วารสารวิทยาศาสตร์ลาดกระบัง ปีที่ 33 ฉบับที่ 1 เดือนมกราคม - มิถุนายน 2567
Journal of Science Ladkrabang Vol. 33 No. 1 January - June 2024

บทความวิจัย (Research article)

การเปรียบเทียบประสิทธิภาพตัวแบบระหว่าง การถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ และการเรียนรู้ของเครื่อง สำหรับราคาน้ำมันดีเซลในประเทศไทย

Comparison of Model Performance between Penalized Regression and Machine Learning for Diesel Prices in Thailand

กุลวดี เวหนรัตน์ ภัชพาภรณ์ ราชวงค์ สมฤทัย แซ่อึ้ง และ พรรณทิพา วาณิชย์จิรัฐติกาล*
Kulwadee Wahanarat Phatcharaphon Ratchawong Somruethai Ze-ueng and Puntipa Wanitjirattikal*
ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ ประเทศไทย
Department of Statistics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
วันที่ส่งบทความ : 8 สิงหาคม 2566 วันที่แก้ไขบทความ : 28 มิถุนายน 2567 วันที่ตอบรับบทความ : 28 มิถุนายน 2567
Received: 8 August 2023, Revised: 28 June 2024, Accepted: 28 June 2024

บทคัดย่อ

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพตัวแบบของราคาน้ำมันดีเซล B7 ราคาน้ำมันดีเซล B10 และราคาน้ำมันดีเซล B20 และหาตัวแบบที่เหมาะสมในการพยากรณ์ราคาน้ำมันดีเซล B7 ราคาน้ำมันดีเซล B10 และราคาน้ำมันดีเซล B20 โดยใช้ข้อมูลรายวัน ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึงวันที่ 31 ธันวาคม พ.ศ. 2565 ซึ่งปัจจัยที่ใช้ในการศึกษาครั้งนี้ ได้แก่ ราคาน้ำมันดิบในตลาดโลก ปริมาณการใช้น้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20 อัตราแลกเปลี่ยน ดัชนีราคาผู้บริโภค อัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20 และราคาน้ำมันปาล์มดิบ และทำการเปรียบเทียบประสิทธิภาพของตัวแบบด้วยการเปรียบเทียบค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) โดยวิธีการหาตัวแบบด้วยการถดถอยแบบขั้นตอน (Stepwise regression) วิธีการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ (Penalized Regression) อันได้แก่ การถดถอยแบบริดจ์ (Ridge regression) การถดถอยแบบลาสโซ่ (Lasso regression) การถดถอยลาสโซ่แบบปรับปรุง (Adaptive lasso regression) การถดถอยแบบอิลาสติคเน็ต (Elastic net regression) และการเรียนรู้ของเครื่อง (Machine learning) ได้แก่ วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน (Support vector regression) และวิธีป่าสุ่ม (Random forest) ผลการวิจัยพบว่าวิธีป่าสุ่มเป็นวิธีที่เหมาะสมที่สุดในการหาตัวแบบที่เหมาะสมในการพยากรณ์ราคาน้ำมันดีเซล B7 ราคาน้ำมันดีเซล B10 และราคาน้ำมันดีเซล B20 โดยมีค่า RMSE เท่ากับ 0.379, 0.3833 และ 0.3539 ตามลำดับ

คำสำคัญ : วิธีการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ การเรียนรู้ของเครื่อง วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน วิธีป่าสุ่ม

*ที่อยู่ติดต่อ E-mail address: puntipa.wa@kmitl.ac.th https://doi.org/10.55003/scikmitl.2024.259980

Abstract

This research study was to compare of model performance for the price of diesel B7, diesel B10, and diesel B20 and find a suitable model to forecast the price of diesel B7, diesel B10, and diesel B20 using daily data from January 1, 2020 to December 31, 2022. The factors used in the study were crude oil prices in the world market, consumption of diesel B7, diesel B10, and diesel B20, exchange rate, consumer price index, oil fuel fund rate of diesel B7, diesel B10, and diesel B20, and crude palm oil prices. Then, compare models to find the best model by measuring the model's performance with the Root Mean Square Error (RMSE) using Stepwise Regression, Penalized Regression such as Ridge Regression, Lasso Regression, Adaptive Lasso Regression, Elastic Net Regression, and Machine Learning such as Support Vector Regression and Random Forest. According to the research results, Random Forest is the most suitable method for forecasting the price of diesel B7, diesel B10, and diesel B20 with RMSE values of 0.379, 0.3833, and 0.3539, respectively.

Keywords: Penalized regression, Machine learning, Support vector regression, Random forest

1. บทนำ

น้ำมันเชื้อเพลิงจัดเป็นทรัพยากรธรรมชาติที่ถือว่าเป็นปัจจัยที่สำคัญต่อการดำรงชีวิต เนื่องจากถูกนำไปใช้เป็นส่วนหนึ่งของกระบวนการผลิตทั้งในภาคอุตสาหกรรม ภาคเกษตรกรรม ภาคบริการ และภาคการค้า ทั้งยังมีความสำคัญต่อการพัฒนาเศรษฐกิจของทุกประเทศรวมถึงประเทศไทยด้วย ทำให้ในการปรับขึ้นราคาน้ำมันดิบส่งผลกระทบอย่างรุนแรงต่อภาคเศรษฐกิจอย่างต่อเนื่อง โดยเป็นอุปสรรคต่อการขยายตัวทางเศรษฐกิจของประเทศไทยทั้งในอดีต ปัจจุบันและอนาคต ซึ่งน้ำมันสำเร็จรูปที่ใช้ในประเทศไทยมาจากการนำเข้าน้ำมันดิบจากต่างประเทศและนำมากลั่นเป็นน้ำมันสำเร็จรูป นอกจากนี้ยังมีการนำเข้าน้ำมันสำเร็จรูปบางส่วนจากต่างประเทศเข้ามาใช้ ทั้งนี้เพราะประเทศไทยมีแหล่งน้ำมันดิบไม่เพียงพอกับความต้องการ ในปี พ.ศ. 2565 ประเทศไทยมีการนำเข้าน้ำมันดิบเฉลี่ย 913,297 บาร์เรล/วัน เพิ่มขึ้น 5.8% เมื่อเทียบกับปี พ.ศ. 2564 และมีการนำเข้าน้ำมันสำเร็จรูปเฉลี่ย 77,971 บาร์เรล/วัน เพิ่มขึ้น 123.9% เมื่อเทียบกับปี พ.ศ. 2564 จะเห็นได้ว่าประเทศไทยมีแนวโน้มการนำเข้าน้ำมันสำเร็จรูปเพิ่มขึ้น โดยเฉพาะน้ำมันกลุ่มดีเซลซึ่งใช้ในภาคการขนส่งเป็นหลักมีปริมาณการใช้เฉลี่ยสูงถึง 73.05 ล้านลิตร/วัน [1] ในปี พ.ศ. 2565 มีการใช้น้ำมันดีเซล B7 เพิ่มขึ้นมาอยู่ที่ 63.88 ล้านลิตร/วัน และน้ำมันดีเซลพื้นฐานเพิ่มขึ้นมาอยู่ที่ 6.70 ล้านลิตร/วัน และน้ำมันดีเซลหมุนเร็ว B20 มีปริมาณการใช้ 0.19 ล้านลิตร/วัน [2] ซึ่งน้ำมันสำเร็จรูปถือว่าเป็นพลังงานสิ้นเปลืองเพราะเป็นพลังงานที่ได้จากทรัพยากรธรรมชาติที่มีอยู่ในปริมาณจำกัดและในวันหนึ่งสามารถหมดไปได้ แต่ไม่สามารถเกิดขึ้นใหม่หรือสังเคราะห์ทดแทนกันได้ เนื่องจากผลิตไม่ทันความต้องการทำให้ทั่วโลกกำลังเผชิญกับปัญหาการขาดแคลนพลังงานจึงเกิดพลังงานทดแทนขึ้นมา ซึ่ง

พลังงานที่ใช้ทดแทนพลังงานจากฟอสซิล เช่น ถ่านหิน ปิโตรเลียมและก๊าซธรรมชาติ เป็นต้น ซึ่งเป็นพลังงานสิ้นเปลืองที่จะหมดลงเมื่อใช้ไป

จากงานวิจัยของ มารุต จำลอง และศักดิ์ชาย นาคนก [3] ศึกษาปัจจัยที่ส่งผลกระทบต่อราคาน้ำมันสำเร็จรูปดีเซลและการพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย ทำการวิเคราะห์การถดถอยพหุคูณด้วยวิธีการหาตัวแบบด้วยการถดถอยแบบขั้นตอน (Stepwise regression) กับตัวแปรอิสระดังต่อไปนี้ อัตราภาษีสรรพสามิต (บาทต่อลิตร) กองทุนน้ำมันเชื้อเพลิง ดัชนีราคาผู้บริโภค และราคาน้ำมันดีเซลในตลาดสิงคโปร์ พบว่าเกิดปัญหาพหุสัมพันธ์ (Multicollinearity) นั่นคือตัวแปรอิสระแต่ละตัวมีความสัมพันธ์กับตัวแปรอิสระ ทำให้มีการตัดตัวแปรบางตัวที่มีความสัมพันธ์กันซึ่งเป็นตัวแปรที่น่าสนใจทิ้งไป และอาจเป็นปัจจัยที่ส่งผลต่อราคาน้ำมันดีเซล และเนื่องจากข้อมูลมีลักษณะเป็นอนุกรมเวลา (Time series) จึงอาจทำให้เกิดค่าสหสัมพันธ์ของค่าคลาดเคลื่อน (Autocorrelation) ซึ่งทำให้การสรุปผลของตัวแบบไม่เหมาะสมเท่าที่ควร ดังนั้นจึงเป็นที่มาของการหาตัวแบบด้วยวิธีอื่น ๆ

จากการศึกษาของนักวิจัยหลายท่านพบว่าการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ (Penalized regression) ช่วยแก้ปัญหาพหุสัมพันธ์และการเกิดสหสัมพันธ์ของค่าคลาดเคลื่อนได้ เช่น จากงานวิจัยของ Hoerl และ Kennard [4] ได้เสนอวิธีการถดถอยแบบริดจ์ (Ridge regression) ซึ่งเป็นวิธีที่นิยมสำหรับสร้างตัวแบบเพื่อแก้ปัญหาพหุสัมพันธ์ที่เกิดขึ้น และในงานวิจัยของ Tibshirani [5] ได้เสนอวิธีถดถอยลาสโซ่ (Least absolute shrinkage and selection operator: Lasso regression) ทั้งนี้งานวิจัยของ Zou และ Hastie [6] ได้เสนอวิธีการถดถอยอิลาสติคเน็ต (Elastic net regression) ที่พัฒนาขึ้นมาเพื่อแก้ไขข้อจำกัดของวิธีลาสโซ่ ซึ่งวิธีอิลาสติคเน็ตจะรวมระหว่างวิธีริดจ์และวิธีลาสโซ่เข้าด้วยกัน ซึ่งเหมาะสำหรับการวิเคราะห์ข้อมูลที่มีจำนวนตัวแปรอิสระมากกว่าขนาดตัวอย่างเป็นจำนวนมาก และเกิดปัญหาพหุสัมพันธ์ ซึ่งสอดคล้องกับงานวิจัยของวรางคณา วัชรเสถียร [7] เบญจมาศ รุ่งศรานนทร์ และอัชฌา อระวีพร [8] ที่ได้ทำการวิจัยเกี่ยวกับการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ

ปัจจุบันการเรียนรู้ของเครื่อง (Machine learning) เป็นแนวทางใหม่ที่ทันสมัยที่มีงานวิจัยทั้งหลายนำประยุกต์ใช้ ทั้งนี้มีงานวิจัยที่เกี่ยวข้องกับสิ่งที่ทางนักวิจัยสนใจ ได้แก่ งานวิจัยของ Yasin และคณะ [9] ทำการทำนายราคาน้ำมันดิบโดยใช้วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression: SVR) ในการวิเคราะห์ความถดถอยสำหรับข้อมูลอนุกรมเวลา

จากงานวิจัยของมารุต จำลอง และศักดิ์ชาย นาคนก [3] ฐิติรัตน์ แพทย์มงคล [10] วิลาสินี หีบแก้ว [11] และณัฐพินท์ เดชขุน [12] ทำให้ผู้วิจัยสนใจที่จะศึกษาปัจจัยที่เกี่ยวข้องกับราคาน้ำมันดีเซล B7 ราคาน้ำมันดีเซล B10 และราคาน้ำมันดีเซล B20 ในประเทศไทย โดยนำตัวแปรอิสระที่ส่งผลต่อราคาน้ำมันในแต่ละงานวิจัยมาศึกษาต่อซึ่งได้แก่ ราคาน้ำมันดิบในตลาดโลก ปริมาณการใช้น้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20 อัตราแลกเปลี่ยน ดัชนีราคาผู้บริโภค อัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20 และราคาน้ำมันปาล์มดิบ เพื่อนำไปสร้างตัวแบบที่เหมาะสม โดยใช้วิธีการหาตัวแบบด้วยการถดถอยแบบขั้นตอน การถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ และการเรียนรู้ของเครื่อง ทั้งนี้ได้ทำการวิเคราะห์ข้อมูลด้วยโปรแกรมภาษาไพธอน

2. วิธีดำเนินงานวิจัย

2.1 ขั้นตอนการดำเนินงานวิจัย

ในงานวิจัยครั้งนี้มีขั้นตอนการดำเนินงานวิจัย ดังรูปที่ 1 มีรายละเอียดดังต่อไปนี้

1. เตรียมข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึงวันที่ 31 ธันวาคม พ.ศ. 2565 โดยมีตัวแปรดังนี้

ตัวแปรตาม ได้แก่

- ราคาน้ำมันดีเซล B7 ราคาน้ำมันดีเซล B10 และราคาน้ำมันดีเซล B20

ตัวแปรอิสระ ได้แก่

- ราคาน้ำมันดิบในตลาดโลก
- ปริมาณการใช้น้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20
- อัตราแลกเปลี่ยน
- ดัชนีราคาผู้บริโภค
- อัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20 และ
- ราคาน้ำมันปาล์มดิบ

2. ทำความสะอาดข้อมูล (Data cleaning)

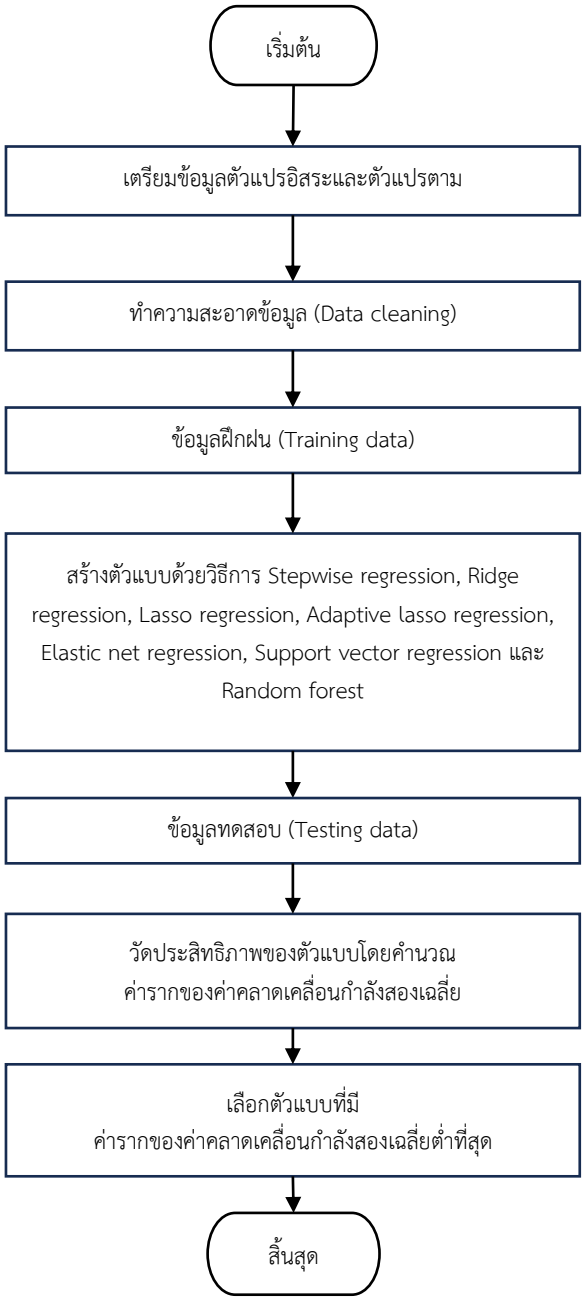
3. แบ่งข้อมูลออกเป็น 2 ส่วน ด้วยอัตราส่วน 70:30 คือ ข้อมูลฝึกฝน และข้อมูลทดสอบ ตามลำดับ

4. สร้างตัวแบบด้วยวิธีการถดถอยแบบขั้นตอน การถดถอยแบบริดจ์ การถดถอยแบบลาสโซ่ การถดถอยลาสโซ่แบบปรับปรุง (Adaptive Lasso Regression) การถดถอยแบบอิลาสติคเน็ต วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน และวิธีป่าสุ่ม (Random Forest) จากข้อมูลฝึกฝน

5. นำข้อมูลทดสอบแทนค่าลงในตัวแบบที่ได้ในแต่ละวิธี และคำนวณค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

6. ทำการเปรียบเทียบประสิทธิภาพของตัวแบบด้วยการเปรียบเทียบค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

7. เลือกตัวแบบที่ให้ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุด



รูปที่ 1. ขั้นตอนการดำเนินงานวิจัย

2.2 แหล่งที่มาของข้อมูล

การเปรียบเทียบประสิทธิภาพตัวแบบระหว่างการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ และการเรียนรู้ของเครื่อง สำหรับราคาน้ำมันดีเซลในประเทศไทยได้ทำการศึกษาจากข้อมูลทุติยภูมิ (Secondary data) ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2563 ถึงวันที่ 31 ธันวาคม พ.ศ. 2565 แบบข้อมูลอนุกรมเวลารายวัน เป็นเวลา 3 ปี และมีจำนวนข้อมูลทั้งหมด 730 ค่าสังเกต ในงานวิจัยนี้ได้แบ่งข้อมูลออกเป็น 2 ส่วน ด้วยอัตราส่วน 70:30 คือ ข้อมูลฝึกฝน และข้อมูลทดสอบ ตามลำดับ ซึ่งได้จากการเก็บข้อมูลจากหน่วยงานต่าง ๆ ที่เกี่ยวข้อง ดังนี้

1.ข้อมูลราคาน้ำมันดีเซล เก็บรวมรวมจากเว็บไซต์ของบริษัท บางจาก คอร์ปอเรชั่น จำกัด (มหาชน) ซึ่งมีข้อมูลราคาน้ำมันชนิดต่าง ๆ ย้อนหลัง

2.ข้อมูลราคาน้ำมันดิบ เก็บรวบรวมจากเว็บไซต์ของบริษัท ฟิวชั่น มีเดีย จำกัด ซึ่งให้บริการข้อมูลเกี่ยวกับตลาดการเงินแบบเรียลไทม์ และราคาซื้อขายล่าสุด

3.ข้อมูลปริมาณการใช้น้ำมันดีเซล เก็บรวบรวมจากเว็บไซต์ของกรมเชื้อเพลิงธรรมชาติ กระทรวงพลังงาน

4.ข้อมูลอัตราแลกเปลี่ยน เก็บรวบรวมจากเว็บไซต์ของธนาคารแห่งประเทศไทย

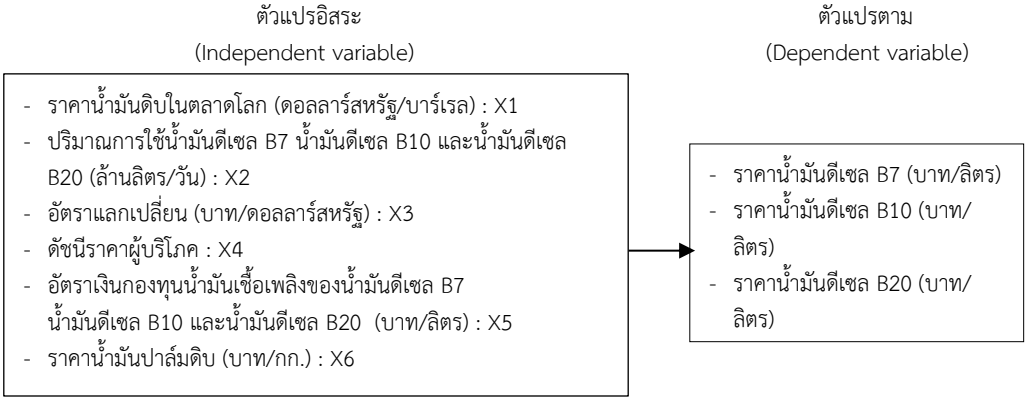
5.ข้อมูลดัชนีราคาผู้บริโภค เก็บรวบรวมจากเว็บไซต์ของกองดัชนีเศรษฐกิจการค้า สำนักงานนโยบายและยุทธศาสตร์การค้า กระทรวงพาณิชย์

6.ข้อมูลอัตราเงินกองทุนน้ำมันเชื้อเพลิง เก็บรวบรวมจากเว็บไซต์ของสำนักงานกองทุนน้ำมันเชื้อเพลิง (สกนช.)

7.ข้อมูลราคาน้ำมันปาล์มดิบ เก็บรวบรวมจากเว็บไซต์ของกรมการค้าภายใน กระทรวงพาณิชย์

2.3 กรอบแนวคิดในการวิจัย

กรอบแนวคิดในการวิจัยเป็นการกำหนดขอบเขตงานวิจัย โดยมีการกำหนดตัวแปรอิสระซึ่งเป็นปัจจัยที่ส่งผลต่อตัวแปรตามหรือผลที่ต้องการทราบ ดังรูปที่ 2



รูปที่ 2. กรอบแนวคิดในการวิจัย

2.4 การถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ

การถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ เป็นวิธีการทางสถิติที่ใช้หาค่าประมาณพารามิเตอร์สัมประสิทธิ์การถดถอยวิธีหนึ่งที่มีแนวคิดแบบ Non-Bayesian โดยมีจุดมุ่งหมายหลักเพื่อประมาณค่าสัมประสิทธิ์การถดถอยที่ใช้ได้กับข้อมูลที่มีมิติสูงและลดปัญหาการเกิดพหุสัมพันธ์ได้

โดยค่าประมาณพารามิเตอร์สัมประสิทธิ์การถดถอยจะหาได้จากการหาค่า β ที่ทำให้ฟังก์ชันเป้าหมาย (Objective function) มีค่าต่ำที่สุด โดยจะมีส่วนที่แตกต่างกันคือ มี $P_\lambda(\beta)$ เพิ่มขึ้นมา ซึ่งเรียกว่าฟังก์ชันนี้ว่า ฟังก์ชันการลงโทษ (Penalty function) ที่ค่าพารามิเตอร์ λ โดยทั่วไปจะใช้วิธี Cross-validation ในการหาค่า λ ที่เหมาะสมสำหรับข้อมูลที่ต้องการวิเคราะห์ โดยค่า λ จะมีค่ามากกว่าหรือเท่ากับศูนย์ ซึ่งฟังก์ชันการลงโทษนั้นมีหลายรูปแบบที่แตกต่างกันไป [8]

2.4.1 การวิเคราะห์การถดถอยแบบริดจ์

Hoerl และ Kennard [4] เสนอการวิเคราะห์การถดถอยแบบริดจ์เพื่อแก้ปัญหาการเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ ซึ่งจะลดค่าความคลาดเคลื่อนกำลังสองของค่าเฉลี่ยให้มีค่าต่ำลง โดยค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีการถดถอยแบบริดจ์สามารถหาได้ จากสมการที่ (1)

$$\hat{\beta}_{Ridge} = \arg\min_{\beta} \left[\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{P} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{P} \beta_j^2 \right] \; ; \; \lambda > 0 \qquad (1)$$

โดยมีฟังก์ชันการลงโทษแบบแอลทูนอร์ม (L2-Norm) ดังสมการที่ (2)

$$P_\lambda(\beta) = \lambda \sum_{j=1}^{P} \beta_j^2 \qquad (2)$$

เมื่อ λ คือ พารามิเตอร์ปรับแต่ง (Tuning parameter) ควบคุมขนาดการหดตัว (Shrinkage) ของตัวประมาณ $\hat{\beta}_{ridge}$ [9]

2.4.2 การวิเคราะห์การถดถอยแบบลาสโซ่

Tibshirani [5] เสนอการวิเคราะห์การถดถอยแบบลาสโซ่ ใช้สำหรับการประมาณค่าสัมประสิทธิ์การถดถอยและการคัดเลือกตัวแปรอิสระเข้าสู่ตัวแบบในคราวเดียวกัน โดยค่าสัมประสิทธิ์การถดถอยพหุคูณของการวิเคราะห์การถดถอยแบบลาสโซ่จะให้ประสิทธิภาพสูงสุดในการพยากรณ์ สำหรับกรณีที่จำนวนตัวแปรอิสระมีความสัมพันธ์กับตัวแปรตามจำนวนไม่มากนักในตัวแบบ ซึ่งจะทำให้ผลบวกระหว่างผลรวมความคลาดเคลื่อนกำลังสองและผลรวมสัมบูรณ์ของค่าสัมประสิทธิ์ถ่วงน้ำหนักให้มีค่าต่ำสุด หากค่าถ่วงน้ำหนักมาก ๆ จะทำให้สัมประสิทธิ์นั้นหดตัวเหลือเท่ากับศูนย์หรือหายไป กล่าวคือวิธีนี้จะทำให้ค่าสัมประสิทธิ์ส่วนใหญ่เป็นศูนย์และค่าสัมประสิทธิ์บางส่วนไม่เท่ากับศูนย์ (Sparse estimator) โดยสามารถประมาณค่าสัมประสิทธิ์การถดถอยของการวิเคราะห์การถดถอยแบบลาสโซ่ ดังสมการที่ (3)

$$\hat{\beta}_{Lasso} = \arg\min_{\beta} \left[\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{P} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{P} |\beta_j| \right] \; ; \; \lambda > 0 \qquad (3)$$

โดยมีฟังก์ชันการลงโทษแบบแอลวันนอร์ม (L1-Norm) ดังสมการที่ (4)

$$P_\lambda(\beta) = \lambda \sum_{j=1}^{p} |\beta_j| \tag{4}$$

เมื่อ λ คือ พารามิเตอร์ปรับแต่งควบคุมขนาดการหดตัวของตัวประมาณ $\hat{\beta}_{Lasso}$

2.4.3 การวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง

เนื่องจากการวิเคราะห์การถดถอยแบบลาสโซ่ยังมีความเอนเอียงในการคัดเลือกตัวประมาณสัมประสิทธิ์การถดถอย ดังนั้น Zou [13] จึงเสนอการวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง โดยเพิ่มค่าถ่วงน้ำหนัก (Weight) อีกหนึ่งพารามิเตอร์เข้ามาในวิธีการวิเคราะห์การถดถอยแบบลาสโซ่ ซึ่งการให้ค่าถ่วงน้ำหนักกับพารามิเตอร์แต่ละตัวแตกต่างกันในฟังก์ชันการลงโทษ $P_\lambda(\beta)$ โดยการกำหนดค่าถ่วงน้ำหนักให้มีค่าสูงสำหรับค่าสัมประสิทธิ์ที่มีค่าน้อย และกำหนดค่าถ่วงน้ำหนักให้มีค่าน้อยสำหรับค่าสัมประสิทธิ์ที่มีค่าสูง เพื่อลดความไม่คงเส้นคงวาที่ทำให้เกิดความเอนเอียงในการประมาณค่าสัมประสิทธิ์การถดถอยแบบลาสโซ่ การใช้ฟังก์ชันการลงโทษยังมีคุณสมบัติอีก 1 ข้อคือ เมื่อขนาดตัวอย่างมีจำนวนมากพอ การวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุงจะมีความสามารถในการเลือกตัวแปรเสมือนกับว่าทราบตัวแบบที่แท้จริง (True model) ซึ่งคุณสมบัตินี้ไม่มีในการวิเคราะห์การถดถอยแบบลาสโซ่ โดยสามารถประมาณค่าสัมประสิทธิ์การถดถอยของการวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง ดังสมการที่ (5)

$$\hat{\beta}_{ALasso} = \arg\min_{\beta} \left[\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \widehat{w}_j |\beta_j| \right] \tag{5}$$

โดยมีฟังก์ชันการลงโทษแบบแอลวันนอร์ม (L1-Norm) ดังสมการที่ (6)

$$P_\lambda(\beta) = \lambda \sum_{j=1}^{p} \widehat{w}_j |\beta_j| \tag{6}$$

เมื่อ $\widehat{w}_j = \frac{1}{|\hat{\beta}_{Lasso}|^\gamma} \; ; j = 1, 2, \ldots, p, \gamma > 0$

โดยทั่วไปกำหนดให้ $\gamma = 1$ และ $\lambda \sum_{j=1}^{p} \widehat{w}_j |\beta_j|$ คือฟังก์ชันการลงโทษของการวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง โดยเป็นฟังก์ชันการลงโทษของการวิเคราะห์การถดถอยแบบลาสโซ่ที่เพิ่มค่าถ่วงน้ำหนักเข้ามา

2.4.4 การวิเคราะห์การถดถอยแบบอิลาสติคเน็ต

Zou และ Hastie [6] เสนอการวิเคราะห์การถดถอยแบบอิลาสติคเน็ต โดยเป็นการรวมกันระหว่างการวิเคราะห์การถดถอยแบบริดจ์และการวิเคราะห์การถดถอยแบบลาสโซ่ จึงสามารถคัดเลือกตัวแปรอิสระและประมาณค่าไปพร้อมกัน วิธีนี้เหมาะสำหรับการวิเคราะห์ข้อมูลที่ตัวแปรอิสระที่มีความสัมพันธ์กันสูงและมีจำนวนตัวแปรอิสระมากกว่าขนาดตัวอย่างมาก ๆ โดยค่าสัมประสิทธิ์การถดถอยพหุคูณของการวิเคราะห์การถดถอยแบบอิลาสติคเน็ต จะอยู่ในรูปผลบวกของผลรวมสัมบูรณ์ของค่าสัมประสิทธิ์ถ่วงน้ำหนักและผลรวมกำลังสองของค่าสัมประสิทธิ์ถ่วงน้ำหนักให้มีค่าต่ำสุด โดยสามารถประมาณค่าสัมประสิทธิ์การถดถอยของการวิเคราะห์การถดถอยแบบอิลาสติคเน็ต ดังสมการที่ (7)

$$\hat{\beta}_{Elastic} = \arg\min_{\beta} \left[\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{P} \beta_j X_{ij} \right)^2 + \lambda_1 \sum_{j=1}^{P} |\beta_j| + \lambda_2 \sum_{j=1}^{P} \beta_j^2 \right] ; 0 < \lambda_1 + \lambda_2 < 0 \quad (7)$$

โดยมีฟังก์ชันการลงโทษ ดังสมการที่ (8)

$$P_\lambda(\beta) = \lambda_1 \sum_{j=1}^{P} |\beta_j| + \lambda_2 \sum_{j=1}^{P} \beta_j^2 \quad (8)$$

ฟังก์ชันการลงโทษของการวิเคราะห์การถดถอยแบบอิลาสติคเน็ต เป็นการรวมกันของฟังก์ชันการลงโทษของการวิเคราะห์การถดถอยแบบริดจ์และการวิเคราะห์การถดถอยแบบลาสโซ่ เมื่อ $\lambda_1 = 0$ การวิเคราะห์การถดถอยแบบอิลาสติคเน็ตจะเป็นการวิเคราะห์การถดถอยแบบริดจ์อย่างง่าย (Simple Ridge Regression) ซึ่งฟังก์ชันการลงโทษของการวิเคราะห์การถดถอยแบบอิลาสติคเน็ตจะมี λ_1 และ λ_2 คือ พารามิเตอร์ปรับแต่ง ซึ่งควบคุมขนาดการหดตัวของ $\hat{\beta}_{Elastic}$

2.5 การเรียนรู้ของเครื่อง

คือกระบวนการออกแบบโปรแกรมให้สามารถเรียนรู้และพัฒนาตัวเองได้จาก “ประสบการณ์” ซึ่งหมายถึงข้อมูลตัวอย่างที่ป้อนเข้าไป โดยหลักการของการเรียนรู้ของเครื่อง คือการนำข้อมูลตัวอย่างที่มีอยู่ที่เรียกว่า ข้อมูลฝึกฝนและเอาท์พุตมาป้อนเข้าไปให้กับคอมพิวเตอร์เพื่อสอนให้คอมพิวเตอร์เรียนรู้และเกิดการพัฒนา “ประสบการณ์” ของตัวโปรแกรม โดยการสร้างตัวแบบการเรียนรู้ให้คอมพิวเตอร์มีมันสมองและปรับตัวตามข้อมูลที่ได้รับ สามารถทำนายหรือตัดสินใจทำงานด้วยตัวเองอย่างอัตโนมัติโดยไม่จำเป็นต้องอาศัยการป้อนโปรแกรมคำสั่งประมวลผล [14]

2.5.1 วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน

เป็นเทคนิคที่ใช้วิธีการของซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) มาวิเคราะห์ความถดถอยระหว่างอินพุตเวกเตอร์ (Input vector) และตัวแปรเอาท์พุต (Output variables) ซึ่งนำมาใช้กับการพยากรณ์อนุกรมเวลา โดยเปลี่ยนการจำแนกคลาสด้วย SVM เป็นการทำนายค่าด้วย SVR เป้าหมาย คือ ต้องการค้นหาความสัมพันธ์เชิงเส้นระหว่างอินพุตเวกเตอร์ในมิติและตัวแปรเอาท์พุต สมการถดถอยของ SVR คล้ายกับสมการไฮเปอร์เพลน (Hyperplane) ของ SVM โดยแสดงดังสมการที่ (9)

$$f(x) = \langle w, \varphi(x) \rangle + b \tag{9}$$

เมื่อ w คือ เวกเตอร์ค่าน้ำหนัก
b คือ ค่าคงที่

ในการหาค่า w, b สามารถทำได้โดยแก้ไขปัญหาการหาค่าที่เหมาะสมภายใต้ข้อจำกัด ดังสมการที่ (10)

$$\underset{w,b,u_i,v_i}{\text{minimize}} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(u_i + v_i) \tag{10}$$

ภายใต้เงื่อนไข
$$\begin{aligned} y_i - \langle w, \varphi(x_i) \rangle - b \leq \varepsilon + u_i, \quad & i = 1,2,\dots,n, \\ \langle w, \varphi(x_i) \rangle + b - y_i \leq \varepsilon + v_i, \quad & i = 1,2,\dots,n, \\ u_i, v_i \geq 0 \quad , \quad & i = 1,2,\dots,n \end{aligned}$$

เมื่อ ε แทนพารามิเตอร์ที่กำหนดขนาดของบริเวณค่าความคลาดเคลื่อนที่ยอมรับได้ โดยค่าความคลาดเคลื่อนที่ตกอยู่ภายในบริเวณนี้จะไม่ถูกนำไปคิดในฟังก์ชันสูญเสีย u, v แทนตัวแปรหย่อน (Slack variable) ซึ่งเป็นค่าที่ยอมให้บางจุดข้อมูลมีค่าความคลาดเคลื่อนมากกว่าค่า ε ที่กำหนดได้ดังแสดงในสมการข้อจำกัด

2.5.2 วิธีป่าสุ่ม

อัลกอริทึมในการหาตัวแบบด้วยวิธีป่าสุ่มเป็นอัลกอริทึมประเภทหนึ่งของวิธีต้นไม้ตัดสินใจที่มีลักษณะแบบไม่ตัดแต่งกิ่ง (Unpruned) หรือต้นไม้ถดถอย (Regression trees) ซึ่งจะสร้างตัวแบบจากการนำข้อมูลฝึกฝนที่ถูกสุ่มเลือกเป็นตัวอย่างข้อมูล และทดสอบตัวแบบด้วยข้อมูลทดสอบ ซึ่งเป็นตัวอย่างที่ไม่ถูกสุ่มเลือกมาสร้างตัวแบบ สำหรับวิธีการนี้จะเรียกว่า Bagging และจะเรียกข้อมูลส่วนนี้ว่า Out-of-Bag (OOB) ผลลัพธ์ที่ได้อย่างอิสระจากต้นไม้ตัดสินใจแต่ละต้นถูกนำมาคิด

หลักการทำวิธีป่าสุ่มเริ่มจากการสุ่มตัวอย่างชุดใหม่จากจำนวนข้อมูลทั้งหมด (Bootstrapping) โดยใช้วิธีสุ่มแบบแทนที่ให้ได้ออกมาชุดจำนวนที่ต้องการ โดยมีลักษณะไม่เหมือนกันมาสร้างตัวแบบต้นไม้ตัดสินใจ (Decision tree) สำหรับแต่ละชุดข้อมูลเพื่อทำนายชุดข้อมูลใหม่โดยการสร้างต้นไม้ตัดสินใจ โดยนิยามเอนโทรปี (Entropy) หรือปริมาณที่บอกถึงความไม่เป็นระเบียบของข้อมูลในเซตของตัวอย่าง S คือ $E(S)$ ดังสมการที่ (11)

$$E(S) = -\sum_{j=1}^{n} Ps(j) \log_2 Ps(j) \tag{11}$$

เมื่อ S คือ ตัวอย่างที่ประกอบด้วยชุดของตัวแปรต้นและตัวแปรตาม
$Ps(j)$ คือ อัตราส่วนของ S ที่ตัวแปรตามหรือผลลัพธ์มีค่า j

หลังจากนั้นทำการหาผลลัพธ์จากแต่ละตัวแบบ เช่น การหาค่าเฉลี่ย (Mean) ของค่าทำนาย การหาค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) สำหรับการวิเคราะห์การถดถอย (Regression analysis) [15]

2.6 การวัดประสิทธิภาพของตัวแบบ

ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

เป็นการวัดค่าความคลาดเคลื่อนแบบมาตรฐานที่นิยมใช้กันอย่างแพร่หลาย โดยค่าที่ได้ยิ่งน้อยแสดงถึงตัวแบบที่ได้มีความแม่นยำมาก โดยคำนวณได้จากการหารากที่สอง (Squared root) ของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งค่าคลาดเคลื่อนคือผลต่างระหว่างค่าสังเกตและค่าพยากรณ์ที่ได้จากตัวแบบ ดังนั้นสามารถหาค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยได้ ดังสมการที่ (12)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (12)$$

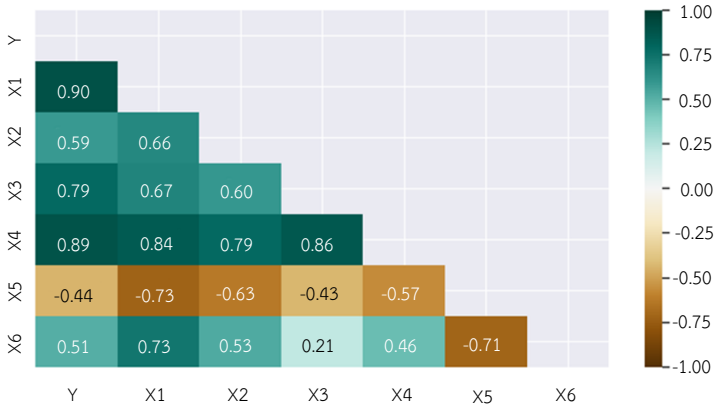
เมื่อ y_i คือ ค่าของข้อมูลที่เกิดขึ้นจริง หรือค่าสังเกต

$\hat{y}_i$ คือ ค่าของข้อมูลที่ได้จากการพยากรณ์

n คือ จำนวนข้อมูลทั้งหมด

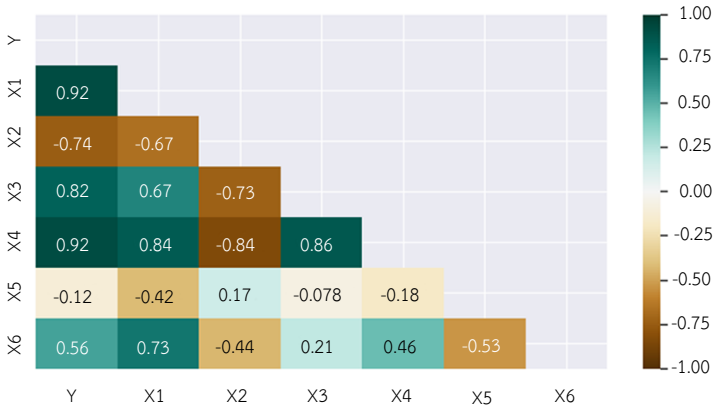
3. ผลการวิจัยและอภิปรายผล

การจากศึกษาหาความสัมพันธ์ระหว่างตัวแปร พบว่าปัจจัยที่นำมาศึกษาทั้ง 6 ตัวแปร มีความสัมพันธ์กับราคาน้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20 ดังรูปที่ 3 – 5



รูปที่ 3. Correlation Heatmap ระหว่างตัวแปรต่าง ๆ ในการพยากรณ์ราคาน้ำมันดีเซล B7

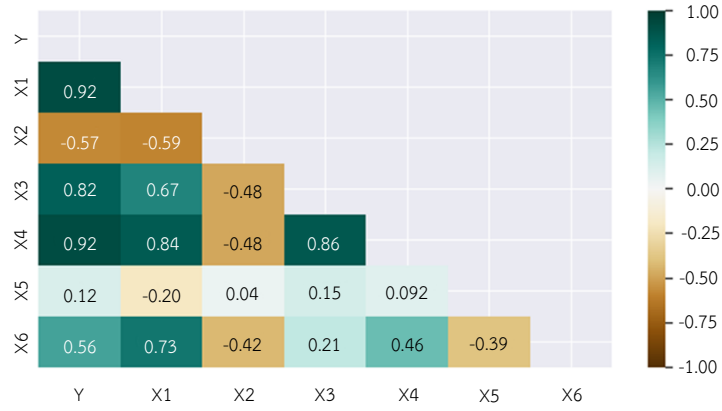
จากรูปที่ 3 พบว่าตัวแปร X5 หรืออัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B7 มีความสัมพันธ์กับตัวแปรอื่น ๆ ในทิศทางตรงกันข้าม รวมถึงตัวแปรราคาน้ำมันดีเซล B7 ด้วย ซึ่งหมายความว่า ถ้าอัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B7 เพิ่มขึ้นจะส่งผลต่อราคาน้ำมันดีเซล B10 ให้มีราคาลดลง



รูปที่ 4. Correlation Heatmap ระหว่างตัวแปรต่าง ๆ ในการพยากรณ์ราคาน้ำมันดีเซล B10

จากรูปที่ 4 พบว่าตัวแปร X2 หรือปริมาณการใช้น้ำมันดีเซล B10 และตัวแปร X5 หรืออัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B10 มีความสัมพันธ์ในทิศทางตรงกันข้ามกับตัวแปร Y หรือราคาน้ำมันดีเซล B10 ซึ่งหมายความว่า ถ้าปริมาณการใช้น้ำมันดีเซล B10 เพิ่มขึ้นจะส่งผลต่อราคาน้ำมันดีเซล B10 ให้มีราคาลดลง ทำนองเดียวกันคือถ้าอัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B10 เพิ่มขึ้นจะส่งผลต่อราคาน้ำมันดีเซล B10 ให้มีราคาลดลง

และพบว่าตัวแปร X2 หรือปริมาณการใช้น้ำมันดีเซล B10 มีความสัมพันธ์กับตัวแปร X5 หรืออัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B10 ในทิศทางเดียวกัน นั่นคือ ถ้ามีการปริมาณการใช้น้ำมันดีเซลเพิ่มขึ้นจะทำให้อัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซลเพิ่มขึ้นเช่นกัน



รูปที่ 5. Correlation Heatmap ระหว่างตัวแปรต่าง ๆ ในการพยากรณ์ราคาน้ำมันดีเซล B20

จากรูปที่ 5 พบว่าตัวแปร X2 หรือปริมาณการใช้น้ำมันดีเซล B20 มีความสัมพันธ์กับตัวแปรอื่น ๆ ในทิศทางตรงกันข้าม รวมถึงตัวแปรราคาน้ำมันดีเซล B20 ด้วย ซึ่งหมายความว่า ถ้าอัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B20 เพิ่มขึ้นจะส่งผลต่อราคาน้ำมันดีเซล B20 ให้มีราคาลดลง

3.1 ผลการวัดประสิทธิภาพของตัวแบบในการพยากรณ์ราคาน้ำมันดีเซล B7

การเปรียบเทียบค่า RMSE ของวิธีต่าง ๆ ดังตารางที่ 1 พบว่าการหาตัวแบบด้วยวิธีป่าสุ่ม ได้ค่า RMSE ต่ำที่สุด จึงสรุปได้ว่า การหาตัวแบบด้วยวิธีป่าสุ่มเป็นวิธีที่เหมาะสมที่สุดในการหาตัวแบบในการพยากรณ์ราคาน้ำมันดีเซล B7

ตารางที่ 1. ค่า RMSE สำหรับแต่ละตัวแบบที่ใช้ในการพยากรณ์ราคาน้ำมันดีเซล B7

ตัวแบบ	RMSE
การถดถอยแบบขั้นตอน	1.2457
การวิเคราะห์การถดถอยแบบริดจ์	1.2456
การวิเคราะห์การถดถอยแบบลาสโซ่	1.2451
การวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง	1.2358
การวิเคราะห์การถดถอยแบบอิลาสติคเน็ต	1.2453
วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน	0.4340
วิธีป่าสุ่ม	**0.3790**

3.2 ผลการวัดประสิทธิภาพของตัวแบบในการพยากรณ์ราคาน้ำมันดีเซล B10

การเปรียบเทียบค่า RMSE ของวิธีต่าง ๆ ดังตารางที่ 2 พบว่าการหาตัวแบบด้วยวิธีป่าสุ่ม ได้ค่า RMSE ต่ำที่สุด จึงสรุปได้ว่าการหาตัวแบบด้วยวิธีป่าสุ่มเป็นวิธีที่เหมาะสมที่สุดในการหาตัวแบบในการพยากรณ์ราคาน้ำมันดีเซล B10

ตารางที่ 2. ค่า RMSE สำหรับแต่ละตัวแบบที่ใช้ในการพยากรณ์ราคาน้ำมันดีเซล B10

ตัวแบบ	RMSE
การถดถอยแบบขั้นตอน	1.2283
การวิเคราะห์การถดถอยแบบริดจ์	1.2283
การวิเคราะห์การถดถอยแบบลาสโซ่	1.2276
การวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง	1.2189
การวิเคราะห์การถดถอยแบบอิลาสติคเน็ต	1.2280
วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน	0.4267
วิธีป่าสุ่ม	**0.3833**

3.3 ผลการวัดประสิทธิภาพของตัวแบบในการพยากรณ์ราคาน้ำมันดีเซล B20

การเปรียบเทียบค่า RMSE ของวิธีต่าง ๆ ดังตารางที่ 3 พบว่าการหาตัวแบบด้วยวิธีป่าสุ่ม ได้ค่า RMSE ต่ำที่สุด จึงสรุปได้ว่าการหาตัวแบบด้วยวิธีป่าสุ่มเป็นวิธีที่เหมาะสมที่สุดในการหาตัวแบบในการพยากรณ์ราคาน้ำมันดีเซล B20

ตารางที่ 3. ค่า RMSE สำหรับแต่ละตัวแบบที่ใช้ในการพยากรณ์ราคาน้ำมันดีเซล B20

ตัวแบบ	RMSE
การถดถอยแบบขั้นตอน	1.2326
การวิเคราะห์การถดถอยแบบริดจ์	1.2328
การวิเคราะห์การถดถอยแบบลาสโซ่	1.2319
การวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง	1.2322
การวิเคราะห์การถดถอยแบบอิลาสติคเน็ต	1.2323
วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน	0.4214
วิธีป่าสุ่ม	**0.3539**

3.4 อภิปรายผล

จากการศึกษาพบว่าตัวแปรอิสระที่อยู่ในตัวแบบที่ใช้ในการพยากรณ์ราคาน้ำมันดีเซลด้วยวิธีการวิเคราะห์การถดถอยแบบลาสโซ่และการวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง ได้แก่ ตัวแปรราคาน้ำมันดิบในตลาดโลก อัตราแลกเปลี่ยน ดัชนีราคาผู้บริโภค อัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล และราคาน้ำมันปาล์มดิบ ซึ่งตัวแปรปริมาณการใช้น้ำมันดีเซลเป็นตัวแปรที่ถูกคัดเลือกออกไปจากตัวแบบ

ตัวแปรที่อยู่ในตัวแบบที่ใช้ในการพยากรณ์ราคาน้ำมันดีเซลด้วยวิธีการวิเคราะห์การถดถอยแบบริดจ์และการวิเคราะห์การถดถอยแบบอิลาสติคเน็ต ได้แก่ ตัวแปรราคาน้ำมันดิบในตลาดโลก ปริมาณการใช้น้ำมันดีเซล อัตราแลกเปลี่ยน ดัชนีราคาผู้บริโภค อัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล และราคาน้ำมันปาล์มดิบ โดยตัวแปรราคาน้ำมันดิบในตลาดโลก อัตราแลกเปลี่ยน อัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล สอดคล้องกับงานวิจัยของฐิติรัตน์ แพทย์มงคล [10] และวิลาสินี หีบแก้ว [11] แต่ตัวแปรอัตราแลกเปลี่ยน ดัชนีราคาผู้บริโภค และอัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล ไม่สอดคล้องกับงานวิจัยของณัฐพินท์ เดชขุน [12] มารุต จำลอง และศักดิ์ชาย นาคนก [3] เนื่องจากในงานวิจัยของณัฐพินท์ เดชขุน [12] ได้ผลสรุปว่าตัวแปรอัตราแลกเปลี่ยน ดัชนีราคาผู้บริโภค และอัตราเงินกองทุนน้ำมันเชื้อเพลิงไม่มีผลต่อราคาน้ำมันสำเร็จรูปดีเซล และงานวิจัยของมารุต จำลอง และศักดิ์ชาย นาคนก [3] ได้ผลสรุปว่าตัวแปรอัตราแลกเปลี่ยนไม่มีผลต่อราคาน้ำมันสำเร็จรูปดีเซล

ทั้งนี้จะพบว่าตัวแบบที่ได้จากวิธีการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษให้ค่า RMSE ที่มากกว่า เนื่องจากจำนวนตัวแปรอิสระใช้ในการสร้างตัวแบบมีจำนวนที่น้อยกว่า เพราะตัวแปรอิสระบางตัวถูกคัดเลือกออกไปจากตัวแบบ ในขณะที่ตัวแบบที่ได้จากวิธีการเรียนรู้ของเครื่องเกิดจากการคำนวณของตัวแปรอิสระทั้งหมด 6 ตัว ซึ่งสอดคล้องกับผลการวิจัยที่ว่าค่า RMSE จากตัวแบบที่ได้จากวิธีการเรียนรู้ของเครื่องให้ค่าน้อยกว่าวิธีการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ

4. สรุปผลการวิจัย

จากการที่ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพตัวแบบระหว่างการถดถอยที่ปรับด้วยฟังก์ชันการลงโทษ และการเรียนรู้ของเครื่อง สำหรับราคาน้ำมันดีเซลในประเทศไทย พบว่ามีปัจจัยที่เกี่ยวข้องดังนี้ ราคาน้ำมันดิบในตลาดโลก ปริมาณการใช้น้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20 อัตราแลกเปลี่ยน ดัชนีราคาผู้บริโภค อัตราเงินกองทุนน้ำมันเชื้อเพลิงของน้ำมันดีเซล B7 น้ำมันดีเซล B10 และน้ำมันดีเซล B20 และราคาน้ำมันปาล์มดิบ และทำการสร้างตัวแบบด้วยการคัดเลือกแบบขั้นตอน การวิเคราะห์การถดถอยแบบริดจ์ การวิเคราะห์การถดถอยแบบลาสโซ่ การวิเคราะห์การถดถอยลาสโซ่แบบปรับปรุง การวิเคราะห์การถดถอยแบบอิลาสติคเน็ต วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน และวิธีป่าสุ่ม เพื่อหาตัวแบบที่เหมาะสมในการพยากรณ์ราคาน้ำมันดีเซล B7 ราคาน้ำมันดีเซล B10 และราคาน้ำมันดีเซล B20 โดยทำการเปรียบเทียบตัวแบบด้วยค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย จากผลการเปรียบเทียบพบว่า วิธีป่าสุ่มเป็นวิธีที่เหมาะสมที่สุดในการหาตัวแบบที่เหมาะสมในการพยากรณ์ราคาน้ำมันดีเซล B7 ราคาน้ำมันดีเซล B10 และราคาน้ำมันดีเซล B20 โดยมีค่า RMSE เท่ากับ 0.379, 0.3833 และ 0.3539 ตามลำดับ

เอกสารอ้างอิง (References)

[1] กรมธุรกิจพลังงาน กระทรวงพลังงาน. 2565. ข่าวสถานการณ์การใช้น้ำมันเชื้อเพลิงของปี 2564 (มกราคม – ธันวาคม), แหล่งข้อมูล : https://www.doeb.go.th/news_activity/2498.pdf ค้นเมื่อวันที่ 4 กุมภาพันธ์ 2566.

[2] กรมธุรกิจพลังงาน กระทรวงพลังงาน. 2566. ข่าวสถานการณ์การใช้น้ำมันเชื้อเพลิงของปี 2565 (มกราคม – ธันวาคม), แหล่งข้อมูล : https://www.doeb.go.th/news_activity/2515.pdf ค้นเมื่อวันที่ 4 กุมภาพันธ์ 2566.

[3] มารุต จำลอง และศักดิ์ชาย นาคนก. 2561. การศึกษาปัจจัยที่ส่งผลกระทบต่อราคาน้ำมันสำเร็จรูปดีเซลและการพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย. Graduate School MINI-Conference 2018 การประชุมวิชาการเสนอผลงานวิจัยระดับชาติ, ครั้งที่ 2, มหาวิทยาลัยราชภัฏสวนสุนันทา กรุงเทพมหานคร, 818-828. [Marut Jamlong and Sakchai Nakanok. 2018. A study of factors affecting and Forecasting the price of diesel oil in Thailand, Graduate School Conference 2018 2nd National Conference, Suan Sunandha Rajabhat University, Bangkok, 818-828. (in Thai)]

[4] Hoerl, A.E. and Kennard, R.W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics,* 12(1), 55-67, https://doi.org/10.2307/1267351.

[5] Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology,* 58(1), 267-288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[6] Zou, H. and Hastie, T. 2005. Regularization and Variable Selection Via the Elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology,* 67(2), 301-320, https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[7] วรางคณา วัชรเสถียร. 2561. การเปรียบเทียบประสิทธิภาพของวิธีการวิเคราะห์การถดถอยแบบพีนอลไลซ์ในตัวแบบการถดถอยลอจิสติกภายใต้ข้อมูลที่มีมิติสูงแบบบางเบาและตัวแปรทำนายมีความสัมพันธ์กันสูง. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์), สาขาวิชาคณิตศาสตร์และสถิต, คณะวิทยาศาสตร์และเทคโนโลยี, มหาวิทยาลัยธรรมศาสตร์. [Warangkhana Watcharasatian. 2018. Performance comparison of penalized regression method in logistic regression for high-dimensional sparse data with multicollinearity. Master of Science (Applied Statistics), Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University (in Thai)]

[8] เบญจมาส รุ่งศรานนท์ และอัชฌา อระวีพร. 2563. การเปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ของการวิเคราะห์การถดถอยที่ปรับด้วยฟังก์ชันการลงโทษภายใต้ข้อมูลที่มีมิติสูง. *วารสารวิทยาศาสตร์และเทคโนโลยี*, 28(8), 1346-1358, https://doi.org/10.14456/tstj.2020.108. [Benjamas Rungsaranon and Autcha Araveeporn. 2020. Comparing Methods of Parameter Estimation with Penalized Regression Analysis under High-Dimensional Data. *Science and Technology Journal,* 28(8), 1346-1358, https://doi.org/10.14456/tstj.2020.108. (in Thai)]

[9] Yasin, H., Caraka, R.E., Tarno, and Hoyyi, A. 2016. Prediction of Crude Oil Prices Using Support Vector Regression (SVR) with grid search - cross validation algorithm. *Global Journal of Pure Applied Mathematics*, 12(4), 3009-3020.

[10] ฐิติรัตน์ แพทย์มงคล. 2552. ปัจจัยที่มีผลตอการปรับตัวของราคาน้ำมันดีเซลในประเทศไทย. วิทยานิพนธ์ปริญญาเศรษฐศาสตร์มหาบัณฑิต, สาขาวิชาเศรษฐศาสตร์ธุรกิจ, บัณฑิตวิทยาลัย มหาวิทยาลัยหอการค้าไทย. [Thitirat Pheatmongkhon. 2009. Factors Affecting the Adjustment of Diesel Prices in Thailand. Master of Economics. Department of Business Economics, Graduate School, University of the Thai Chamber of Commerce. (in Thai)]

[11] วิลาสินี หีบแก้ว. 2550. การกำหนดราคาและปัจจัยที่มีผลกระทบต่อการเปลี่ยนแปลงราคาน้ำมันดีเซล. วิทยานิพนธ์ปริญญาเศรษฐศาสตร์มหาบัณฑิต, สาขาวิชาเศรษฐศาสตร์ธุรกิจ มหาวิทยาลัยเกษตรศาสตร์. [Wilasinee Heepkaew. 2007. Price Determination and Factors Affecting Diesel Oil Price Changes. Master of Economics. Department of Business Economics, Kasetsart University. (in Thai)]

[12] ณัฐพินท์ เดชขุน. 2553. การศึกษาปัจจัยทางเศรษฐกิจที่มีผลต่อราคาน้ำมันสำเร็จรูปดีเซล. บริหารธุรกิจมหาบัณฑิต, สาขาวิชาการเงิน, มหาวิทยาลัยหอการค้าไทย. [Natthaphin Dechkhun. 2010. Study of Economic Factors Affecting Diesel Finished Oil Prices. Master of

Business Administration, Department of Finance, University of the Thai Chamber of Commerce. (in Thai)]

[13] Zou, H. 2006. The Adaptive Lasso and Its Oracle Properties. *Journal of the. American. Statistical. Journal of the American Statistical Association*, 101(476), 1418-1429, https://doi.org/10.1198/016214506000000735.

[14] อรพิน ประวัติบริสุทธิ์. 2564. Python สำหรับงาน Data Science Data Visualization และ Machine Learning. พิมพ์ครั้งที่ 1, บริษัทโปรวิชั่นจำกัด, กรุงเทพฯ. [Orapin Prawatborisud. 2021. Python for Data Science, Data Visualization and Machine Learning. 1st ed., Provision Co., Ltd., Bangkok. (in Thai)]

[15] สห ธิติถามวัต. 2562. การพัฒนารูปแบบการแนะนำงานสำหรับองค์กรและผู้สมัคร ตามทักษะการเรียนรู้ด้วยเทคนิคป่าแบบสุ่ม. วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต, สาขาเทคโนโลยีสารสนเทศ, สถาบันเทคโนโลยีไทย-ญี่ปุ่น.