



มหาวิทยาลัยหาดใหญ่
HATYAI UNIVERSITY

การพยากรณ์ราคาน้ำมันแก๊สโซฮอล์ 95 ในอำเภอหาดใหญ่ จังหวัดสงขลา โดยใช้วิธีวิเคราะห์อนุกรมเวลา

Forecasting the Price of Gasohol 95 in Hatyai District, Songkhla Province Using Time Series Analysis.

พงศ์พัฒน์ ฉายศิริพันธ์^{1*}, เดชานูชิต กตัญญูทวีทิพย์², ณรงค์ฤทธิ์ สุกนธสิงห์³ และอรรณพ กางกั้น⁴
Pongpat Chaisiripan^{1*}, Dechanucit Katanyutaveetip², Narongrit Sukonthasing³, and Unnop Kangkan⁴

^{1,2,3,4} อาจารย์สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยสยาม

^{1,2,3,4} Lecturer in Information Technology Faculty of Information Technology Siam University

*Corresponding author, E-mail: pongpat.cha@siam.edu

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างและเปรียบเทียบแบบจำลองการพยากรณ์ราคาน้ำมันแก๊สโซฮอล์ 95 ที่จำหน่ายในอำเภอหาดใหญ่ จังหวัดสงขลา โดยใช้วิธีวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมืองข้อมูล 3 วิธี ได้แก่ วิธีการถดถอยเชิงเส้น แบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ข้อมูลที่นำมาใช้เป็นชุดข้อมูลราคาน้ำมันแก๊สโซฮอล์ 95 ที่จำหน่ายในอำเภอหาดใหญ่ จังหวัดสงขลา ตั้งแต่เดือนมกราคม พ.ศ.2566 ถึงมกราคม พ.ศ.2567 ผลการวิจัยพบว่า ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (SMOreg) มีความเหมาะสมกับชุดข้อมูลราคาน้ำมันแก๊สโซฮอล์ 95 ด้วยชุดข้อมูลย้อนหลัง 25 วัน (Lag 25) ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 0.153%

คำสำคัญ: พยากรณ์, น้ำมัน, อนุกรมเวลา

Abstract

The objective of this research is to create and compare a price prediction model for Gasohol 95 sold in Hat Yai District, Songkhla Province, using time series analysis using three data mining techniques: linear regression method. Multi-layer perceptron neural network model. and support vector machines for regression. The data used as the oil price dataset includes the price of gasohol 95 sold in Hat Yai District, Songkhla Province from January 2023 to January 2024. The research results found that the vector machine supports For the regression (SMOreg), it is suitable for the gasohol price data set 95 with the historical data set of 25 days (Lag 25). The mean relative error is equal to 0.153%.

Keywords: prediction, oil, time series



บทนำ

น้ำมันถือเป็นแหล่งพลังงานที่มีความสำคัญต่อเศรษฐกิจโลก ซึ่งในปัจจุบันนี้พลังงานจากน้ำมันเป็นส่วนสำคัญต่อการดำรงชีวิต ทั้งการคมนาคมและการขนส่งสินค้าทั้งหมด จากปริมาณการใช้น้ำมันกลุ่มเบนซิน เดือนมกราคม - สิงหาคม 2566 เฉลี่ยอยู่ที่ 31.91 ล้านลิตร/วัน เพิ่มขึ้นจากช่วงเดียวกันของปี 2565 ร้อยละ 5.5 การใช้น้ำมันแก๊สโซฮอล์ 95 เพิ่มขึ้นมาอยู่ที่ 18.10 ล้านลิตร/วัน และแก๊สโซฮอล์ อี 20 เพิ่มขึ้นมาอยู่ที่ 5.94 ล้านลิตร/วัน ขณะที่การใช้แก๊สโซฮอล์ 91 แก๊สโซฮอล์ อี85 และเบนซิน ลดลงมาอยู่ที่ 6.83 ล้านลิตร/วัน 0.19 ล้านลิตร/วัน และ 0.47 ล้านลิตร/วัน ตามลำดับ ทั้งนี้ส่วนต่างราคาแก๊สโซฮอล์ 95 ใกล้เคียงกับแก๊สโซฮอล์ 91 ประกอบกับแก๊สโซฮอล์ อี 20 มีราคาต่ำกว่าแก๊สโซฮอล์ อี 85 (กองบริหารธุรกิจและการสำรองน้ำมันเชื้อเพลิง กรมธุรกิจพลังงาน, 2566) จากปริมาณการใช้น้ำมันกลุ่มเบนซินที่เพิ่มขึ้นเมื่อทำการพิจารณาราคาน้ำมันแก๊สโซฮอล์ในปัจจุบันยังคงมีราคาสูง โดยเฉพาะแก๊สโซฮอล์ 95 ซึ่งราคาในปี 2564 ที่บริษัท ปตท. จำกัด (มหาชน) รายงานไว้มีค่าเพียง 31.47 บาท/ลิตร แต่ราคาในปัจจุบัน (ณ วันที่ 31 มกราคม 2567) มีค่าสูงถึง 37.87 บาท/ลิตร ซึ่งมีค่าสูงกว่าในอดีต

จากการทบทวนวรรณกรรม พบว่า งานวิจัยที่เกี่ยวกับการพยากรณ์ (สุภาพร บรรดาศักดิ์, จุฑามาศ แป้นทอง และนุทยา กัลป์ลา, 2559) ได้พัฒนาแบบจำลองการพยากรณ์การใช้ลูกกอล์ฟในสนามฝึกซ้อมโดยใช้ตัวแปรอนุกรมเวลา โดยโปรแกรม Weka ด้วยเทคนิคเหมือนข้อมูล 2 เทคนิค ได้แก่ 1) การถดถอยเชิงเส้น 2) การถดถอยเวกเตอร์ โดยใช้ข้อมูลในปี 2557 - 2559 จากผลการทดลอง งานวิจัยนี้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยเทคนิค 2 กับชุดข้อมูลฝึกฝน และชุดข้อมูลทดสอบเป็นชุดข้อมูลย้อนหลัง 1) 3 วัน 2) 5 วัน 3) 7 วัน 4) 11 วัน 5) 15 วัน 6) 21 วัน 7) 31 วัน งานวิจัยนี้แสดงให้เห็นถึงการแบ่งประเภทเป็นรวมชนิดกับแยกชนิด จะสามารถได้ค่าความคลาดเคลื่อนที่เหมาะสมที่สุดกับแต่ละชนิด เพื่อนำไปปรับใช้ให้เกิดผลลัพธ์ที่ดีที่สุด (พิชญากร เลค, 2561) ได้ศึกษาเกี่ยวกับการพยากรณ์อัตราแลกเปลี่ยนสกุลเงินตราต่างประเทศโดยใช้การวิเคราะห์อนุกรมเวลา โดยใช้เทคนิคเหมือนข้อมูล 3 เทคนิค ได้แก่ 1) การถดถอยเชิงเส้น 2) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยใช้ข้อมูลอัตราแลกเปลี่ยนเงินตราต่างประเทศ ปี 2552 - 2560 จำนวน 108 เดือน งานวิจัยนี้แสดงให้เห็นถึงการเลือกใช้เทคนิคเหมือนข้อมูลแต่ละชนิด ข้อมูลย้อนหลังที่เหมาะสมกับชนิดของข้อมูลเป็นสิ่งที่ทำให้เห็นถึงประสิทธิภาพของแบบจำลองแต่ละแบบ (ปัทมญา บุญรักษา และจารี ทองคำ, 2560) ได้พัฒนาแบบจำลองการเกิดอุบัติเหตุทางถนนโดยใช้เทคนิคอนุกรมเวลา โดยการใช้เทคนิคเหมือนข้อมูล 5 เทคนิค ได้แก่ 1) linear regression 2) Artificial Neural Network 3) Sequential Minimal Optimization for Regression 4) Support Vector Machine Regression 5) Gaussian Process โดยใช้ข้อมูลการเกิดอุบัติเหตุช่วงเดือนมกราคม พ.ศ. 2554 - ธันวาคม พ.ศ. 2559 เป็นจำนวน 7 ปี ทั้งสิ้น 78 เดือน แบ่งชุดข้อมูลออกเป็น 2 ส่วน 1) ชุดข้อมูลสำหรับฝึกสอน 2) ชุดข้อมูลสำหรับทดสอบ โดยทำการเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยค่า Mean Absolute Error (MAE) และ Root Mean Square Error (RMSE) งานวิจัยนี้แสดงให้เห็นถึงกราฟข้อมูลการเกิดอุบัติเหตุย้อนหลัง มีลักษณะเป็นการผันแปรตามฤดูกาล ทั้งนี้ผู้วิจัยยังได้แนะนำถึงปัจจัยภายนอกอื่น ๆ เช่น แสงส่องสว่างที่อาจเกิดชำรุดในภายหลังอาจเป็นปัจจัยเพิ่มเติมได้ เป็นต้น (ฉริยะ อัครวรรณ และจารี ทองคำ, 2561) ได้ศึกษาเกี่ยวกับการสร้างตัวแบบจำลองการพยากรณ์การใช้หนังสือด้วยเทคนิคอนุกรมเวลา โดยมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของเทคนิคอนุกรมเวลาการใช้หนังสือของสมาชิกห้องสมุดนำข้อมูลการใช้หนังสือระหว่างช่วงปี พ.ศ. 2557 - 2560 โดยใช้โปรแกรม Weka สร้างแบบจำลองด้วย



เทคนิคเหมืองข้อมูล 5 เทคนิค ได้แก่ 1) โครงข่ายประสาทเทียมแบบหลายชั้นแบบถดถอย 2) โครงข่ายประสาทเทียมแบบถดถอย 3) เทคนิคสนับสนุนเส้นสมมุติแบบถดถอย 4) การวิเคราะห์การถดถอยโลจิสติก 5) การเรียนรู้ต้นไม้ตัดสินใจแบบถดถอยอัลกอริทึมการตัดกิ่งต้นไม้ โดยวัดค่าด้วยการหาค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ และค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง ผลการวิจัยชุดข้อมูลหนังสือมีการใช้งานสูงในช่วงเดือนกุมภาพันธ์ - มีนาคม และกันยายน - ตุลาคม และการใช้งานต่ำในช่วงพฤษภาคม - มิถุนายน และธันวาคมของทุกปี งานวิจัยนี้แสดงให้เห็นถึงการแบ่งชุดข้อมูลการเรียนรู้และชุดข้อมูลสำหรับการทดสอบในรูปแบบ Slide window แล้วจึงวัดประสิทธิภาพของแบบจำลองด้วยค่า MAE และ RMSE เพื่อหาค่าเฉลี่ยที่ดีที่สุด ซึ่งได้ผลการทดลองเป็นเทคนิคสนับสนุนเส้นสมมุติแบบถดถอยได้ค่าเฉลี่ยน้อยที่สุด จึงเป็นแบบจำลองที่มีประสิทธิภาพมากที่สุด

ดังนั้นเพื่อทราบทิศทางราคาน้ำมันแก๊สโซฮอล์ 95 ในอนาคต จึงนำมาสู่ความสนใจของผู้วิจัยในการสร้างแบบจำลองพยากรณ์ด้วยเทคนิคเหมืองข้อมูล (Data Mining) จำนวน 3 วิธี ได้แก่ 1) การถดถอยเชิงเส้น (Linear Regression) 2) แบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น (Mutilayer Perceptron) และ 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) และเปรียบเทียบประสิทธิภาพทั้ง 3 วิธีด้วยการวัดค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean Magnitude of Relative Error: MMRE) และเพื่อหาแบบจำลองพยากรณ์ที่เหมาะสมใช้เป็นเครื่องมือในการตัดสินใจวางแผนด้านพลังงานและปริมาณการใช้น้ำมันภายในประเทศต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างแบบจำลองพยากรณ์ราคาน้ำมันแก๊สโซฮอล์ 95 ด้วยเทคนิคเหมืองข้อมูล จำนวน 3 วิธี ได้แก่ 1) วิธีการถดถอยเชิงเส้น 2) แบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น และ 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย
2. เพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์ และวิธีที่เหมาะสมในการพยากรณ์ราคาน้ำมันแก๊สโซฮอล์ 95

สมมติฐาน

การพยากรณ์ราคาน้ำมันแก๊สโซฮอล์ 95 ที่จำหน่ายในอำเภอหาดใหญ่ จังหวัดสงขลา ด้วยเทคนิคเหมืองข้อมูล (Time Series Data Mining Techniques) ด้วยข้อมูลชุดเดียวกันทั้ง 3 วิธี สามารถพยากรณ์หาแนวโน้มราคาราคาน้ำมันแก๊สโซฮอล์ 95 ได้อย่างแม่นยำมีประสิทธิภาพเท่าเทียมกัน

ขอบเขตการวิจัยและกรอบแนวคิดในการวิจัย

1. ข้อมูลที่นำมาวิเคราะห์ เป็นข้อมูลราคาน้ำมันแก๊สโซฮอล์ 95 ที่จำหน่ายในอำเภอหาดใหญ่ จังหวัดสงขลา จากเว็บไซต์ของบริษัท ปตท. จำกัด (มหาชน) ตั้งแต่เดือนมกราคม 2566 ถึงเดือนมกราคม 2567 โดยแบ่งชุดข้อมูลออกเป็น 2 ส่วนคือ ชุดข้อมูลสำหรับฝึก และข้อมูลที่ใช้ทดสอบ
2. วิธีพยากรณ์หาแนวโน้มราคาราคาน้ำมันแก๊สโซฮอล์ 95 ที่จำหน่ายในอำเภอหาดใหญ่ จังหวัดสงขลา จำนวน 3 วิธี ได้แก่ 1) การถดถอยเชิงเส้น (Linear Regression) 2) แบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น (Mutilayer Perceptron) และ 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression)



วิธีดำเนินการวิจัย

การดำเนินการวิจัยประกอบขึ้นตอนดังนี้

- 1) ทบทวนวรรณกรรมที่เกี่ยวข้องกับการพยากรณ์ด้วยเทคนิคเหมืองข้อมูล
- 2) ดำเนินการเก็บรวบรวมข้อมูลราคาน้ำมันแก๊สโซฮอล์ 95 ที่จำหน่ายในอำเภอหาดใหญ่ จังหวัดสงขลา จากเว็บไซต์ของบริษัท ปตท. จำกัด (มหาชน) ตั้งแต่เดือนมกราคม 2566 ถึงเดือนมกราคม 2567
- 3) ศึกษาตัวแบบจำลองและทฤษฎีที่เกี่ยวข้อง

3.1 การทำเหมืองข้อมูล (Data Mining)

เหมืองข้อมูล (Data Mining) คือกระบวนการวิเคราะห์เพื่อค้นหาความสัมพันธ์หรือรูปแบบจากฐานข้อมูลขนาดใหญ่ โดยอาศัยหลักการทางสถิติและการเรียนรู้ของเครื่อง เพื่อนำข้อมูลสารสนเทศที่ได้มาประกอบการตัดสินใจหรือวางแผนทางธุรกิจในอนาคต (อดุลย์ ยิ้มงาม, 2551) ขั้นตอนการทำเหมืองข้อมูลมีกระบวนการวิเคราะห์ข้อมูลแบบมาตรฐานคือ CRISP - DM (Cross Industry Standard Process For Data Mining) มี 6 ขั้นตอนดังนี้ (เอกสิทธิ์ พัทธวงศ์ศักดิ์, 2564)

ขั้นตอนที่ 1. ทำความเข้าใจในธุรกิจ (Business understanding)

ขั้นตอนที่ 2. ทำความเข้าใจข้อมูล (Data understanding)

ขั้นตอนที่ 3. การเตรียมข้อมูล (Data Preparation)

ขั้นตอนที่ 4. การจัดทำตัวแบบ (Modelling)

ขั้นตอนที่ 5. การประเมินผล (Evaluation)

ขั้นตอนที่ 6. การนำเอาไปใช้งาน (Deployment)

3.2 ข้อมูลอนุกรมเวลา (Time Series Data)

ข้อมูลอนุกรมเวลา คือชุดข้อมูลหรือค่าสังเกตเชิงปริมาณ (X) ที่มีการเก็บรวบรวม โดยเรียงลำดับของเวลาอย่างต่อเนื่องกัน การจัดเรียงข้อมูลมีลักษณะเป็นรายปี (Yearly) รายไตรมาส (Quarterly) รายเดือน (Monthly) รายอาทิตย์ (Weekly) รายวัน (Daily) หรือรายชั่วโมง (Hourly) ขึ้นอยู่กับความเหมาะสมในการนำไปใช้งาน (เฉลิมพล จตุพร, 2562) ส่วนประกอบของอนุกรมเวลาประกอบด้วย 4 ส่วน คือ

1. แนวโน้ม (Trend) คือส่วนที่ทำให้อนุกรมเวลามีค่าเพิ่มหรือลดลงเรื่อย ๆ เมื่อเวลาผ่านไป (ภูมิฐาน รังकुณวัฒน์, 2562)

2. วัฏจักร (Cycle) คือส่วนที่ทำให้อนุกรมเวลาที่เก็บข้อมูลมีค่าเพิ่มขึ้น และลดลงสลับกันไปรอบ ๆ ค่าแนวโน้ม การนับระยะเวลาของส่วนวัฏจักรจะนับจุดสูงสุดหนึ่ง (peak) ไปยังอีกจุดสูงสุดหนึ่ง หรือจากจุดต่ำสุดหนึ่ง (trough) ไปยังอีกจุดต่ำสุดหนึ่ง เมื่อส่วนของวัฏจักรอยู่ในช่วงที่อนุกรมเวลามีค่าลดลงเรียกว่าช่วงถดถอย (Recession) ส่วนของวัฏจักรที่ทำให้อนุกรมเวลามีค่าเพิ่มขึ้นเรียกว่าช่วงฟื้นตัว (Recovery)

3. ความผันแปรจากฤดูกาล (Seasonal Variations) คือรูปแบบในช่วงเวลาหนึ่งของอนุกรมเวลาที่จะเป็นภายใน 1 ปีและเกิดขึ้นซ้ำกันทุกปี

4. ความผันผวนจากเหตุการณ์ไม่ปกติ (Irregular Fluctuations) คือ ส่วนที่ทำให้อนุกรมเวลามีค่าที่ผิดปกติไปจากรูปแบบที่เป็นมักเกิดจากเหตุการณ์ไม่คาดฝัน (Shock) คำนวณจากการนำค่าของส่วนแนวโน้ม ค่าของส่วนวัฏจักร และค่าของความผันแปรจากฤดูกาล ไปหักล้างค่าอนุกรมเวลานั้น



3.3 การวิเคราะห์การถดถอย (Regression analysis)

การวิเคราะห์การถดถอย เป็นเทคนิคทางสถิติที่ใช้ในการประมาณค่าความสัมพันธ์ระหว่างตัวแปรแบ่งออกเป็น 2 แบบ คือ 1) การวิเคราะห์การถดถอยเชิงเส้น (Linear regression analysis) เป็นการวิเคราะห์การถดถอยที่ตัวแปรอิสระส่วนใหญ่เป็นตัวแปรเชิงปริมาณ ส่วนตัวแปรตามเป็นตัวแปรเชิงปริมาณเท่านั้น รูปแบบของความสัมพันธ์สามารถแทนได้ด้วยสมการทางคณิตศาสตร์ที่เป็นเชิงเส้น (Linear model) 2) การวิเคราะห์การถดถอยแบบไม่เป็นเชิงเส้น (Non linear regression) เป็นการวิเคราะห์การถดถอยที่รูปแบบของความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม สามารถแทนได้ด้วยสมการทางคณิตศาสตร์ที่ไม่เป็นเชิงเส้น (non – Linear model) การวิเคราะห์การถดถอยแบ่งออกเป็น

1. แบบจำลองการถดถอยอย่างง่าย ประกอบด้วยตัวแปรตาม (Y) หนึ่งตัวแปร และตัวแปรอิสระ (X) เพียงหนึ่งตัวแปร ดังนี้ (เฉลิมพล จตุพร, 2562)

$$Y_i = \alpha_0 + \beta_1 X_i + \epsilon_i$$

$$Y_t = \alpha_0 + \beta_1 X_t + \epsilon_t$$

2. แบบจำลองการถดถอยพหุคูณ ประกอบด้วยตัวแปรตาม (Y) หนึ่งตัวแปร และตัวแปรอิสระ (X) มากกว่าหนึ่งตัวแปร

$$Y_t = \alpha_0 + \beta_1 X_{1t} + \alpha_0 + \beta_2 X_{2t} + \dots + \alpha_0 + \beta_n X_{nt} + \epsilon_t$$

$$Y_t = \alpha_0 + \sum_{i=1}^n \beta_i X_{it} + \epsilon_t$$

3.4 โครงข่ายประสาทเทียม (Artificial Neural Network: ANN)

โครงข่ายประสาทเทียม (Artificial Neural Network, ANN) คือ โมเดลทางคณิตศาสตร์สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนกชันนิสต์ (Connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ หลักการทำงานในการประมวลผลเกิดขึ้นในหน่วยประมวลผลย่อย (Node) เมื่อมีข้อมูลนำเข้า (Input Data) ส่งเข้ามายังโครงข่าย ข้อมูลนำเข้าแต่ละค่าจะถูกนำมาคูณกับค่าถ่วงน้ำหนัก (weight) แทนด้วยสัญลักษณ์ W_n ผลลัพธ์ที่ได้ถูกนำมาบวกกันแล้วส่งไปยังฟังก์ชันการแปลง (Transfer function) หรือเรียกว่า “ฟังก์ชันกระตุ้น (Activation function)” เพื่อหาข้อมูลผลลัพธ์ (Output Data) ส่งไปยังหน่วยประมวลผลย่อยถัดไป มีองค์ประกอบดังนี้

1. ชั้นอินพุต (Input Layer) ประกอบด้วยนิวรอนรับข้อมูล (Input Neurons) ทำหน้าที่เป็นตัวรับข้อมูลเข้าสู่แบบจำลองโครงข่ายประสาทเทียม ชั้นรับข้อมูล (Input Layer) ทำหน้าที่รับข้อมูลเข้าสู่นิวรอนจำนวนข้อมูลกับจำนวนนิวรอนเท่ากัน

2. ชั้นส่งออกหรือแสดงผลลัพธ์ (Output Layer) ทำหน้าที่แสดงผลลัพธ์ของข้อมูล จำนวนนิวรอนข้อมูลจะเท่ากับนิวรอนแสดงผลลัพธ์

3. ชั้นซ่อน (Hidden Layer) ทำหน้าที่รวมข้อมูลจากชั้นรับข้อมูลและค่าน้ำหนักส่งต่อไปยังชั้นแสดงผลลัพธ์ อาจมีหลายชั้นและมีนิวรอนจำนวนมาก

กระบวนการเรียนรู้ของโครงข่ายประสาทเทียม มี 2 ส่วนคือ การเรียนรู้แบบมีผู้สอน (Supervised Learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

3.5 โครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น (Multi-layer perceptron: MLP)

โครงข่ายประสาทเทียมแบบหลายชั้น (Feedforward neural network) เป็นตัวแบบประเภทการพยากรณ์ (prediction) ใช้พยากรณ์ข้อมูลอนุกรมเวลา มีโครงสร้าง 3 ชั้น ได้แก่ Input Layer, Hidden layer และ Output layer แต่ละชั้นจะประกอบด้วยโหนดที่มีจำนวนต่างกัน (ภัทร วรภู,



2556) โหนดในชั้น Input Layer จะเชื่อมไปข้างหน้ากับทุกโหนดในชั้นถัดไปตามลำดับจนถึง Output Layer โดยค่า Inputs ถูกส่งผ่านจาก Input neurons ไปยัง Hidden neurons แบบทั่วถึงกัน (Fully connected) จากนั้นเข้าสู่ Output neurons ค่าผลรวมของโหนดในชั้นซ่อน คำนวณโดยนำค่าของข้อมูลนำเข้าคูณกับค่าถ่วงน้ำหนักในแต่ละเส้นเชื่อมโยงไปยังโหนดที่ j นำมารวมกัน ดังนั้นค่าผลรวมของโหนดที่ j ในชั้นซ่อน คือ

$$net_j = \sum_{i=1}^N x_i w_{ij} + bias$$

เมื่อ X_i แทนข้อมูลนำเข้าที่ i , $i=1,...,N$ W_{ij} แทนค่าถ่วงน้ำหนักของเส้นเชื่อมโยงกับข้อมูลนำเข้าที่ i ไปยังโหนดที่ j , $j=1,...,J$ ปรับค่าผลรวมด้วย Sigmoid/logistic Activation Function จะได้ค่าของโหนดในชั้นซ่อนที่อยู่ในช่วง $(0,1)$

$$y_i = f(net_j) = \frac{1}{1 + e^{-net_j}}$$

พิจารณาโหนดในชั้นผลลัพธ์ คำนวณค่าผลรวมของโหนดในชั้นผลลัพธ์ (Output Layer) ซึ่งมี k โหนด

$$net_k = \sum_{j=1}^J y_j w_{jk} + bias$$

เมื่อ w_{jk} แทนค่าถ่วงน้ำหนักของเส้นเชื่อมโยงจากโหนดที่ j ในชั้น hidden layer ไปยังโหนดที่ k ในชั้น Output layer, $k=1,...,K$ ทำการปรับค่าผลรวมด้วยฟังก์ชันกระตุ้นโดยใช้ Linear Activation Function ซึ่งจะได้ค่าของโหนดในชั้น Output

$$y_k = f(net_k) = net_k$$

3.6 แบบจำลองพอร์ทเวกเตอร์แมชชีนสำหรับการถดถอย (Support vector machine model for regression)

หลักการของพอร์ทเวกเตอร์แมชชีนสำหรับการถดถอย (Support vector regression :SVR) คือการลดความเสี่ยงเชิงโครงสร้างให้มีค่าต่ำที่สุด ด้วยกระบวนการปรับรูปแบบข้อมูลจากข้อมูลที่มีมิติต่ำ (low dimension dataset) บนพื้นที่ข้อมูลนำเข้า (input space) ให้อยู่ในรูปแบบของข้อมูลที่มีมิติสูง (high dimension dataset) บนพื้นที่ข้อมูลคุณลักษณะ (Feature space) เมื่อกำหนดเซตข้อมูลนำเข้า โดยที่ฟังก์ชัน $f(x)$ จะต้องมียุทธศาสตร์ ϵ มากที่สุด ในขณะที่เดียวกันมีค่าคลาดเคลื่อน (Minimize error) น้อยที่สุด

$$f(x) = (w, x) + b, \quad w \in R^d, b \in R$$

การหาฟังก์ชันการถดถอยที่เหมาะสมที่สุด (optimal regression function) เป็นการหาค่าน้อยที่สุดของฟังก์ชันต่อไปนี้

$$\text{Minimize } -R_{svm}(C) = \frac{1}{2} ||w||^2 + C \sum_{i=1}^n L_\epsilon(y_i - f(x_i, w)) \text{----(1)}$$

$$L_\epsilon(y_i, f(x_i, w)) = \begin{cases} 0, & |y_i - f(x_i, w)| \leq \epsilon \\ |y_i - f(x_i, w)| - \epsilon, & \text{otherwise} \end{cases}$$

$$\text{Subject to } \begin{cases} y_i - (w, x) - b \leq \epsilon \\ (w, x) + b - y_i \leq \epsilon \end{cases}$$

พารามิเตอร์ ϵ เป็นตัวควบคุมความกว้างของ ϵ -intensive zone ใช้ในการ fit ข้อมูล training ส่งผลกระทบต่อการสร้างสมการ regression กล่าวคือหาก ϵ มีค่าน้อยทำให้ training error มีค่าต่ำ แต่อาจทำให้เกิดปัญหา overfitting ทำให้การพยากรณ์ในชุด testing data ไม่แม่นยำ หาก ϵ มีค่ามาก



เกินไปทำให้ training error มีค่าสูง ทำให้ไม่เกิดปัญหา overfitting มากนัก (Chen & Wang, 2007) ส่วน errors ที่มีค่ามากกว่า $\pm \epsilon$ จะถูกเรียกหรือกำหนดให้เป็น “Slack variable : $\xi_i^{(*)}$ ” ξ_i^* (เหนือ ϵ) และ ξ_i^* (ต่ำกว่า ϵ) แสดงระยะห่างระหว่างค่าจริงที่อยู่นอก ϵ -intensive zone และสมการ regression ดังสมการที่ (1) เปลี่ยนไปเป็นสมการที่ (2)

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2)$$

$$\text{Subject to } \begin{cases} y_i - \langle w, x \rangle - b \leq \epsilon + \xi_i \\ \langle w, x \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{cases}$$

3.7 เกณฑ์ที่ใช้ในการตัดสินใจ

เกณฑ์ที่ใช้ในการคัดเลือกตัวแบบในการพยากรณ์ในงานวิจัยนี้ ใช้ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย Mean Absolute Error (MAE) เป็นค่าวัดความถูกต้องของการพยากรณ์ที่วัดจากขนาดของค่าความคลาดเคลื่อนของการพยากรณ์ โดยไม่คำนึงถึงทิศทางของความคลาดเคลื่อน

$$MAE = \frac{\sum_{i=1}^n |x_t - \hat{x}_t|}{n}$$

3.8 การประมาณค่าความแม่นยำ (Evaluation Criterion)

การประมาณค่าความแม่นยำ (Evaluation Criterion) จากการใช้วิธีต่าง ๆ ที่สร้างขึ้นต้องมีความแม่นยำเข้ากันได้ (Model best fit) จะถูกนำไปทดสอบกับกลุ่มข้อมูลอีกชุดหนึ่งที่ทราบค่าจริง (Actual Data) ผลจากการพยากรณ์ข้อมูลชุดใหม่ (Predicted Data) จะถูกนำมาคำนวณหาความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error: MRE) โดยสามารถคำนวณได้จากสมการดังนี้

$$MRE_i = \frac{\text{Actual}_i - \text{predicted}_i}{\text{Actual}_i}$$

หากมีข้อมูลหลายชุด การคำนวณให้นำค่าความคลาดเคลื่อนสัมพัทธ์ (MRE) ของแต่ละชุดข้อมูลมารวมกันและหาค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean-MRE: MMRE) ค่า MMRE มากกว่า 0 แสดงว่ามีความคลาดเคลื่อนสูง ค่า MMRE = 0 (ศูนย์) แสดงว่าค่าของการพยากรณ์เท่ากับค่าจริงทุกค่า นั่นคือ MMRE ยิ่งมีค่าน้อยแสดงว่ามีความแม่นยำมาก ในการประมาณค่าสูญหายของข้อมูลคำนวณได้จากสมการดังนี้

$$MMRE = \frac{1}{2} \sum_i^n = 1 \frac{\text{Actual}_i - \text{predicted}_i}{\text{Actual}_i} \times 100$$

4) สร้างแบบจำลองสำหรับการพยากรณ์โดยใช้วิธีการวิเคราะห์อนุกรมเวลาด้วยเทคนิคเหมือนข้อมูล จำนวน 3 วิธี

5). คำนวณหาค่า MAE และค่าเฉลี่ยของ RMSE โดยใช้ชุดข้อมูลสำหรับเรียนรู้ แบ่งช่วงเวลาย้อนหลัง (Lagged) 10 วัน, 12 วัน, 25 วัน, 28 วัน และ 30 วัน

6) พิจารณาคัดเลือกค่า MAE ที่ค่าใกล้เคียงกันของแบบจำลองพยากรณ์ทั้ง 3 วิธี

7) พยากรณ์ราคาน้ำมันแก๊สโซฮอล์ 95 ด้วยชุดข้อมูลสำหรับการทดสอบ เพื่อคำนวณหาค่า MRE และ MMRE

8) เปรียบเทียบความแม่นยำของแบบจำลองพยากรณ์ทั้ง 3 วิธี



9) วิเคราะห์และสรุปผลการวิจัย

ผลการวิจัย

จากการเก็บข้อมูลและสร้างแบบจำลองพยากรณ์ เพื่อทำการเปรียบเทียบประสิทธิภาพการพยากรณ์ ข้อมูลราคาน้ำมันแก๊สโซฮอล์ 95 จำนวน 3 วิธี ด้วยการแบ่งชุดข้อมูลออกเป็นชุดข้อมูลการเรียนรู้และชุดข้อมูลสำหรับการทดสอบ (ฉริยะ อัครวรรณ และจารี ทองคำ, 2561) ผลจากการเปรียบเทียบดังตารางที่ 1

ตารางที่ 1 ผลการเปรียบเทียบประสิทธิภาพการพยากรณ์ข้อมูลราคาน้ำมันแก๊สโซฮอล์ 95

Lagged		Time Series Data Mining Techniques					
		Linear Regression		Multilayer Perceptron		SMOreg	
		MAE	RMSE	MAE	RMSE	MAE	RMSE
แก๊สโซฮอล์ 95	Lag 10	0.3502	0.3959	0.3134	0.4028	0.333	0.4092
	Lag 12	0.333	0.3822	0.2804	0.3385	0.3289	0.4207
	Lag 25	0.2815	0.3344	0.2427	0.295	0.2806	0.3765
	Lag 28	0.2809	0.3315	0.2196	0.2675	0.2746	0.358
	Lag 30	0.269	0.3122	0.209	0.2534	0.2602	0.343

จากตารางที่ 1 พบว่า การแบ่งชุดข้อมูลย้อนหลัง (Lag) ที่แตกต่างกัน และหาค่า MAE และ RMSE สำหรับวัดประสิทธิภาพของแบบจำลองทั้ง 3 วิธี ได้ผลดังนี้ วิธีการถดถอยเชิงเส้นด้วยข้อมูลย้อนหลัง 28 วัน มีค่า MAE เท่ากับ 0.2809 และค่า RMSE เท่ากับ 0.3315 แบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซ็ปตรอนหลายชั้นด้วยข้อมูลย้อนหลัง 12 วัน มีค่า MAE เท่ากับ 0.2804 และค่า RMSE เท่ากับ 0.3385 และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยด้วยข้อมูลย้อนหลัง 25 วัน มีค่า MAE เท่ากับ 0.2806 และค่า RMSE เท่ากับ 0.3765 ซึ่งค่าความคลาดเคลื่อนที่ได้มีค่าที่ใกล้เคียงกัน (สุภาพร บรรดาศักดิ์, จุฬามาศ แป้นทอง และนุทยา กัลป์ลา, 2559)

จากสมมุติฐานที่ว่าแบบจำลองพยากรณ์ด้วยข้อมูลชุดเดียวกันทั้ง 3 วิธี สามารถพยากรณ์หาแนวโน้มราคาราคาน้ำมันแก๊สโซฮอล์ 95 ได้อย่างแม่นยำมีประสิทธิภาพเท่าเทียมกัน เมื่อทำการเปรียบเทียบแบบจำลองการพยากรณ์ทั้ง 3 วิธีด้วยชุดข้อมูลสำหรับการทดสอบเดือนมกราคม 2567 จำนวน 30 วัน ด้วยการวัดค่าความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error: MRE) ผลการเปรียบเทียบดังตารางที่ 2



ตารางที่ 2 ตารางเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายราคาแก๊สโซฮอล์ 95

Data Set			Time Series Data Mining Techniques					
Data	1 month data		Linear Regression		Multilayer Perceptron		SMOreg	
			Lag 28		Lag 12		Lag 25	
	Data	Actual	Predict	MRE	Predict	MRE	Predict	MRE
น้ำมัน แก๊ส โซฮอล์ 95	1	35.57	35.005	0.0162	34.894	0.0194	35.425	0.0041
	2	35.57	28.131	0.2644	35.221	0.0099	35.465	0.0030
	3	35.57	28.877	0.2318	35.032	0.0154	35.457	0.0032
	4	35.57	29.612	0.2012	34.866	0.0202	35.451	0.0034
	5	35.07	30.334	0.1561	34.724	0.0100	35.447	-0.0106
	6	35.07	31.148	0.1259	34.767	0.0087	35.065	0.0002
	7	35.07	31.837	0.1016	34.711	0.0104	35.065	0.0002
	8	35.07	32.505	0.0789	34.682	0.0112	35.067	0.0001
	9	35.07	33.149	0.0579	34.682	0.0112	35.071	0.0000
	10	35.07	33.769	0.0385	34.712	0.0103	35.077	-0.0002
	11	35.07	34.359	0.0207	34.773	0.0085	35.086	-0.0005
	12	35.07	34.919	0.0043	34.866	0.0059	35.098	-0.0008
	13	35.07	35.445	-0.0106	34.992	0.0022	35.112	-0.0012
	14	35.07	35.936	-0.0241	35.153	-0.0023	35.129	-0.0017
	15	35.07	36.387	-0.0362	35.349	-0.0079	35.149	-0.0022
	16	35.57	36.797	-0.0334	35.581	-0.0003	35.172	0.0113
	17	35.57	37.057	-0.0401	35.306	0.0075	35.576	-0.0002
	18	35.87	37.377	-0.0403	35.579	0.0082	35.605	0.0075
	19	35.87	37.584	-0.0456	35.520	0.0098	35.863	0.0002
	20	36.17	37.803	-0.0432	35.853	0.0089	35.899	0.0075
	21	36.17	35.005	0.0162	34.894	0.0194	35.425	0.0041
	22	36.17	37.904	-0.0457	35.810	0.0100	36.165	0.0002
	23	36.67	38.013	-0.0485	36.205	-0.0010	36.207	-0.0010
	24	36.67	38.062	-0.0366	36.642	0.0008	36.254	0.0115
	25	37.17	37.943	-0.0335	36.319	0.0097	36.681	-0.0003
	26	37.17	37.866	-0.0184	36.808	0.0098	36.735	0.0118
	27	37.57	37.615	-0.0118	36.463	0.0194	37.170	0.0000
	28	37.57	37.401	0.0045	37.004	0.0153	37.232	0.0091
	29	37.57	37.030	0.0146	36.828	0.0201	37.599	-0.0008
	30	37.57	36.669	0.0246	37.428	0.0038	37.670	-0.0027



ตารางที่ 2 (ต่อ)

Data Set			Time Series Data Mining Techniques					
Data	1 month data		Linear Regression		Multilayer Perceptron		SMOreg	
			Lag 28		Lag 12		Lag 25	
	Data	Actual	Predict	MRE	Predict	MRE	Predict	MRE
	31	37.87	36.231	0.0369	38.069	-0.0131	37.745	-0.0046
MMRE			3.131%		0.707%		0.153%	

จากตารางที่ 2 แสดงผลการเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายทั้ง 3 วิธีกับชุดข้อมูลสำหรับทดสอบ โดยใช้วิธีประมาณการความแม่นยำในการพยากรณ์ด้วยค่าความคลาดเคลื่อนสัมพัทธ์ (Mean Magnitude of Relative Error: MMRE) พบว่า แบบจำลองโครงข่ายประสาทเทียมซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยให้ประสิทธิภาพสูงสุดด้วยชุดข้อมูลราคาน้ำมันแก๊สโซฮอล์ 95 ย้อนหลัง 25 วัน (Lag 25) มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ เท่ากับ 0.153% แบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้นด้วยชุดข้อมูลย้อนหลัง 12 วัน (Lag 12) มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 0.707% และวิธีการถดถอยเชิงเส้นด้วยชุดข้อมูลย้อนหลัง 28 วัน (Lag 28) มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 3.131% ตามลำดับ

อภิปรายผล

ผลจากการสร้างและเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ราคาน้ำมันแก๊สโซฮอล์ 95 ที่จำหน่ายในอำเภอหาดใหญ่ จังหวัดสงขลา ด้วยเทคนิคเหมืองข้อมูล พบว่า ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยด้วยชุดข้อมูลย้อนหลัง 25 วัน (Lag 25) เมื่อทำการวัดด้วยค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ มีค่าเท่ากับ 0.153% ส่วนแบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้นด้วยชุดข้อมูลย้อนหลัง 12 วัน (Lag 12) เมื่อทำการวัดด้วยค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์มีค่าเท่ากับ 0.707% และแบบจำลองการถดถอยเชิงเส้นด้วยชุดข้อมูลย้อนหลัง 28 วัน (Lag 28) เมื่อทำการวัดด้วยค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์มีค่าเท่ากับ 3.131%

เมื่อพิจารณาเลือกใช้เทคนิคเหมืองข้อมูลแต่ละชนิด ด้วยข้อมูลย้อนหลังที่เหมาะสมกับชนิดของข้อมูล (พิชญากร เลค, 2561) พบว่า แบบจำลองโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้นมีความเหมาะสมที่สุดในการนำไปวิเคราะห์กับชุดข้อมูลราคาน้ำมันแก๊สโซฮอล์ 95 ด้วยชุดข้อมูลย้อนหลัง 25 วัน

ดังนั้นผลจากการวิจัยสามารถนำแบบจำลองดังกล่าวไปพัฒนาต่อเพื่อดูแนวโน้มของราคาน้ำมันในการตัดสินใจวางแผนด้านพลังงานและปริมาณการใช้น้ำมัน

ข้อเสนอแนะ

การที่แบบจำลองจะสามารถทำนายผลได้อย่างแม่นยำต้องอาศัยข้อมูลเป็นจำนวนมาก และมีความสมบูรณ์ของข้อมูล ดังนั้นจึงควรศึกษาข้อมูลจากหลากหลายแหล่งก่อนเริ่มทำงานวิจัยในหัวข้อดังกล่าว เพื่อศึกษาถึงความเป็นไปได้ต่าง ๆ จะทำให้การพยากรณ์มีความแม่นยำมีประสิทธิภาพมากขึ้น



เอกสารอ้างอิง

- กองบริหารธุรกิจและการสำรอน้ำมันเชื้อเพลิง กรมธุรกิจพลังงาน กระทรวงพลังงาน. (2565).
สถานการณ์การใช้น้ำมันเชื้อเพลิงรอบ 8 เดือนของปี 2566 (มกราคม-สิงหาคม). สืบค้นเมื่อ 18
กุมภาพันธ์ 2567 จาก https://www.doeb.go.th/news_activity/2528.pdf.
- ฉริยะ อัครวรรณ และจารี ทองคำ. (2561). การพยากรณ์การใช้น้ำมันด้วยเทคนิคอนุกรมเวลา.
มหาสารคาม : มหาวิทยาลัยมหาสารคาม.
- เฉลิมพล จตุพร. (2562) การวิเคราะห์อนุกรมเวลาเบื้องต้น (Basic of Time Series Analysis). สืบค้น
เมื่อ 16 กุมภาพันธ์ 2567 จาก [https://cj007blog.files.wordpress.com/2020/04/03-
basic-of-time-series-analysis.pdf](https://cj007blog.files.wordpress.com/2020/04/03-basic-of-time-series-analysis.pdf)
- ปัทมธนา บุญรักษา และจารี ทองคำ. (2560). การเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิด
อุบัติเหตุทางถนน โดยใช้เทคนิคอนุกรมเวลา. มหาสารคาม : มหาวิทยาลัยมหาสารคาม :
- พิชญากร เลค. (2561). พยากรณ์อัตราการแลกเปลี่ยนสกุลเงินต่างประเทศ. กรุงเทพฯ : มหาวิทยาลัย
สยาม :
- ภัทร วรภู (2556). การเปรียบเทียบความแม่นยำของการพยากรณ์อนุกรมเวลาระหว่างตัวแบบผสม และ
ตัวแบบเดี่ยว. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- ภูมิฐาน รังคกุลวัฒน์. (2556) การวิเคราะห์อนุกรมเวลาสำหรับเศรษฐศาสตร์และธุรกิจ (พิมพ์ครั้งที่ 1).
กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- สุภาพร บรรดาศักดิ์, จุฑามาศ แป้นทอง และ นุตยา กัลป์ลา. (2559). พยากรณ์การใช้ลูกกอล์ฟในสนาม
ฝึกซ้อมโดยใช้ตัวแปรอนุกรมเวลา. กรุงเทพฯ : มหาวิทยาลัยเกษตรศาสตร์.
- อดุลย์ ยิ้มงาม (2565). การทำเหมืองข้อมูล (Data Mining). สืบค้นเมื่อ 16 กุมภาพันธ์ 2567 จาก
<https://sites.google.com/a/bu.ac.th/compcenter/news-information/data-mining>
- เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2565) เอกสารประกอบการอบรมหลักสูตรอบรม Practical Data Mining with
RapidMiner Studio 9. สืบค้นเมื่อ 16 กุมภาพันธ์ 2567 จาก [https://datacubeth.ai/crisp-
dm](https://datacubeth.ai/crisp-dm)