

การเปรียบเทียบตัวแบบในการทำนายปริมาณยอดขายเบียร์ :
กรณีศึกษาเบียร์ยี่ห้อหนึ่ง
Model Comparison for Predicting Beer Sales Volume:
a Case Study of a Certain Brand of Beer

จันทร์จิรา พิลาดัง^{1*} ดวงเดือน พรหมวัฒน์² และ วิกานดา ผาพันธุ์^{2,3}

Janjira Piladaeng^{1*} Duangduen Promwat² and Wikanda Phaphan^{2,3}

¹ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา จ.ชลบุรี ประเทศไทย

²ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพฯ ประเทศไทย

³กลุ่มวิจัยการเรียนรู้เชิงสถิติและการอนุมาน มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพฯ ประเทศไทย

¹Department of Mathematics, Faculty of Science, Burapha University, Chon Buri, Thailand

²Department of Applied Statistics, Faculty of Applied Science,

King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

³Research Group in Statistical Learning and Inference,

King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

วันที่ส่งบทความ : 14 กรกฎาคม 2566 วันที่แก้ไขบทความ : 28 มิถุนายน 2567 วันที่ตอบรับบทความ : 28 มิถุนายน 2567

Received: 14 July 2023, Revised: 28 June 2024, Accepted: 28 June 2024

บทคัดย่อ

ตลาดเบียร์เป็นตลาดขนาดใหญ่ที่มีความสำคัญในอุตสาหกรรมเครื่องดื่มแอลกอฮอล์ การทำนายปริมาณยอดขายเบียร์ได้อย่างแม่นยำจึงเป็นสิ่งสำคัญสำหรับผู้ประกอบการนำไปใช้ในการวางแผนการผลิต และกำหนดกลยุทธ์ทางการตลาด ด้วยเหตุนี้ งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาตัวแบบในการทำนายปริมาณยอดขายเบียร์ยี่ห้อหนึ่ง ด้วยวิธีการวิเคราะห์การถดถอยพหุคูณ วิธีต้นไม้ตัดสินใจ และวิธีแบบป่าสุ่ม พร้อมเปรียบเทียบความแม่นยำของตัวแบบด้วยค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ข้อมูลที่ใช้ในงานวิจัยเป็นข้อมูลทุติยภูมิรายเดือน ระหว่างปี พ.ศ. 2561 – 2564 ซึ่งมีตัวแปรทำนายทั้งหมด 12 ตัวแปร และทำการวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรมภาษา R จากผลลัพธ์ พบว่า ตัวแบบที่เหมาะสมและมีประสิทธิภาพในการทำนายดีที่สุด คือ ตัวแบบการถดถอยเชิงเส้นพหุคูณ ที่มีการคัดเลือกตัวแปรทำนายแบบขั้นตอน ด้วยเกณฑ์ข้อสนเทศของอาไคเคะ และมีตัวแปรทำนายที่มีผลต่อการทำนายปริมาณยอดขายเบียร์ ได้แก่ 1) ปริมาณยอดขายเบียร์คู่แข่งที่ 1 2) ราคาขายทอดที่ 3 ของคู่แข่งที่ 1 3) ราคาขายทอดที่ 3 ของคู่แข่งที่ 2 4) ดัชนีค่าปลีก 5) ดัชนีราคาผู้บริโภค หมวดยาสูบและเครื่องดื่มมีแอลกอฮอล์ และ 6) จำนวนประชากรไทยที่มีอายุ 15 ปีขึ้นไป

คำสำคัญ : การบริโภคเบียร์ อิทธิพล การถดถอยพหุคูณ ต้นไม้ตัดสินใจ ป่าสุ่ม

*ที่อยู่ติดต่อ E-mail address: janjira.pi@buu.ac.th

<https://doi.org/10.55003/scikmitl.2024.259727>

Abstract

The beer market is a sizable and important market in the alcoholic beverage industry. Accurate prediction of beer sales volume is crucial for entrepreneurs to employ in production planning and develop marketing strategies. As a result, this research aims to study models for predicting the sales volume of a certain brand of beer using multiple linear regression, decision trees, and random forest methods. The accuracy of the models is compared using the Root Mean Square Error (RMSE). The study utilizes monthly data from 2018 to 2021, consisting of 12 independent variables. The statistical analysis is conducted using the R programming language. The results reveal that the most suitable and efficient model for prediction is the multiple linear regression model, which uses a stepwise variable selection based on the Akaike Information Criterion (AIC). The significant independent variables influencing beer sales prediction include 1) the sales volume of the 1st competitor, 2) the third-tier selling price of the 1st competitor, 3) the third-tier selling price of the 2nd competitor, 4) the retail trade index, 5) the consumer price index for tobacco and alcoholic beverages, and 6) the population of Thailand aged 15 and above.

Keywords: Beer consumption, Influence, Multiple regression, Decision tree, Random forest

1. บทนำ

เบียร์เป็นเครื่องดื่มที่ได้รับความนิยมสูงทั่วโลก รวมทั้งในประเทศไทย การบริโภคเบียร์มีบทบาทสำคัญในวัฒนธรรมและรูปแบบการดำเนินชีวิตของผู้คนในหลายประเทศ นอกจากนี้ยังมีบทบาทสำคัญในเศรษฐกิจของประเทศด้วย การผลิตและจำหน่ายเบียร์เป็นอุตสาหกรรมใหญ่ที่มีความสำคัญของผู้ประกอบการในหลายประเทศ ในประเทศไทยนั้น เบียร์ถือเป็นผลิตภัณฑ์เครื่องดื่มแอลกอฮอล์ที่มีปริมาณการบริโภคและมูลค่าตลาดสูงสุด ซึ่งคิดเป็นร้อยละ 71.3 และ 54.3 ในเชิงปริมาณและมูลค่าของตลาดเครื่องดื่มแอลกอฮอล์ทั้งหมด ในอุตสาหกรรมเบียร์ ตลาดเบียร์เกือบทั้งหมดถูกครอบครองโดยผู้ผลิตรายใหญ่ 2 ราย ซึ่งมีศักยภาพทางธุรกิจสูง มีความได้เปรียบในด้านเงินทุน และมีการกระจายสินค้าผ่านช่องทางจำหน่ายที่มีประสิทธิภาพ ซึ่งมีส่วนแบ่งตลาดรวมกันมากถึง 95% ของปริมาณการจำหน่ายเบียร์ในประเทศ [1]

นอกจากนี้ สถาบันอาหารได้รายงานตลาดเครื่องดื่มแอลกอฮอล์ [2] ในช่วงปี พ.ศ. 2561 – 2565 พบว่าปี พ.ศ. 2563 เป็นปีที่ตลาดเบียร์หดตัวมากที่สุด เนื่องจากมาตรการการควบคุมการระบาดของโรคโควิด-19 ที่ห้ามจำหน่ายเครื่องดื่มแอลกอฮอล์ในร้านอาหาร สถานบันเทิง หรือสถานที่อื่น ๆ ที่เป็นช่องทางการจัดจำหน่ายที่สำคัญ หลังจากที่มีการผ่อนคลายมาตรการต่าง ๆ ในปลายปี พ.ศ. 2564 จึงทำให้ตลาดเบียร์ขยายตัวเพิ่มขึ้นในปี พ.ศ. 2565 ร้อยละ 13.9 (จากปี พ.ศ. 2563) ซึ่งแสดงให้เห็นถึงการเพิ่มขึ้นของปริมาณยอดขายเบียร์ในตลาด ดังนั้น การทำนายปริมาณยอดขายเบียร์เป็นสิ่งสำคัญอย่างมากสำหรับ

ผู้ประกอบการในตลาด เนื่องจากการทำนายที่แม่นยำจะช่วยให้ผู้ประกอบการสามารถวางแผนกลยุทธ์ทางการตลาดและการผลิตได้อย่างมีประสิทธิภาพ ปริมาณยอดขายเบียร์เป็นตัวชี้วัดที่สำคัญสำหรับการบริโภคเบียร์ในตลาด และเป็นตัวบ่งบอกถึงค่านิยมของเบียร์ในช่วงเวลานั้น ๆ การที่ผู้ประกอบการสามารถทำนายตลาดในธุรกิจนี้ได้แม่นยำจำเป็นจะต้องศึกษาปัจจัยที่เกี่ยวข้องที่มีผลต่อปริมาณยอดขาย รวมถึงการเลือกใช้ตัวแบบทำนายที่เหมาะสม

ในการศึกษาตัวแบบทำนายที่มีปัจจัยหรือตัวแปรอิสระเข้ามาเกี่ยวข้องนั้น มีเทคนิคทางสถิติมากมายให้เลือกใช้ในการวิเคราะห์ข้อมูล เพื่อสร้างตัวแบบทำนายที่มีความแม่นยำ ในปัจจุบันได้มีการใช้เทคโนโลยีการเรียนรู้ของเครื่อง (Machine learning) เพื่อช่วยในการสร้างตัวแบบเหล่านั้น โดยทั่วไป วิธีที่เป็นที่รู้จักและนิยมใช้ในการศึกษาปัจจัยและสร้างตัวแบบการทำนาย คือ การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression: MLR) นอกจากนี้ยังมีเทคนิคการเรียนรู้อื่น ๆ เช่น การถดถอยต้นไม้ตัดสินใจ (Decision Tree Regression: DTR) และการถดถอยแบบป่าสุ่ม (Random Forest Regression: RFR) เป็นต้น ในการศึกษาที่ผ่านมา ได้มีการประยุกต์ใช้การเรียนรู้ของเครื่องในงานวิจัยหลายประเภท เช่น งานวิจัยด้านการตลาด การแพทย์ และวิทยาศาสตร์ เป็นต้น ในการศึกษาของ พรทิศา วิเศษสุวรรณ์ และ เอกสิทธิ์ พัทธวงศ์ศักดิ์ [3] ได้ทำการเปรียบเทียบระหว่างตัวแบบ MLR ตัวแบบ DTR และตัวแบบ RFR โดยใช้ข้อมูลยอดขายของร้านกาแฟและข้อมูลรายงานสถานการณ์โควิด-19 ระลอก 3 พบว่า ตัวแบบ RFR ให้ค่าเฉลี่ยร้อยละความผิดพลาดสัมบูรณ์ (Mean Absolute Percent Error: MAPE) ที่ต่ำที่สุด นอกจากนี้ในการศึกษาเกี่ยวกับการเปรียบเทียบรายได้รายวันของโรงพยาบาลแห่งหนึ่ง โดยใช้การเรียนรู้ของเครื่องและตัวแบบอาร์มา (ARIMA) ของสิทธิชัย นาคพิทักษ์ และวิภาดา ผาพันธ์ [4] ยังพบว่า เทคนิคการเรียนรู้ของเครื่องที่มีตัวแปรทำนายมาเกี่ยวข้องนั้น มีความผิดพลาดของตัวแบบการทำนายที่ต่ำกว่าตัวแบบอาร์มาที่ไม่มีตัวแปรอิสระมาเกี่ยวข้อง ในขณะที่ Čeh และคณะ [5] ได้ศึกษาประสิทธิภาพของตัวแบบ RFR เทียบกับตัวแบบ MLR ในการทำนายราคาอะพาร์ตเมนต์ ซึ่งพบว่าเทคนิค RFR มีประสิทธิภาพในการทำนายมากกว่า และยังพบว่าตัวแบบ MLR มีค่า R^2 ต่ำมาก แสดงว่าตัวแบบการถดถอยไม่เชิงเส้นพหุคูณ (Multiple Nonlinear Regression) น่าจะมีความเหมาะสมมากกว่าตัวแบบ MLR ในทำนองเดียวกัน Krkač และคณะ [6] ได้ทำการเปรียบเทียบวิธี RFR กับ MLR ในการทำนายความเร็วของแผ่นดินถล่ม ซึ่งผลลัพธ์ที่ได้นั้น พบว่า เทคนิค RFR นั้นดีกว่า MLR เพียงเล็กน้อย ซึ่งพิสูจน์ได้ว่าการวิเคราะห์ด้วยตัวแบบ MLR ยังมีศักยภาพในการพยากรณ์การเคลื่อนที่ของแผ่นดินถล่ม จากที่กล่าวมาข้างต้น จะเห็นได้ว่าวิธีการเรียนรู้ของเครื่องโดยใช้ตัวแบบ RFR นั้นมีประสิทธิภาพดีกว่าตัวแบบ MLR และตัวแบบ DTR แต่อย่างไรก็ตาม การเลือกใช้ตัวแบบที่มีความเหมาะสมกับข้อมูลเป็นสิ่งที่สำคัญที่สุด เพื่อให้การทำนายข้อมูลมีประสิทธิภาพและความแม่นยำสูง

สำหรับการศึกษาปัจจัยที่เกี่ยวข้องกับปริมาณยอดขายหรือปริมาณการบริโภคเบียร์นั้น เบญจวรรณ ยุทธโท [7] ได้ทำการวิเคราะห์ความต้องการปริมาณการบริโภคสินค้าแอลกอฮอล์ ซึ่งใช้แบบจำลอง Vector Autoregression (VAR) ในการวิเคราะห์ถึงสาเหตุที่ทำให้ความต้องการเบียร์ ABC ขวดใหญ่ ในช่วงเวลาที่ผ่านมามีความผันผวนสูง โดยใช้ข้อมูลอนุกรมเวลารายเดือนและตัวแปรหุ่น (Dummy variable) ร่วมกับตัวแปรอื่น ๆ ที่ส่งผลต่อการเปลี่ยนแปลงปริมาณความต้องการบริโภคเบียร์ จากผลการศึกษา พบว่า โปรโมชัน 120 ลิ้ง แกม 10 ลิ้ง การปรับรสชาติและการปรับฉลากบรรจุภัณฑ์ และการขึ้นราคาสินค้าหรือการขึ้นภาษี

สินค้าแอลกอฮอล์ ส่งผลต่อความต้องการปริมาณเบียร์ ABC อย่างมีนัยสำคัญทางสถิติ นอกจากนี้ อธิวัฒน์ ลากหลาย และคณะ [8] ได้ใช้การวิเคราะห์ MLR ในการสร้างตัวแบบเพื่อศึกษาถึงปัจจัยที่มีผลต่อการบริโภคเบียร์ในประเทศไทย โดยตัวแปรในตัวแบบอ้างอิงจากทฤษฎีอุปสงค์เป็นหลัก จากการศึกษา พบว่าราคาตลาดเบียร์ รายได้ผู้บริโภค ช่วงเวลาแต่ละไตรมาส มาตรการของรัฐบาลบางมาตรการ และปริมาณการบริโภคเบียร์ในไตรมาสที่ผ่านมา ที่มีผลต่อปริมาณการบริโภคเบียร์ในประเทศไทยอย่างมีนัยสำคัญทางสถิติ

ในงานวิจัยนี้ ผู้วิจัยจึงมุ่งเน้นไปที่การศึกษาตัวแบบการทำนายปริมาณยอดขายเบียร์ของผู้ประกอบการเบียร์ยี่ห้อหนึ่ง เนื่องจากความแตกต่างกันของเทคนิคการเรียนรู้ของเครื่องที่นำมาใช้ในการทำนาย โดยตัวแบบ MLR เป็นวิธีการทั่วไปที่นิยมใช้ศึกษาถึงความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรทำนาย ในขณะที่ DTR และ RFR เป็นวิธีการทำนายด้วยการจำแนกข้อมูล โดยการสร้างเงื่อนไขเพื่อช่วยตัดสินใจจากข้อมูล และทำนายผลที่จะเกิดขึ้น [9] จึงได้ทำการเปรียบเทียบตัวแบบการทำนายตามวิธีการเรียนรู้ของเครื่อง 3 ตัวแบบ ได้แก่ ตัวแบบ MLR ตัวแบบ DTR และตัวแบบ RFR เพื่อหาตัวแบบที่เหมาะสมที่สุด และตัวแปรใดบ้างที่มีอิทธิพลต่อปริมาณยอดขายดังกล่าว โดยตัวแบบที่เหมาะสมที่สุดควรมีค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Squares Error: RMSE) ที่ต่ำที่สุด ซึ่งแสดงถึงความผิดพลาดในการทำนายที่น้อยที่สุด

2. วิธีดำเนินการวิจัย

ในการดำเนินการวิจัย ผู้วิจัยได้แบ่งชุดข้อมูลออกเป็น 2 ชุด คือ ชุดข้อมูลที่ใช้สำหรับสอน (Training data) ซึ่งเป็นชุดข้อมูลที่ใช้ในการสร้างตัวแบบ โดยใช้ข้อมูล 80% ของข้อมูลทั้งหมด และชุดข้อมูลที่ใช้ทดสอบ (Test data) ซึ่งเป็นชุดข้อมูลที่ใช้ในการตรวจสอบความแม่นยำของตัวแบบ โดยใช้ข้อมูล 20% ที่เหลือจากชุดข้อมูลที่ใช้สำหรับสอน ทั้งนี้ใช้เกณฑ์ RMSE ที่ต่ำที่สุดในการประเมินความแม่นยำของตัวแบบ และใช้โปรแกรม R เวอร์ชัน 4.3.1 ในการประมวลผล

2.1 ข้อมูลที่ใช้ในการวิจัย

การทำนายปริมาณยอดขายเบียร์ยี่ห้อหนึ่งเกี่ยวข้องกับตัวแปรต่าง ๆ ที่อาจจะมีผลกระทบต่อยุทธศาสตร์เบียร์ โดยต้องพิจารณาปัจจัยทั้งภายในและภายนอกบริษัท เน้นอนว่าข้อมูลของคู่แข่งทางการตลาดก็เป็นสิ่งสำคัญ หรือเป็นปัจจัยที่คาดว่าจะมีผลกระทบต่อยุทธศาสตร์เบียร์มากที่สุด รวมไปถึงการจัดทำกลยุทธ์ของทางบริษัทที่อาจจะมีผลต่อความสนใจของลูกค้า และความไม่พร้อมของสินค้าที่อาจทำให้ปริมาณยอดขายลดลงได้ หากมีปริมาณสินค้าไม่เพียงพอต่อความต้องการ และข่าวการปรับราคาสินค้าหรือปรับภาษีแอลกอฮอล์ก็อาจเป็นอีกสาเหตุหนึ่งที่ทำให้ปริมาณการบริโภคลดลงได้เช่นเดียวกัน

นอกจากนี้ ตัวชี้วัดที่บอกถึงการเปลี่ยนแปลงราคาสินค้าและบริการโดยเฉลี่ยที่ผู้บริโภคจ่ายไปสำหรับกลุ่มสินค้าและบริการที่กำหนด และการวัดค่าการเปลี่ยนแปลงในมูลค่ายอดขายทั้งหมดในระดับการค้าปลีก ยังเป็นดัชนีที่สำคัญมากที่บ่งชี้ถึงการใช้จ่ายของผู้บริโภค ซึ่งคิดเป็นส่วนสำคัญมากที่สุดของกิจกรรมทางเศรษฐกิจโดยรวม

อย่างไรก็ตาม สถานการณ์การบริโภคเครื่องดื่มประเภทแอลกอฮอล์ในปัจจุบัน ได้มีการเปลี่ยนแปลงไปตามการแพร่ระบาดของโรคโควิด-19 ไม่ว่าจะเป็นช่วงที่มีเหตุการณ์ห้ามจำหน่ายหรือห้ามดื่มแอลกอฮอล์

ในร้าน และเปลี่ยนแปลงตามโครงสร้างของประชากร โดยจากการสำรวจพฤติกรรมด้านสุขภาพของประชากร พ.ศ. 2564 โดยสำนักงานสถิติแห่งชาติ พบว่า ในปี พ.ศ. 2564 ประชากรไทยอายุตั้งแต่ 15 ปีขึ้นไป มีจำนวน 15.96 ล้านคน หรือร้อยละ 28.00 เป็นนักดื่มในปัจจุบัน [10]

จากที่กล่าวมาข้างต้น ผู้วิจัยจึงได้มีการรวบรวมข้อมูลจากแหล่งต่าง ๆ หลายแหล่งเพื่อใช้ในการศึกษาตัวแปรในการทำนายปริมาณยอดขายเบียร์ยี่ห้อหนึ่ง ได้แก่ ข้อมูลจากผู้ประกอบการ ศูนย์วิจัยปัญหาสุรา [11] ธนาคารแห่งประเทศไทย [12] สำนักดัชนีเศรษฐกิจการค้า กระทรวงพาณิชย์ [13] และสำนักบริหารการทะเบียน กระทรวงมหาดไทย [14] โดยข้อมูลที่รวบรวมมานั้นเป็นข้อมูลทุติยภูมิแบบรายเดือน ตั้งแต่เดือนมกราคม พ.ศ. 2561 – เดือนธันวาคม พ.ศ. 2564 ทั้งหมด 48 รายการ ซึ่งประกอบไปด้วย ตัวแปรตาม (Y) แทนปริมาณยอดขายเบียร์ยี่ห้อหนึ่ง และตัวแปรอิสระหรือตัวแปรทำนาย (X) ที่ใช้ในการศึกษาครั้งนี้ ทั้งหมด 12 ตัวแปร ดังแสดงในตารางที่ 1

ตารางที่ 1. ตัวแปรตามและตัวแปรทำนาย

ตัวแปร	คำอธิบายตัวแปร	หน่วย	ประเภทของข้อมูล	แหล่งที่มา
Y	ปริมาณยอดขายเบียร์ยี่ห้อหนึ่ง	ลิตร	เชิงปริมาณ	ผู้ประกอบการ
X ₁	ปริมาณยอดขายเบียร์คู่แข่งที่ 1	ลิตร	เชิงปริมาณ	ผู้ประกอบการ
X ₂	ราคาขายทอดที่ 3 ของคู่แข่งที่ 1	บาท	เชิงปริมาณ	ผู้ประกอบการ
X ₃	ราคาขายทอดที่ 3 ของคู่แข่งที่ 2	บาท	เชิงปริมาณ	ผู้ประกอบการ
X ₄	ราคาขายทอดที่ 3 ของคู่แข่งที่ 3	บาท	เชิงปริมาณ	ผู้ประกอบการ
X ₅	การทำกลยุทธ์ที่เกี่ยวกับราคา (1 = ช่วงที่เกิดเหตุการณ์ 0 = ช่วงที่ไม่เกิดเหตุการณ์)	-	เชิงคุณภาพ	ผู้ประกอบการ
X ₆	การทำกลยุทธ์ที่เกี่ยวกับการให้เครดิตพิเศษ (1 = ช่วงที่เกิดเหตุการณ์ 0 = ช่วงที่ไม่เกิดเหตุการณ์)	-	เชิงคุณภาพ	ผู้ประกอบการ
X ₇	ความพร้อมของสต็อกสินค้า	-	เชิงปริมาณ	ผู้ประกอบการ
X ₈	ข่าวขึ้นราคาสินค้าหรือปรับภาษีแอลกอฮอล์รวมถึงข่าวที่เกี่ยวข้อง (1 = ช่วงที่เกิดเหตุการณ์ 0 = ช่วงที่ไม่เกิดเหตุการณ์)	-	เชิงคุณภาพ	ผู้ประกอบการ
X ₉	ช่วงที่มีเหตุการณ์ห้ามจำหน่ายหรือห้ามดื่มแอลกอฮอล์ในร้าน (1 = ช่วงที่เกิดเหตุการณ์ 0 = ช่วงที่ไม่เกิดเหตุการณ์)	-	เชิงคุณภาพ	ศูนย์วิจัยปัญหาสุรา [11]
X ₁₀	ดัชนีค่าปลีก	-	เชิงปริมาณ	ธนาคารแห่งประเทศไทย [12]
X ₁₁	ดัชนีราคาผู้บริโภค หมวดยาสูบและเครื่องดื่มมีแอลกอฮอล์	-	เชิงปริมาณ	สำนักดัชนีเศรษฐกิจการค้า กระทรวงพาณิชย์ [13]
X ₁₂	จำนวนประชากรไทยที่มีอายุ 15 ปีขึ้นไป	คน	เชิงปริมาณ	สำนักบริหารการทะเบียน กระทรวงมหาดไทย [14]

2.2 ตัวแบบที่ศึกษา

ในส่วนนี้จะอธิบายเกี่ยวกับตัวแบบที่ถูกศึกษาภายใต้งานวิจัยนี้ ซึ่งประกอบไปด้วย ตัวแบบ MLR ตัวแบบ DTR และตัวแบบ RFR โดยมีรายละเอียดของแต่ละตัวแบบ ดังต่อไปนี้

2.2.1 ตัวแบบการถดถอยเชิงเส้นพหุคูณ

ตัวแบบการถดถอยเชิงเส้นเป็นแบบจำลองที่ใช้ในการอธิบายความสัมพันธ์ระหว่างตัวแปรตาม (Y) และตัวแปรทำนาย (X) ถ้าในตัวแบบมีตัวแปรทำนายมากกว่าหนึ่งตัว จะเรียกตัวแบบนี้ว่า ตัวแบบการถดถอยเชิงเส้นพหุคูณ โดยมีรูปแบบทั่วไปดังสมการที่ (1)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{12} X_{12} + \varepsilon \quad (1)$$

โดยที่	Y	คือ	ปริมาณยอดขายเบียร์ยี่ห้อหนึ่ง
	β_0	คือ	จุดตัดแกน Y
	X_i	คือ	ตัวแปรทำนายที่ i เมื่อ $i = 1, 2, \dots, 12$
	β_i	คือ	สัมประสิทธิ์การถดถอยของตัวแปรทำนาย X_i
และ	ε	คือ	ค่าความคลาดเคลื่อนสุ่ม ที่มีการแจกแจงปกติ ค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนคงที่

ซึ่งการประมาณค่าสัมประสิทธิ์ถดถอย β_0 และ β_i จะประมาณค่าด้วยวิธีกำลังสองน้อยที่สุด (Ordinary Least Squares: OLS)

โดยทั่วไปแล้ว ในการใช้ตัวแบบการถดถอยมุ่งหวังที่จะใช้ตัวทำนายให้น้อยที่สุด เพื่ออธิบายความผันแปรในตัวแปรตามที่เกิดขึ้น เทคนิคที่นิยมใช้มากที่สุดในการคัดเลือกตัวแปรทำนายที่เกี่ยวข้องในตัวแบบถดถอย คือ การคัดเลือกแบบขั้นตอน (Stepwise selection method) โดยใช้เกณฑ์ต่าง ๆ เพื่อการเลือกตัวแปรทำนายที่แตกต่างกัน ตัวอย่างเกณฑ์ที่ใช้ ได้แก่ เกณฑ์ข้อสนเทศของอาไคเคะ (Akaike Information Criterion: AIC) และเกณฑ์ข้อสนเทศของเบส์ (Bayesian Information Criteria: BIC) เป็นต้น

2.2.2 ตัวแบบการถดถอยต้นไม้ตัดสินใจ

เทคนิคต้นไม้ตัดสินใจ (Decision tree) เป็นเครื่องมือที่ใช้ในการประมวลผลข้อมูลและการทำนาย ด้วยการสร้างโครงสร้างต้นไม้จากข้อมูลที่มีอยู่ เทคนิคนี้ใช้การเรียนรู้แบบมีผู้สอน (Supervised learning) โดยมีสองประเภทหลัก คือ ต้นไม้ตัดสินใจแบบจำแนก (Classification trees) และต้นไม้ตัดสินใจแบบถดถอย (Regression trees) โดยต้นไม้ตัดสินใจแบบจำแนกใช้ในการทำนายผลลัพธ์ที่อยู่ในรูปแบบกลุ่มหรือประเภทต่าง ๆ ในขณะที่ต้นไม้ตัดสินใจแบบถดถอยใช้ในการทำนายผลลัพธ์ที่เป็นค่าต่อเนื่อง

การใช้งานต้นไม้ตัดสินใจเริ่มต้นที่โหนดราก (Root node) ซึ่งเป็นโหนดแรกที่ตรวจสอบคุณลักษณะ (Attribute: A) ของข้อมูล แล้วแบ่งข้อมูลออกเป็นกลุ่มย่อยโดยใช้เงื่อนไขที่เกี่ยวข้องกับคุณลักษณะนั้น ๆ กลุ่มย่อยจะแบ่งออกเป็นโหนดกิ่ง (Branch node) ซึ่งจะเป็นโหนดถัดไปในต้นไม้ กระบวนการนี้จะทำซ้ำไป

เรื่อย ๆ จนกว่าข้อมูลในกลุ่มย่อยไม่สามารถแบ่งได้แล้ว และโหนดสุดท้ายจะเรียกว่า โหนดใบ (Leaf node) ซึ่งในโหนดนี้จะเป็นผลลัพธ์ที่เกี่ยวข้องกับปัญหาของต้นไม้ตัดสินใจนั้น ๆ ในแต่ละขั้นตอนของการสร้างโหนด จะใช้เกณฑ์การคำนวณค่าเกนสารสนเทศ (Information Gain: IG) เพื่อเลือกคุณลักษณะหรือตัวแปรอิสระที่สามารถจัดหมวดหมู่ของข้อมูลได้ดีที่สุด โดยค่าเกนสารสนเทศสามารถคำนวณได้จากสมการที่ (2) [15]

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^a \frac{|S_i|}{|S|} \times \text{Entropy}(S_i) \quad (2)$$

โดยที่ $S = \{S_1, S_2, \dots, S_a\}$ คือ การแบ่งข้อมูล S ออกเป็นกลุ่มย่อยตามค่าของคุณลักษณะ A
 a คือ จำนวนคุณลักษณะ A
 $|S_i|$ คือ จำนวนข้อมูลในกลุ่มย่อย S_i
 และ $|S|$ คือ จำนวนข้อมูลทั้งหมดใน S

นอกจากนี้ ยังสามารถคำนวณหาเอนโทรปี (Entropy) ของข้อมูลใน S และกลุ่มย่อย S_i ซึ่งเป็นการวัดความไม่บริสุทธิ์ (Impurity) ของข้อมูล ได้จากสมการที่ (3)

$$\text{Entropy}(S) = \sum_{j=1}^g -\pi_j \log_2(\pi_j) \quad \text{และ} \quad \text{Entropy}(S_i) = \sum_{j=1}^g -\pi_{ij} \log_2(\pi_{ij}) \quad (3)$$

เมื่อ π_j คือ สัดส่วนของข้อมูล S ที่ผลลัพธ์มีค่า j และ π_{ij} คือ สัดส่วนของข้อมูลในกลุ่มย่อย S_i ที่ผลลัพธ์มีค่า j โดยที่เอนโทรปีจะมีค่าอยู่ระหว่าง 0 ถึง 1 ทั้งนี้ เอนโทรปีจะมีค่าเป็นศูนย์ เมื่อข้อมูลใน S หรือ S_i มีผลลัพธ์แบบเดียวกัน และจะมีค่ามากขึ้น เมื่อมีผลลัพธ์ที่แตกต่างกันมากขึ้น (หรือมีค่ามากขึ้น หากข้อมูลไม่บริสุทธิ์)

2.2.3 ตัวแบบการถดถอยแบบป่าสุ่ม

เทคนิคแบบป่าสุ่ม (Random forest) เป็นวิธีการเรียนรู้ของเครื่องใช้ในการจัดหมวดหมู่และทำนายข้อมูล โดยสร้างจากต้นไม้ตัดสินใจจำนวนมาก แล้วหาค่าเฉลี่ยจากผลลัพธ์ของต้นไม้ตัดสินใจทั้งหมดของตัวแปรตามเพื่อลดความแปรปรวนของการทำนาย ซึ่งจะทำให้ได้ผลลัพธ์ที่มีความแม่นยำมากกว่าการใช้ต้นไม้ตัดสินใจเพียงต้นเดียว วิธีแบบป่าสุ่มเป็นวิธีการสร้างตัวแบบที่มีความหลากหลาย และมีความสามารถในการทำนายที่แม่นยำ โดยที่กระบวนการสร้างตัวแบบถดถอยแบบป่าสุ่ม ประกอบไปด้วยขั้นตอนต่อไปนี้

- 1) เลือกตัวอย่างจากชุดข้อมูลที่มีอยู่ โดยการสุ่มเลือกตัวอย่างมาจำนวน n ตัว
- 2) เลือกตัวแปรทำนาย k ตัว จากทั้งหมด $p = 12$ ตัว แบบสุ่ม โดยที่ตัวแปรทำนายเหล่านี้จะนำมาใช้ในการสร้างต้นไม้ตัดสินใจ ซึ่งช่วยทำให้ต้นไม้แต่ละต้นไม่ในป่าสุ่มมีความหลากหลาย

3) ทำซ้ำขั้นตอนที่ 1 และ 2 จำนวน m ครั้ง เพื่อสร้างต้นไม้ตัดสินใจ m ต้น และรวมต้นไม้เหล่านี้เข้าด้วยกันเป็นป่าสุ่ม

4) เมื่อมีข้อมูลใหม่ที่ต้องการทำนาย ต้นไม้ตัดสินใจแต่ละต้นในป่าสุ่มจะตัดสินใจตามความหลากหลายของตัวเอง และรวมผลลัพธ์ที่ได้จากแต่ละต้นเพื่อการทำนายโดยใช้วิธีการคำนวณค่าเฉลี่ยของตัวแปรตาม

2.3 การเปรียบเทียบตัวแบบ

ในการเปรียบเทียบแม่นยำของตัวแบบจากการวิเคราะห์ MLR DTR และ RFR จะเลือกตัวแบบที่มีค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำที่สุดเป็นตัวเลือกหลักในการประมาณค่าและทำนายปริมาณยอดขายเปียร์รี่ห่อหนึ่ง โดยที่ RMSE สามารถคำนวณได้จากสมการที่ (4)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (4)$$

เมื่อกำหนดให้ y_t คือ ค่าที่เกิดขึ้นจริง ณ เวลาที่ t , $\hat{y}_t$ คือ ค่าพยากรณ์ ณ เวลาที่ t และ n คือ จำนวนข้อมูลทั้งหมด

3. ผลการวิจัย

ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคการวิเคราะห์ MLR, DTR และ RFR มีดังนี้

3.1 ผลการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

สำหรับการวิเคราะห์ MLR ผู้วิจัยได้ดำเนินการตรวจสอบข้อตกลงเบื้องต้นของชุดข้อมูลที่ใช้สำหรับสอน โดยการตรวจสอบความเป็นอิสระต่อกันของตัวแปรทำนาย การตรวจสอบการแจกแจงของส่วนเหลือ (Residual) ตรวจสอบความเท่ากันของความแปรปรวนของส่วนเหลือ การตรวจสอบความเป็นอิสระต่อกันของส่วนเหลือ และการตรวจสอบค่าผิดปกติ

เพื่อตรวจสอบความเป็นอิสระต่อกันของตัวแปรทำนาย ในการวิเคราะห์นี้ใช้ค่า Variance Inflation Factor (VIF) ในการตรวจสอบ โดยหากค่า VIF เกิน 10 แสดงว่าตัวแปรทำนายมีความสัมพันธ์กัน จากผลลัพธ์ในตารางที่ 2 พบว่า ค่า VIF ของตัวแปรทำนายทั้งหมดน้อยกว่า 10 แสดงว่าไม่มีความสัมพันธ์กันระหว่างตัวแปรทำนายแต่ละตัว นอกจากนี้ รูปที่ 1(ก) แสดงให้เห็นว่าส่วนเหลือกระจายอยู่รอบ ๆ ค่าศูนย์ และจากการทดสอบสมมติฐานด้วยสถิติทดสอบของ Breusch-Pagan ในตารางที่ 3 พบว่า ค่าสถิติทดสอบเท่ากับ 8.347 และ p-value เท่ากับ 0.758 แสดงว่าความแปรปรวนของส่วนเหลือคงที่ ที่ระดับนัยสำคัญ 0.05 นอกจากนี้ จากการตรวจสอบค่ามาตรฐานของส่วนเหลือในรูปที่ 1(ก) พบว่า มีค่าไม่เกินค่า |3| ซึ่งแสดงว่าข้อมูลไม่มีค่าผิดปกติ [16]

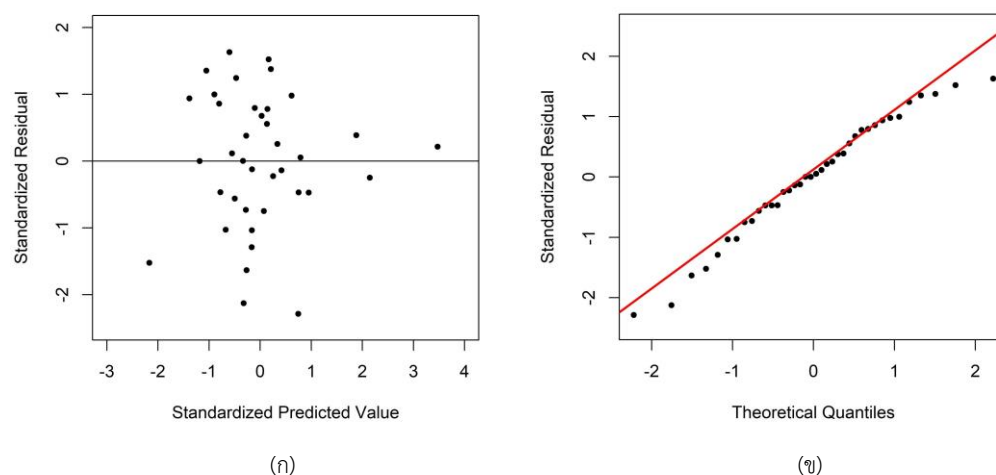
จากรูปที่ 1(ข) แสดงให้เห็นว่าค่ามาตรฐานของส่วนเหลือมีการกระจายตามแนวเส้นอ้างอิง ซึ่งสอดคล้องกับผลการทดสอบสมมติฐานด้วยสถิติทดสอบของ Shapiro-Wilk ในตารางที่ 3 (เนื่องจากมีจำนวนตัวอย่างน้อยกว่า 50) ซึ่งพบว่า ส่วนเหลือมีการแจกแจงปกติที่ระดับนัยสำคัญ 0.05 โดยมีค่าสถิติทดสอบเท่ากับ 0.972 นอกจากนี้ ในการตรวจสอบความเป็นอิสระต่อกันของส่วนเหลือด้วยการทดสอบสมมติฐาน พบว่า ส่วนเหลือไม่มีปัญหาเรื่องสหสัมพันธ์ในตัว (Autocorrelation) หรือส่วนเหลือเป็นอิสระต่อกันที่ระดับนัยสำคัญ 0.05 โดยมีค่าสถิติทดสอบของ Durbin-Watson เท่ากับ 1.826

ตารางที่ 2. ค่า VIF ของตัวแปรทำนายแต่ละตัว

ตัวแปรทำนาย											
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
2.605	7.069	3.621	1.576	2.578	1.572	1.541	1.859	2.295	4.456	2.060	4.420

ตารางที่ 3. ตรวจสอบการแจกแจงปกติ ความเท่ากันของความแปรปรวน และความเป็นอิสระกันของส่วนเหลือ

การทดสอบสมมติฐาน	สถิติทดสอบ	p-value
1. ตรวจสอบการแจกแจงปกติ	Shapiro-Wilk = 0.972	0.453
2. ตรวจสอบความเท่ากันของความแปรปรวน	Breusch-Pagan = 8.347	0.758
3. ตรวจสอบความเป็นอิสระกัน	Durbin-Watson = 1.826	0.564



รูปที่ 1. การวิเคราะห์ส่วนเหลือ (ก) แผนภาพการกระจายระหว่างค่ามาตรฐานของส่วนเหลือและค่ามาตรฐานของค่าทำนาย และ (ข) แผนภาพ Normal quantile-quantile (Normal q-q plot)

จากการตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์ MLR พบว่าเป็นไปตามข้อตกลงเบื้องต้นทุกประการ ดังนั้นผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลด้วยการวิเคราะห์ MLR โดยการนำเข้าตัวแปรทำนายทั้งหมด (วิธี Enter) และใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC และ BIC แสดงไว้ในตารางที่ 4

ตารางที่ 4. ค่าประมาณพารามิเตอร์และความคลาดเคลื่อนมาตรฐาน จากการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

ตัวแปรอิสระ	Enter		AIC		BIC	
	ค่าประมาณพารามิเตอร์	ความคลาดเคลื่อนมาตรฐาน	ค่าประมาณพารามิเตอร์	ความคลาดเคลื่อนมาตรฐาน	ค่าประมาณพารามิเตอร์	ความคลาดเคลื่อนมาตรฐาน
ค่าคงที่	-601,600,000.00 ***	122,600,000.00	-562,500,000.00 ***	84,740,000.00	-618,400,000.00 ***	80,310,000.00
X_1	0.06 ***	0.01	0.06 ***	0.01	0.06 ***	0.01
X_2	-3,301,000.00	2,125,000.00	-3,602,000.00 *	1,701,000.00	-	-
X_3	7,075,000.00 **	2,362,000.00	6,518,000.00 **	1,830,000.00	4,243,000.00 **	1,343,000.00
X_4	-240,500.00	356,700.00	-	-	-	-
X_5	-259,300.00	525,200.00	-	-	-	-
X_6	-347,700.00	436,000.00	-	-	-	-
X_7	-6,774,000.00	5,862,000.00	-8,498,000.00	5,163,000.00	-	-
X_8	-47,700.00	1,080,000.00	-	-	-	-
X_9	713,500.00	860,300.00	-	-	-	-
X_{10}	28,300.00 *	11,290.00	24,790.00 **	7,812.00	21,280.00 **	8,095.00
X_{11}	2,202,000.00 **	627,600.00	2,156,000.00 ***	517,500.00	2,306,000.00 ***	536,500.00
X_{12}	3.73 ***	0.71	3.71 ***	0.60	2.76 ***	0.37
R^2	0.9272		0.9223		0.9071	
Adjusted R^2	0.8922		0.9041		0.8925	
F	26.52 ***		50.86 ***		62.46 ***	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

ผลลัพธ์ในตารางที่ 4 พบว่าตัวแบบที่มีค่าสัมประสิทธิ์แสดงการตัดสินใจที่ปรับค่าแล้ว (Adjusted R^2) มากที่สุด คือ ตัวแบบที่ใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC ซึ่งมีค่า Adjusted R^2 เท่ากับ 0.9041 หมายความว่า ตัวแปรทำนายทั้งหมดในตัวแบบนี้สามารถอธิบายความผันแปรทั้งหมดของปริมาณยอดขายเบียร์ได้ร้อยละ 90.41 รองลงมา คือ ตัวแบบที่ใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ BIC และตัวแบบที่นำเข้าตัวแปรทำนายทั้งหมด ซึ่งมีค่า Adjusted R^2 เท่ากับ 0.8925 และ 0.8922 ตามลำดับ จากวิธีการคัดเลือกตัวแปรทำนายทั้ง 3 วิธี สามารถเขียนสมการถดถอยได้ ดังนี้

1) นำเข้าตัวแปรทำนายทั้งหมด

$$Y = -601,600,000.00 + (0.06)(X_1) - (3,301,000.00)(X_2) + (7,075,000.00)(X_3) \\ - (240,500.00)(X_4) - (259,300.00)(X_5) - (347,700.00)(X_6) \\ - (6,774,000.00)(X_7) - (47,700.00)(X_8) + (713,500.00)(X_9) \\ + (28,300.00)(X_{10}) + (2,202,000.00)(X_{11}) + (3.73)(X_{12})$$

2) คัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC

$$Y = -562,500,000.00 + (0.06)(X_1) - (3,602,000.00)(X_2) + (6,518,000.00)(X_3) \\ - (8,498,000.00)(X_7) + (24,790.00)(X_{10}) + (2,156,000.00)(X_{11}) + (3.71)(X_{12})$$

3) คัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ BIC

$$Y = -618,400,000.00 + (0.06)(X_1) + (4,243,000.00)(X_3) + (21,280.00)(X_{10}) \\ + (2,306,000.00)(X_{11}) + (2.76)(X_{12})$$

โดยที่สมการถดถอยข้างต้น ไม่ได้ตัดตัวแปรที่ไม่มีนัยสำคัญออกจากสมการ เนื่องจากจะทำให้ค่าประมาณของสัมประสิทธิ์ R^2 และ Adjusted R^2 เปลี่ยนแปลงไป

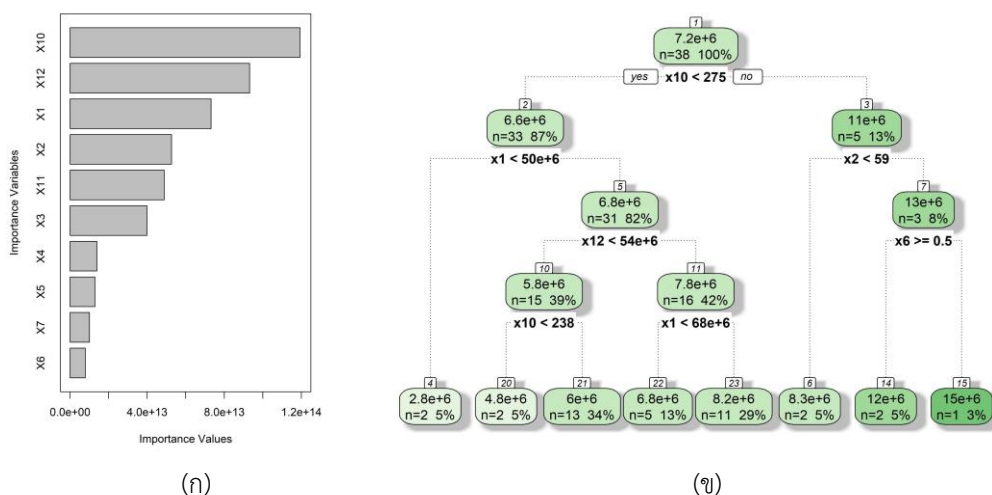
ทั้งนี้ จากการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC มีตัวแปรทำนายที่ถูกคัดเลือกเข้ามาในตัวแบบทั้งหมด 7 ตัวแปร และมีเพียง 6 ตัวแปร ที่มีความสัมพันธ์กับปริมาณยอดขายเบียร์อย่างมีนัยสำคัญทางสถิติ ได้แก่ ปริมาณยอดขายเบียร์คู่แข่งที่ 1 (X_1) ราคาขายทอดที่ 3 ของคู่แข่งที่ 1 (X_2) ราคาขายทอดที่ 3 ของคู่แข่งที่ 2 (X_3) ดัชนีค่าปลีก (X_{10}) ดัชนีราคาผู้บริโภค หมวดยาสูบและเครื่องดื่มมีแอลกอฮอล์ (X_{11}) และจำนวนประชากรไทยที่มีอายุ 15 ปีขึ้นไป (X_{12})

ในขณะที่การคัดเลือกตัวแปรทำนายด้วยเกณฑ์ BIC จะมีตัวแปรทำนายที่ถูกคัดเลือกเข้ามาในตัวแบบจำนวน 5 ตัวแปร และตัวแปรที่ถูกคัดเลือกเข้ามานี้มีความสัมพันธ์กับปริมาณยอดขายเบียร์อย่างมีนัยสำคัญทางสถิติทุกตัว ซึ่งประกอบไปด้วย ปริมาณยอดขายเบียร์คู่แข่งที่ 1 (X_1) ราคาขายทอดที่ 3 ของคู่แข่งที่ 2 (X_3) ดัชนีค่าปลีก (X_{10}) ดัชนีราคาผู้บริโภค หมวดยาสูบและเครื่องดื่มมีแอลกอฮอล์ (X_{11}) และจำนวนประชากรไทยที่มีอายุ 15 ปี ขึ้นไป (X_{12})

3.2 ผลการวิเคราะห์การถดถอยต้นไม้ตัดสินใจ

ในการวิเคราะห์ข้อมูลด้วยต้นไม้ตัดสินใจ ผู้วิจัยได้เลือกใช้เทคนิคต้นไม้ตัดสินใจแบบจำแนกและแบบถดถอย (Classification and Regression Tree: CART) โดยใช้ชุดคำสั่ง rpart ในโปรแกรมภาษา R มาช่วยในการสร้างตัวแบบ

เพื่อให้ได้แบบตัวแบบที่มีประสิทธิภาพที่ดี ผู้วิจัยได้ทำการปรับแต่งไฮเปอร์พารามิเตอร์ (Hyperparameter tuning) ด้วยวิธีการค้นหาแบบกริด (Grid search) โดยกำหนดให้ช่วงของจำนวนข้อมูลต่ำสุดที่จำเป็นต้องมีก่อนที่จะทำการสร้างโหนด (minsplit) อยู่ระหว่าง 2 ถึง 20 และช่วงของจำนวนโหนดภายในสูงสุดระหว่างโหนดรากและโหนดปลาย (maxdepth) อยู่ระหว่าง 2 ถึง 15 จากการค้นหาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสม พบว่า ไฮเปอร์พารามิเตอร์ที่ทำให้ค่าความคลาดเคลื่อนต่ำที่สุด คือ minsplit = 3 และ maxdepth = 10 ซึ่งทำให้ได้ผลลัพธ์ดังแสดงในรูปที่ 2



รูปที่ 2. ตัวแปรทำนายที่สำคัญและต้นไม้ตัดสินใจ (ก) แผนภูมิแท่งแสดงค่าความสำคัญของตัวแปรทำนายที่ได้จากการวิเคราะห์การถดถอยต้นไม้ตัดสินใจ และ (ข) แผนภาพต้นไม้ตัดสินใจ

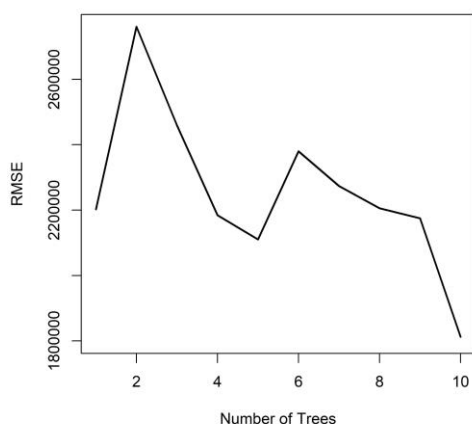
จากรูปที่ 2(ก) พบว่า ตัวแปรทำนายที่มีความสำคัญสำหรับการวิเคราะห์ข้อมูลด้วย DTR มีทั้งหมด 10 ตัวแปร โดยตัวแปรที่มีความสำคัญมากที่สุด คือ ดัชนีค่าปลิก (X_{10}) และจากตัวแปรทั้งหมด 10 ตัวแปรที่มีความสำคัญดังกล่าว จะมีตัวแปรทำนายที่ถูกนำมาใช้ในการสร้างต้นไม้ตัดสินใจเพียง 5 ตัวแปร ซึ่งประกอบไปด้วย ปริมาณยอดขายสินค้าเบียร์คู่แข่งที่ 1 (X_1) ราคาขายทอดที่ 3 ของคู่แข่งที่ 1 (X_2) การทำกลยุทธ์ที่เกี่ยวกับการให้เครดิตพิเศษ (X_6) ดัชนีค่าปลิก (X_{10}) และจำนวนประชากรไทยที่มีอายุ 15 ปีขึ้นไป (X_{12}) ดังแสดงในรูปที่ 2(ข)

3.3 ผลการวิเคราะห์การถดถอยแบบป่าสุ่ม

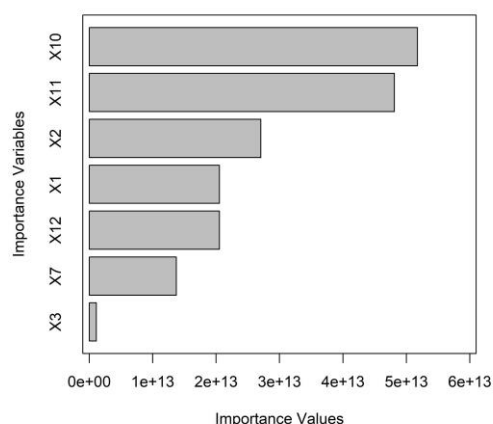
ในการวิเคราะห์ RFR ผู้วิจัยใช้ชุดคำสั่ง randomForest ในโปรแกรม R เพื่อสร้างแบบตัวแบบ ในทำนองเดียวกับการวิเคราะห์ DTR เพื่อให้ได้แบบตัวแบบที่มีประสิทธิภาพที่ดี จึงได้ทำการปรับแต่งไฮเปอร์พารามิเตอร์ด้วยวิธีการค้นหาแบบกริด โดยกำหนดให้ช่วงของจำนวนต้นไม้ (ntree) อยู่ระหว่าง 5 ถึง 500 ต้น โดยให้เพิ่มขึ้นทีละ 5 ต้น ช่วงของจำนวนตัวแปรที่สุ่มเลือกในแต่ละการแบ่ง (mtry) อยู่ระหว่าง 1 ถึง 12 ช่วงของจำนวนตัวอย่างขั้นต่ำในโหนดปลาย (nodesize) มีค่าอยู่ระหว่าง 3 ถึง 10 และช่วงของจำนวนสูงสุดของโหนดปลาย (maxnodes) มีค่าอยู่ระหว่าง 3 ถึง 10 และจากการค้นหา พบว่า ไฮเปอร์

พารามิเตอร์ที่ทำให้ค่า out-of-bag (OOB) RMSE ต่ำที่สุด คือ $n_{tree} = 10$, $m_{try} = 8$, $nodesize = 3$ และ $maxnodes = 4$ ซึ่งทำให้ได้ผลลัพธ์ดังแสดงในรูปที่ 3

จากรูปที่ 3(ก) พบว่า จำนวนต้นไม้ที่ให้ค่า RMSE ต่ำที่สุด คือ จำนวน 10 ต้น โดยมีค่าเท่ากับ 1,812,497 นอกจากนี้ ตัวแปรทำนายที่มีความสำคัญในตัวแบบ RFR นั้น ประกอบไปด้วย ปริมาณยอดขาย เปียรีคู่แข่งที่ 1 (X_1) ราคาขายทอดที่ 3 ของคู่แข่งที่ 1 (X_2) ราคาขายทอดที่ 3 ของคู่แข่งที่ 2 (X_3) ความพร้อมของสต็อกสินค้า (X_7) ดัชนีค่าปลิก (X_{10}) ดัชนีราคาผู้บริโภค หมวดยาสูบและเครื่องดื่มมีแอลกอฮอล์ (X_{11}) และจำนวนประชากรไทยที่มีอายุ 15 ปีขึ้นไป (X_{12}) ดังแสดงในรูปที่ 3(ข)



(ก)



(ข)

รูปที่ 3. จำนวนต้นไม้ที่เหมาะสมและตัวแปรทำนายที่สำคัญ (ก) แผนภาพแสดงระหว่างค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองและจำนวนต้นไม้ และ (ข) แผนภูมิแท่งแสดงค่าความสำคัญของตัวแปรทำนายที่ได้จากการวิเคราะห์การถดถอยแบบป่าสุ่ม

3.4 ผลการเปรียบเทียบตัวแบบการพยากรณ์

เพื่อประเมินประสิทธิภาพของตัวแบบที่ได้จากการวิเคราะห์ MLR, DTR และ RFR ในการทำนายข้อมูลในชุดทดสอบ และค่า RMSE ถูกนำมาใช้ในการวัดความแม่นยำของการทำนายข้อมูล ซึ่งสามารถสรุปผลการเปรียบเทียบประสิทธิภาพตัวแบบที่ใช้ในการทำนายได้ ดังนี้

ตารางที่ 5. ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของแต่ละวิธีที่ใช้ในการทำนาย

วิธีที่ใช้ในการทำนาย	RMSE
1. การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (MLR)	
1.1 นำเข้าตัวแปรอิสระทั้งหมด	331,281.70
1.2 คัดเลือกตัวแปรทำนายแบบขั้นตอน ด้วยเกณฑ์ AIC	225,698.50
1.3 คัดเลือกตัวแปรทำนายแบบขั้นตอน ด้วยเกณฑ์ BIC	428,344.90
2. การวิเคราะห์การถดถอยต้นไม้ตัดสินใจ (DTR)	1,727,588.00
3. การวิเคราะห์การถดถอยแบบป่าสุ่ม (RFR)	514,340.00

จากตารางที่ 5 พบว่า วิธีการวิเคราะห์ MLR ที่ใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC มีค่า RMSE ต่ำที่สุด นั่นคือ $RMSE = 225,698.50$ แสดงว่า วิธีการวิเคราะห์ MLR ที่ใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC มีประสิทธิภาพในการทำนายได้มากกว่าวิธีการอื่น ๆ ที่ทำการเปรียบเทียบ รองลงมา คือ การวิเคราะห์ MLR ที่นำเข้าตัวแปรทำนายทั้งหมด การวิเคราะห์ MLR ที่ใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ BIC การวิเคราะห์ RFR และการวิเคราะห์ DTR ซึ่งมีค่า RMSE เท่ากับ 331,281.70, 428,344.90, 514,340.00 และ 1,727,588.00 ตามลำดับ

จากผลลัพธ์ข้างต้น ผู้วิจัยจึงตัดสินใจเลือกใช้ตัวแบบ MLR ที่ใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC มาใช้ในการทำนายปริมาณยอดขายเบียร์ยี่ห้อนี้ และเมื่อนำข้อมูลทั้งหมดมาใช้ในการสร้างตัวแบบ จะได้ผลลัพธ์ดังตารางที่ 6 ซึ่งพบว่า ตัวแปรทำนายในตัวแบบสามารถอธิบายความผันแปรทั้งหมดของปริมาณยอดขายเบียร์ได้ร้อยละ 91.20 ($Adjusted R^2 = 0.9120$) โดยจะได้สมการถดถอยจากตัวแปรทำนายทั้ง 7 ตัวแปร ดังสมการที่ (5)

$$Y = -576,300,000.00 + (0.06)(X_1) - (3,456,000.00)(X_2) + (6,498,000.00)(X_3) - (7,510,000.00)(X_7) + (24,650.00)(X_{10}) + (2,190,000.00)(X_{11}) + (3.75)(X_{12}) \quad (5)$$

ตารางที่ 6. ค่าประมาณพารามิเตอร์ ค่ามาตรฐานของค่าประมาณพารามิเตอร์ และค่าความคลาดเคลื่อนมาตรฐาน จากการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ ที่ใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC

ตัวแปรอิสระ		ค่าประมาณพารามิเตอร์	ค่ามาตรฐานของค่าประมาณพารามิเตอร์	ค่าความคลาดเคลื่อนมาตรฐาน	p-value
ค่าคงที่		-576,300,000.00	-	71,430,000.00	< 0.001***
X ₁	ปริมาณยอดขายเบียร์คู่แข่งที่ 1	0.06	0.50	0.01	< 0.001***
X ₂	ราคาขายทอดที่ 3 ของคู่แข่งที่ 1	-3,456,000.00	-0.25	1,445,000.00	0.022*
X ₃	ราคาขายทอดที่ 3 ของคู่แข่งที่ 2	6,498,000.00	0.28	1,577,000.00	< 0.001***
X ₇	ความพร้อมของสต็อกสินค้า	-7,510,000.00	-0.08	4,522,000.00	0.105
X ₁₀	ดัชนีค่าปลีก	24,650.00	0.24	6,982.00	0.001**
X ₁₁	ดัชนีราคาผู้บริโภค หมวดยาสูบและเครื่องดื่มแอลกอฮอล์	2,190,000.00	0.26	455,100.00	< 0.001***
X ₁₂	จำนวนประชากรไทยที่มีอายุ 15 ปีขึ้นไป	3.75	0.61	0.52	< 0.001***
R ²		0.9251			
Adjusted R ²		0.9120			
F		70.62			< 0.001***

* p < 0.05, ** p < 0.01, *** p < 0.001

นอกจากนี้ยังพบว่า มีตัวแปรทำนายที่มีอิทธิพลต่อการทำนายปริมาณยอดขายเบียร์อย่างมีนัยสำคัญทางสถิติเพียง 6 ตัวแปร และเมื่อเรียงลำดับความสำคัญของตัวแปรทำนายที่มีอิทธิพลต่อปริมาณยอดขายเบียร์ยี่ห้อหนึ่งจากมากไปน้อย โดยการพิจารณาค่ามาตรฐาน (Standardized) ของค่าประมาณพารามิเตอร์พบว่า จำนวนประชากรไทยที่มีอายุ 15 ปีขึ้นไป (X_{12}) มีอิทธิพลต่อปริมาณยอดขายเบียร์ยี่ห้อหนึ่งมากที่สุด เนื่องจากค่าสัมบูรณ์ของค่ามาตรฐานของสัมประสิทธิ์การถดถอยเท่ากับ 0.61 รองลงมา คือ ปริมาณยอดขายเบียร์คู่แข่งที่ 1 (X_1) ราคาขายทอดที่ 3 ของคู่แข่งที่ 2 (X_3) ดัชนีราคาผู้บริโภค หมวดยาสูบและเครื่องดื่มมีแอลกอฮอล์ (X_{11}) ราคาขายทอดที่ 3 ของคู่แข่งที่ 1 (X_2) และดัชนีค่าปลีก (X_{10}) ตามลำดับ

4. สรุปและอภิปรายผลการวิจัย

ในการศึกษาการเปรียบเทียบตัวแบบในการทำนายปริมาณยอดขายเบียร์ยี่ห้อหนึ่ง ผู้วิจัยได้ทำการสร้างตัวแบบที่ใช้ในการทำนายด้วยวิธีการถดถอยเชิงเส้นพหุคูณ (MLR) วิธีการถดถอยต้นไม้ตัดสินใจ (DTR) และวิธีการถดถอยแบบป่าสุ่ม (RFR) และประเมินประสิทธิภาพของตัวแบบด้วยค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ผลการศึกษาพบว่า ตัวแบบที่ได้จากการวิเคราะห์ MLR ที่ใช้วิธีการคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ข้อสนเทศของอาไคเคะ (AIC) มีประสิทธิภาพในการทำนายปริมาณยอดขายเบียร์ยี่ห้อหนึ่งมากที่สุด โดยมีค่า RMSE เท่ากับ 225,698.50 นอกจากนี้ เมื่อนำข้อมูลทั้งหมดมาใช้ในการวิเคราะห์ตัวแปรทำนายในตัวแบบ MLR ที่ได้จากวิธีการคัดเลือกตัวแปรดังกล่าวสามารถอธิบายความผันแปรทั้งหมดของปริมาณยอดขายเบียร์ได้ร้อยละ 91.20 และมีตัวแปรทำนายที่มีอิทธิพลต่อการทำนายปริมาณยอดขายเบียร์ยี่ห้อหนึ่งจำนวน 6 ตัวแปร ได้แก่ 1) ปริมาณยอดขายเบียร์คู่แข่งที่ 1 2) ราคาขายทอดที่ 3 ของคู่แข่งที่ 1 3) ราคาขายทอดที่ 3 ของคู่แข่งที่ 2 4) ดัชนีค่าปลีก 5) ดัชนีราคาผู้บริโภค หมวดยาสูบและเครื่องดื่มมีแอลกอฮอล์ และ 6) จำนวนประชากรไทยที่มีอายุ 15 ปีขึ้นไป โดยผลการวิจัยครั้งนี้ พบว่า ราคาขายทอดที่ 3 ของคู่แข่งที่ 1 และคู่แข่งที่ 2 นั้นสอดคล้องกับราคาตลาดเบียร์ในงานวิจัยของ อธิวัฒน์ ลาภหลาย และคณะ [8] ซึ่งมีผลต่อปริมาณการบริโภคเบียร์ในประเทศไทยอย่างมีนัยสำคัญทางสถิติ

จากการศึกษาในครั้งนี้ ตัวแบบ DTR และตัวแบบ RFR ยังไม่เหมาะสมสำหรับนำมาใช้ในการทำนายปริมาณยอดขายเบียร์ของผู้ประกอบการเบียร์ยี่ห้อหนึ่ง ซึ่งต่างจากการศึกษาตัวแบบของ พรทิwa วิเศษสุวรรณธร และเอกสิทธิ์ พัทธวงศ์ศักดิ์ [3] สิทธิชัย นาคพิทักษ์ และวิภาดา ผาพันธ์ [4] และ Čeh และคณะ [5] ที่ตัวแบบ RFR มีประสิทธิภาพในการทำนายดีกว่า ทั้งนี้ ตัวแบบที่เหมาะสมที่สุดนั้นขึ้นอยู่กับข้อมูลที่นำมาใช้ในการสร้างตัวแบบ ซึ่งในงานวิจัยของ Krkač และคณะ [6] ตัวแบบ MLR นั้น ดีกว่าตัวแบบ RFR เพียงเล็กน้อย สำหรับในงานวิจัยนี้ ตัวแบบ MLR ที่ใช้การคัดเลือกตัวแปรทำนายแบบขั้นตอนด้วยเกณฑ์ AIC เป็นตัวแบบที่เหมาะสมกับข้อมูลชุดนี้มากที่สุด ซึ่งทำให้ประสิทธิภาพในการทำนายปริมาณยอดขายเบียร์มีความแม่นยำมากกว่าตัวแบบอื่น ถึงแม้ว่าข้อมูลที่นำมาใช้ในการวิเคราะห์ในครั้งนี้ จะมีคาบเกี่ยวจากการระบาดของโรคโควิด-19 อย่างไรก็ตาม ตัวแบบนี้สามารถนำไปใช้เป็นแนวทางในการนำไปประยุกต์ใช้ทำนายปริมาณยอดขายเบียร์ของผู้ประกอบการในอนาคตได้

เพื่อเพิ่มความแม่นยำในการทำนายปริมาณยอดขายเบียร์ในอนาคต การปรับปรุงตัวแบบยังคงเป็นปัจจัยสำคัญที่จะช่วยให้ตัวแบบมีความทันสมัย และมีความแม่นยำในการทำนายมากยิ่งขึ้น ดังนั้นในอนาคต

การศึกษาวิธีการทำนายที่มีการคัดเลือกตัวแปรด้วยวิธีแบบปรับโทษ (Penalty) เช่น Least Absolute Shrinkage and Selection Operator (LASSO) adaptive LASSO หรือ Elastic Net เป็นต้น หรือศึกษาวิธีการทำนายโดยใช้วิธีการเรียนรู้เชิงลึก (Deep learning) อาจจะทำให้ได้ตัวแบบที่เหมาะสมกับข้อมูลมากยิ่งขึ้น

เอกสารอ้างอิง (References)

- [1] ศูนย์วิจัยกรุงศรี. 2565. แนวโน้มธุรกิจ/อุตสาหกรรม ปี 2565-67 อุตสาหกรรมเครื่องดื่ม. แหล่งข้อมูล : <https://www.krungsri.com/th/research/industry/industry-outlook/food-beverage/beverage/io/io-beverage-2022>. ค้นเมื่อวันที่ 23 มกราคม 2567.
- [2] สถาบันอาหาร. 2566. รายงานตลาดอาหารในประเทศไทย ตลาดเครื่องดื่มแอลกอฮอล์. แหล่งข้อมูล : https://fic.nfi.or.th/upload/market_overview/pdf379.pdf. ค้นเมื่อวันที่ 23 มกราคม 2567.
- [3] พรทิwa วิศิษฐ์สรอรรถ และเอกสิทธิ์ พัทรวงศ์ศักดิ์. 2564. ระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิคทางสถิติและการเรียนรู้ของเครื่อง. *วารสารบัณฑิตวิทยาลัย มหาวิทยาลัยธุรกิจบัณฑิตย์*, 10(2), 133-147. [Porntiwa Wisitsora-At and Eakasit Pacharawongsakda. 2021. The time series analysis system through statistic and machine learning techniques. *Dhurakij Pundit University Journal of Graduate Studies*, 10(2), 133-147. (in Thai)]
- [4] สิทธิชัย นาคพิทักษ์ และวิกานดา ผาพันธ์. 2565. การเปรียบเทียบรายได้รายวันของโรงพยาบาลแห่งหนึ่ง 1 โดยใช้การเรียนรู้ของเครื่องและตัวแบบอาร์มา. *วารสารวิทยาศาสตร์ เทคโนโลยีและนวัตกรรม มหาวิทยาลัยกาฬสินธุ์*, 1(1), 20-42. [Sittichai Nakphithak and Wikanda Pha-phan. 2022. A comparison of the daily income of a hospital by using machine learning methods and the ARIMA model. *Kalasin University Journal of Science Technology and Innovation*, 1(1), 20-42. (in Thai)]
- [5] Čeh, M., Milan, K., Anka, L. and Branislav, B. 2018. Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS International Journal of Geo-Information*, 7(5), 1-16, <https://doi.org/10.3390/ijgi7050168>.
- [6] Krkač, M., Bernat Gazibara, S., Arbanas, Ž., Sećanj, M. and Mihalić Arbanas, S. 2020. A comparative study of random forests and multiple linear regression in the prediction of landslide velocity. *Landslides*, 17, 2515-2531, <https://doi.org/10.1007/s10346-020-01476-6>.
- [7] เบญจวรรณ ยุทโท. 2561. วิเคราะห์ความต้องการปริมาณสินค้าแอลกอฮอล์: กรณีศึกษาเบียร์ ABC ขนาดใหญ่. วิทยานิพนธ์ปริญญาเศรษฐศาสตรมหาบัณฑิต, สาขาวิชาเศรษฐศาสตร์ธุรกิจ, สถาบันบัณฑิตพัฒนบริหารศาสตร์. [Benjawan Yutto. 2018. Demand analysis for alcohol

- products: a case study of ABC big bottle beer. Master Thesis, National Institute of Development Administration, Bangkok. (in Thai)]
- [8] อธิวัฒน์ ลาภหลาย, อภิญา วนเศรษฐ และวสุ สุวรรณวิหค. 2563. ปัจจัยที่มีผลต่อการบริโภคเบียร์ในประเทศไทยและนโยบายภาครัฐ. *วารสารเกษตรศาสตร์ธุรกิจประยุกต์*, 14(21), 1-14. [Atiwat Laplai, Apinya Wanaset and Vasu Suvanvihok. 2020. Factors affecting the consumption of beer in Thailand and government policies. *Kasetsart Applied Business Journal*, 14(21), 1-14. (in Thai)]
- [9] ประเมษฐ์ ฐันวานนท์, ชัยกร ยิ่งเสรี, วรพล พงษ์เพชร และธนภัทร ชังคะจิตร. 2560. การประยุกต์ใช้โมเดลการเรียนรู้แบบรวมกลุ่มเพื่อพยากรณ์แนวโน้มของราคาหลักทรัพย์ในตลาดหลักทรัพย์แห่งประเทศไทย. *วารสารวิทยาการและเทคโนโลยีสารสนเทศ*, 7(1), 12-21. [Poramet Tunwanont, Chaiyakorn Yingsaeree, Woraphol Phongphet and Thanapat Kangkachit. 2017. Predict stock price trends in Stock Exchange of Thailand using Ensemble Model. *Journal of Information Science and Technology*, 7(1), 12-21. (in Thai)]
- [10] ศูนย์วิจัยปัญหาสุรา. 2565. แบบแผนและแนวโน้มพฤติกรรมกรรมการดื่มสุราของประชากรไทย. สงขลา: ศูนย์วิจัยปัญหาสุรา, สาขาวิชาระบาดวิทยา, คณะแพทยศาสตร์ มหาวิทยาลัยสงขลานครินทร์.
- [11] กนิษฐา ไทยกล้า, นิษฐา หุ่นเกษม และวิทย์ วิชัยดิษฐ์. 2563. ผลกระทบและการแก้ไขปัญหาของธุรกิจเครื่องดื่มแอลกอฮอล์ในสังคมไทยในสภาวะการระบาดโควิด-19. ศูนย์วิจัยปัญหาสุรา, แหล่งข้อมูล : <https://cas.or.th/?p=8240>. ค้นเมื่อวันที่ 12 เมษายน 2566.
- [12] ธนาคารแห่งประเทศไทย. 2566. ดัชนีค้าปลีก. แหล่งข้อมูล : https://app.bot.or.th/BTWS_STAT/statistics/BOTWEBSTAT.aspx?reportID=830&language=TH. ค้นเมื่อวันที่ 12 เมษายน 2566.
- [13] สำนักดัชนีเศรษฐกิจการค้า กระทรวงพาณิชย์. 2566. ดัชนีราคาผู้บริโภค. แหล่งข้อมูล : http://www.price.moc.go.th/price/cpi/index_new.asp. ค้นเมื่อวันที่ 12 เมษายน 2566.
- [14] สำนักบริหารการทะเบียน กระทรวงมหาดไทย. 2566. สถิติประชากรทางการทะเบียนราษฎร (รายเดือน). แหล่งข้อมูล : <https://stat.bora.dopa.go.th/stat/statnew/statMONTH/statmonth/#/view>. ค้นเมื่อวันที่ 12 เมษายน 2566.
- [15] Dang, T. K., and Nguyen, H. H. X. 2022. A hybrid approach using decision tree and multiple linear regression for predicting students' performance based on learning progress and behavior. *SN Computer Science*, 3(5), 393, <https://doi.org/10.1007/s42979-022-01251-5>.
- [16] ฉัตรศิริ ปิยะพิมลสิทธิ์. 2547. ค่าผิดปกติ. *วารสาร มจร.วิชาการ*, 8(15), 106-110. [Chatsiri Piyapimonsit. 2004. Outlier. *HCU Journal*, 8(15), 106-110. (in Thai)]