า ซึ่งเครื่องดื่มที่เป็นที่ต้องการสูงที่สุดของนักศึกษาคือ **น้ำเปล่า**

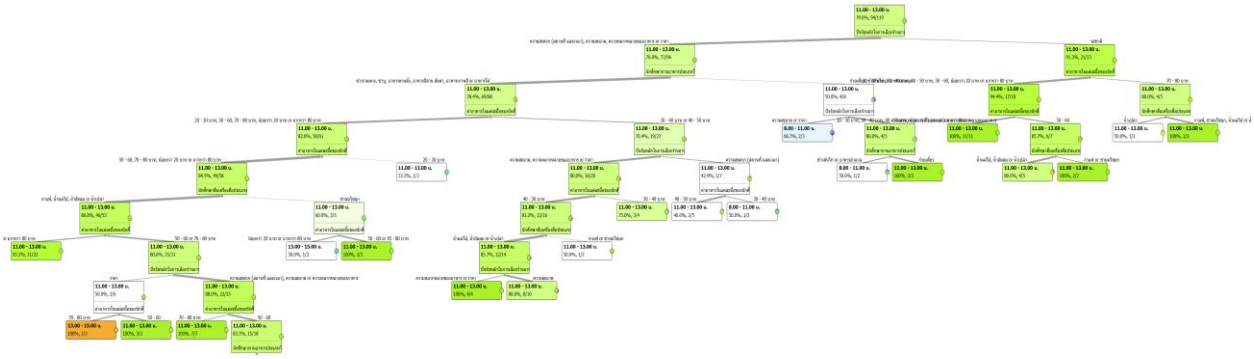
และเมื่อทำการทดลองสร้างต้นแบบเพื่อหาช่วงเวลาในการรับประทานอาหาร พบว่าผลลัพธ์ไปในทิศทางเดียวกัน คือ อัลกอริทึมต้นไม้ตัดสินใจให้ค่าประเมินความแม่นยำสูงกว่าอัลกอรึทึมอื่น การแบ่งข้อมูลแบบ Percentage Split (70/30) ให้ค่าความแม่นยำสูงกว่า การแบ่งข้อมูลแบบ 5-Fold Validation ดังตารางที่ 3

ตารางที่ 3 เปรียบเทียบค่าการประเมินในการหาช่วงเวลา

Technique and Evaluation	ค่าความถูกต้อง Accuracy	ค่าประสิทธิภาพโดยรวม F1	ค่าความแม่นยำ Precision	ค่าความครบถ้วน Recall
Decision Tree				
• Percentage Split (70/30)	**82.40**	**78.40**	**77.80**	**82.40**
• 5-Fold Cross Validation	72.80	72.00	71.20	72.80
K-Nearest Neighbor				
• Percentage Split (70/30)	79.00	69.70	62.40	79.00
• 5-Fold Cross Validation	81.60	73.40	66.60	81.60

Technique and Evaluation	ค่าความถูกต้อง Accuracy	ค่าประสิทธิภาพโดยรวม F1	ค่าความแม่นยำ Precision	ค่าความครบถ้วน Recall
Naïve Bayes				
• Percentage Split (70/30)	42.90	50.80	67.40	42.90
• 5-Fold Cross Validation	37.50	48.70	69.40	37.50

จากตารางที่ 3 แสดงค่าประเมินของการใช้อัลกอริทึมต้นไม้ตัดสินใจสูงที่สุด โดยมีค่าความถูกต้องร้อยละ 82.40 ค่าประเมินประสิทธิภาพร้อยละ 78.40 ค่าความแม่นยำร้อยละ 77.80 และค่าความครบถ้วน 82.40 ส่วน Naive Bayes ให้ค่าความถูกต้องและค่าประสิทธิภาพการประเมินต่ำที่สุด



ภาพที่ 4 แผนภูมิต้นไม้ตัดสินใจของช่วงเวลา

อัลกอริทึมที่มีประสิทธิภาพสูงสุดคือ ต้นไม้ตัดสินใจ โดยแสดงออกมาเป็นแผนภูมิ ดังภาพที่ 4 แผนภูมิต้นไม้ตัดสินใจของช่วงเวลา และมีค่าที่ถูกทำนายทั้งหมด 24 ค่า ซึ่งช่วงเวลาที่นักศึกษานิยมรับประทานอาหารมากที่สุดคือ **ช่วงเวลา 11:00-13:00 น.**

จากผลลัพธ์ในการหาความต้องการทางความหลากหลาย ทางทีมผู้วิจัยจึงต้องการหาค่าความสัมพันธ์เพิ่มเติมกรณีถ้านักศึกษาที่รับประทานอาหารตามสั่งนั้นจะมีแนวโน้มที่จะรับประทานอาหารอย่างอื่น และเครื่องดื่มด้วยหรือไม่ ดังนั้นจึงทำการทดลองโดยการทำเหมืองข้อมูลแบบกฎความสัมพันธ์ และใช้อัลกอริทึมอะพริโอริ มาหาคำตอบ ด้วยโมเดลกฎความสัมพันธ์ โดยกำหนดให้มีค่าสนับสนุน (Minimum Support) ร้อยละ 30 ค่าความเชื่อมั่น (Confidence) ร้อยละ 80 และค่าสหสัมพันธ์ (Lift) มากกว่า 1 ทำให้ได้กฎความสัมพันธ์มาจำนวน 30 กฎ โดยมีรายละเอียดดังตารางที่ 4 ตัวอย่างแสดงกฎความสัมพันธ์ที่สำคัญพร้อมค่าสนับสนุน (Support) ค่าความเชื่อมั่น (Confidence) และ ค่าสหสัมพันธ์ (Lift)

ตารางที่ 4 แสดงกฎความสัมพันธ์ที่สำคัญ พร้อมค่าสนับสนุน (Support) ค่าความเชื่อมั่น (Confidence) และ ค่าสหสัมพันธ์ (Lift)

ข้อ	กฎของความสัมพันธ์	ค่าสนับสนุน	ค่าความเชื่อมั่น	ค่าสหสัมพันธ์
1	น้ำอัดลม=1 ⟶ 11.00 - 13.00 น.=1	32.40	91.70	1.097
2	ข้าวราดแกง=1, 11.00 - 13.00 น.=1 ⟶ น้ำเปล่า=1	34.70	90.80	1.087
3	อาหารตามสั่ง=1, ข้าวราดแกง=1 ⟶ น้ำเปล่า=1	31.80	90.00	1.077
4	อาหารตามสั่ง=1, ก๋วยเตี๋ยว=1 ⟶ น้ำเปล่า=1	41.80	89.90	1.076
5	ก๋วยเตี๋ยว=1, น้ำเปล่า=1 ⟶ อาหารตามสั่ง=1	41.80	89.90	1.14
6	อาหารตามสั่ง=1, ก๋วยเตี๋ยว=1, 11.00 - 13.00 น.=1 ⟶ น้ำเปล่า=1	36.50	89.90	1.076
7	ก๋วยเตี๋ยว=1, น้ำเปล่า=1, 11.00 - 13.00 น.=1 ⟶ อาหารตามสั่ง=1	36.50	89.90	1.14
8	ข้าวราดแกง=1 ⟶ น้ำเปล่า=1	41.20	89.70	1.074
9	ก๋วยเตี๋ยว=1 ⟶ น้ำเปล่า=1	46.50	88.80	1.063
10	ก๋วยเตี๋ยว=1 ⟶ อาหารตามสั่ง=1	46.50	88.80	1.126

เมื่อนำมาเทียบกับกฎที่ได้จากอัลกอริทึมต้นไม้ตัดสินใจในการเลือกประเภทอาหาร มีความสอดคล้องกัน 5 กฎ คือ

1. ถ้าดื่มน้ำเปล่า และรับประทานอาหารช่วงเวลา 11.00 - 13.00 น. จะรับประทานอาหารตามสั่ง
2. ถ้าดื่มชานมไข่มุก และรับประทานอาหารช่วงเวลา 11.00 - 13.00 น. จะรับประทานอาหารตามสั่ง
3. ถ้ารับประทานอาหาช่วงเวลา 11.00 - 13.00 น. จะรับประทานอาหารตามสั่ง
4. ถ้าดื่มน้ำเปล่า จะรับประทานอาหารตามสั่ง
5. ถ้าดื่มชานมไข่มุก จะรับประทานอาหารตามสั่ง

เมื่อนำมาเทียบกับกฎที่ได้จากอัลกอริทึมต้นไม้ตัดสินใจในการเลือกเครื่องดื่ม มีความสอดคล้องกัน 6 กฎ คือ

1. ถ้ารับประทานก๋วยเตี๋ยว จะดื่มน้ำเปล่า
2. ถ้ารับประทานอาหารตามสั่ง จะดื่มน้ำเปล่า
3. ถ้ารับประทานข้าวราดแกง จะดื่มน้ำเปล่า
4. ถ้ารับประทานก๋วยเตี๋ยว ในช่วงเวลา 11:00-13:00 น. จะดื่มน้ำเปล่า
5. ถ้ารับประทานอาหารตามสั่ง ในช่วงเวลา 11:00-13:00 น. จะดื่มน้ำเปล่า

6. ถ้ารับประทานข้าวราดแกง ในช่วงเวลา 11:00-13:00 น. จะดื่มน้ำเปล่า

เมื่อนำมาเทียบกับกฎที่ได้จากอัลกอริทึมต้นไม้ตัดสินใจในการเลือกช่วงเวลา มีความสอดคล้องกัน 5 กฎ คือ

1. ถ้ารับประทานอาหารตามสั่งและดื่มน้ำเปล่า จะรับประทานอาหารช่วง 11.00 - 13.00 น.
2. ถ้ารับประทานอาหารตามสั่งและดื่มชานมไข่มุก จะรับประทานอาหารช่วง 11.00 - 13.00 น.
3. ถ้ารับประทานอาหารตามสั่ง จะรับประทานอาหารช่วง 11.00 - 13.00 น.
4. ถ้าดื่มน้ำเปล่า จะรับประทานอาหารช่วง 11.00 - 13.00 น.
5. ถ้าดื่มชานมไข่มุก จะรับประทานอาหารช่วง 11.00 - 13.00 น.

สรุป และอภิปรายผลการวิจัย

จากการทดลองการทำเหมืองข้อมูลโดยการจำแนกข้อมูล และ การใช้กฎความสัมพันธ์ มาหาความสอดคล้องนั้น สามารถสรุปได้ว่า การทำเหมืองข้อมูลแบบจำแนกข้อมูล ด้วยอัลกอรึทึมต้นไม้ตัดสินใจนั้นให้ค่าความถูกต้องออกมาดีที่สุดในทั้ง 3 แบบจำลอง ได้แก่

1. แบบจำลองจำแนกประเภทอาหาร โดยมีค่าความถูกต้องร้อยละ 84.90 ค่าประเมินประสิทธิภาพร้อยละ 83.10 ค่าความแม่นยำร้อยละ 81.90 และค่าความครบถ้วน 84.90 ได้กฎ 24 กฎ

2. แบบจำลองจำแนกเครื่องดื่ม โดยมีค่าความถูกต้องร้อยละ 88.20 ค่าประเมินประสิทธิภาพร้อยละ 87.60 ค่าความแม่นยำร้อยละ 86.90 และค่าความครบถ้วน 88.20 ได้กฎ 22 กฎ

3. แบบจำลองจำแนกช่วงเวลา โดยมีค่าความถูกต้องร้อยละ 82.40 ค่าประเมินประสิทธิภาพร้อยละ 78.40 ค่าความแม่นยำร้อยละ 77.80 และค่าความครบถ้วน 82.40 ได้กฎ 24 กฎ

ส่วนการใช้กฎความสัมพันธ์ เพื่อหาความสอดคล้องระหว่างข้อมูล โดยกำหนดให้มีค่าสนับสนุน (Minimum Support) ร้อยละ 30 ค่าความเชื่อมั่น (Confidence) ร้อยละ 80 และค่าสหสัมพันธ์ (Lift) มากกว่า 1 ทำให้ได้กฎความสัมพันธ์มาจำนวน 30 กฎ และพบกฎที่สอดคล้องกัน 16 กฎ

จากการทดลองทำให้ทราบถึงประเภทของอาหารและเครื่องดื่มที่เป็นที่ต้องการของนักศึกษาและสามารถนำไปใช้ต่อยอดในการลงทุนได้ว่าหากต้องการที่จะทำร้านอาหารในโรงอาหารของมหาวิทยาลัย หรือตามคณะต่างๆ จะเห็นได้ว่าอาหารตามสั่ง ก๋วยเตี๋ยว ข้าวราดแกง น้ำเปล่า และชาไข่มุก ยังมีอุปสงค์ในการบริโภคอาหารและเครื่องดื่มเหล่านี้อยู่ ในขณะเดียวกันทางมหาวิทยาลัยก็สามารถที่จะจัดสรรโควต้าร้านอาหารและเครื่องดื่มในการประมูลเช่าได้ถูกต้องตรงตามความต้องการของนักศึกษา ในส่วนของเรื่องราคาอาหารและเครื่องดื่ม จากการทดลองความสอดคล้องราคาไม่ได้มีผลต่อการติดสินใจในการบริโภคมากนัก เนื่องจากราคาอาหารที่จำหน่ายอยู่ในร้านค้าในมหาวิทยาลัยมีราคาอยู่ที่ประมาณ 20-80 บาท ซึ่งอยู่ในวิสัยที่นักศึกษามีกำลังซื้อตามปกติอยู่แล้ว

ข้อเสนอแนะ

ในการทดลองการทำเหมืองข้อมูลแบบจำแนกประเภทข้อมูลและการประเมินโมเดล ทำให้พบว่า ในกรณีที่ค่าความถูกต้องมีแนวโน้มสูง ค่าความครบถ้วนสูง ก็จะทำให้ค่าประสิทธิภาพโดยรวมก็จะมีแนวโน้มสูงตามไปด้วยตามหลักของ Confusion Matrix ซึ่งเป็นตารางแบบสมมาตร ที่ใช้ในการแสดงค่าการวิเคราะห์ในการทำนายของการทำเหมืองข้อมูลแบบจำแนกข้อมูล (ธาดา จันตะคุณ, 2559)

ในงานวิจัยครั้งนี้ข้อมูลที่ได้ทำการรวบรวม และนำมาใช้หลังจากเข้ากระบวนการ Data Pre-processing แล้วพบว่า ข้อมูลมีการกระจายตัวที่ไม่เท่าเทียมกันและไม่สมดุล หรือที่เรียกว่า Imbalanced Data ทำให้ผลลัพธ์ส่วนใหญ่ที่ออกมาอยู่ที่ "อาหารตามสั่ง" ดังนั้นจึงควรมีการทำการ Data Wrangling ใหม่ โดย เทคนิค Sampling Method ซึ่งเป็นวิธีทางสถิติ จากปัญหาที่พบในข้อมูลนี้นั้นจึงควรใช้ Sampling Method แบบ Over-Sampling คือการสุ่มเพิ่มจำนวนกลุ่มน้อยให้เยอะขึ้นพอๆ กับกลุ่มหลัก เพื่อทำให้ข้อมูลมีความสมดุลและสามารถทำนายค่าได้ถูกต้องแม่นยำขึ้น

สำหรับคำแนะนำในกรณีการศึกษาวิจัยต่อเนื่องในประเด็นที่เกี่ยวข้อง ทางผู้วิจัยเห็นว่า ด้วยข้อจำกัดในเรื่องระยะเวลาการทำวิจัยและกลุ่มตัวอย่างครั้งนี้ ทำให้ผลลัพธ์ที่ได้ออกมายังไม่ละเอียดเพียงพออย่างที่คาดหวัง แต่สามารถมองเห็นช่องทางในการพัฒนาต่อยอดงานวิจัย ด้วยการทำให้ข้อมูลมีความจำเพาะเจาะจงและชัดเจนยิ่งขึ้น โดยศึกษาต่อในรายละเอียด จำแนกข้อมูลที่ได้จากผลลัพธ์ของการวิจัยครั้งนี้ ให้มีความแยกย่อยชัดเจนเพิ่มเติม โดยการสำรวจระบุชนิดของ อาหารตามสั่ง ที่เจาะจงมากขึ้น เช่น ข้าวราดผัดกระเพรา หมู ไก่ กุ้ง ฯลฯ เป็นต้น หรือในส่วนของระยะเวลารับประทานอาหาร เช่น แบ่งช่วงเวลาให้สั้นลงสอดคล้องกับการใช้เวลารับประทานจริง ประมาณ 1 ชั่วโมง เป็นช่วง 11.00 - 12.00 และ 12.00 -13.00 น. เป็นต้น ก็น่าจะทำให้ได้ข้อมูลที่เป็นประโยชน์ ให้ผู้ค้าสามารถเข้าใจพฤติกรรมของนักศึกษาและนำไปปรับเมนูอาหาร การเตรียมวัตถุดิบให้ตรงกับความต้องการของลูกค้า และการเตรียมการรองรับช่วงเวลาที่ลูกค้าหนาแน่น ได้อย่างเหมาะสมต่อไป

เอกสารอ้างอิง

ธาดา จันตะคุณ. (2559). *การพยากรณ์ความเป็นไปได้ในการเลือกสมัครสาขาวิชาโดยใช้เทคนิคเหมืองข้อมูล*. (งานวิจัย). มหาวิทยาลัยราชภัฏมหาสารคาม, คณะวิทยาศาสตร์และเทคโนโลยี, สาขาวิชาระบบสารสนเทศเพื่อการจัดการ.

ศักดิ์อนันต์ รัตนสาครชัย.(2558). *ปัจจัยที่มีผลต่อพฤติกรรมการบริโภคอาหารของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในเขตอำเภอเมืองชุมพร จังหวัดชุมพร*. ศักดิ์อนันต์ รัตนสาครชัย สำนักงานสาธารณสุขจังหวัดชุมพร.

วทัญญุตา นีลาภาตระกูล, ชุติมา เบี้ยวไข่มุข. (2562). *การศึกษาปัจจัยที่สัมพันธ์กับการตัดสินใจลาออกและการเปรียบเทียบประสิทธิภาพตัวแบบพยากรณ์การลาออกของพนักงาน กรณีศึกษา บริษัทประกันภัย.* สมาคมสถาบันอุดมศึกษาเอกชนแห่งประเทศไทย 2562. 8(1): 46-63.

ศศิมา มณฑาสุวรรณ. (2557). *การพัฒนาระบบค้นคว้ารหัส ICD - 10 สำหรับงานเวชระเบียน* (วิทยานิพนธ์ปริญญา มหาบัณฑิต). มหาวิทยาลัยศิลปากร, คณะศิลปศาสตร์, สาขาวิชาสนเทศศาสตร์เพื่อการศึกษา

ศุภามณ จันทร์สกุล. (2561). *เทคนิคเหมืองข้อมูลในการวิเคราะห์ข้อมูลทางการพยาบาล.* วารสารวิชาการมหาวิทยาลัยอีสเทิร์นเอเชีย, 12(2).