การเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลสำหรับพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย

Comparative Efficient Evaluation of Data Mining for Forecasting Diesel Prices in Thailand

ธมลวรรณ พูลเจริญ[1], ภัชษร สิ่วสำแดงเดช[2] และอนุพงศ์ สุขประเสริฐ[3*]
[1,2,3*]คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม

Thamonwan Phuncharern[1], Patsorn Siewsamdangdet[2] and Anupong Sukprasert[3*]
[1,2,3*]Mahasarakham Business School, Mahasarakham University

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย ด้วยเทคนิคการทำเหมืองข้อมูล 4 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้น เทคนิคโครงข่ายประสาทเทียม เทคนิคเพื่อนบ้านใกล้ที่สุด และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยใช้ข้อมูลปัจจัยที่ส่งกระทบต่อราคาน้ำมันดีเซล ตั้งแต่เดือนมกราคม ปี พ.ศ.2560 ถึงเดือนธันวาคม ปี พ.ศ.2565 รวมทั้งสิ้น 72 เดือน ผลการวิจัยพบว่า เทคนิคที่มีความหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ราคาน้ำมันดีเซล คือเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 0.320 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง เท่ากับ 0.566 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 0.182 ค่าความคลาดเคลื่อนกำลังสองเท่ากับ 0.483 ซึ่งมีค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับเทคนิคอื่นๆ และให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ สูงที่สุดเมื่อเทียบกับเทคนิคอื่นๆ ซึ่งมีค่าเท่ากับ 98.50%

คำสำคัญ: 1) ราคาน้ำมันดีเซล 2) เทคนิคการถดถอยเชิงเส้น 3) เทคนิคโครงข่ายประสาทเทียม 4) เทคนิคเพื่อนบ้านใกล้ที่สุด 5) เทคนิคซัพพอร์ตเวกเตอร์แมชชีน

Abstract

This research aims to compare the performance of models for diesel price forecasting in Thailand using four data mining techniques: linear regression, artificial neural network, k-nearest neighbors, and support vector machine. It utilizes data on factors affecting diesel oil prices from January 2017 to December 2022, totaling 72 months. The research findings indicate that the most suitable technique for building a diesel price forecasting model is Support Vector Machines, with Mean Squared Error (MSE) of 0.320, Root Mean Squared Error (RMSE) of 0.566, Mean Absolute Error (MAE) of 0.182, and Squared Error (SE) of 0.483, demonstrating the least error compared to other techniques and the highest R-squared of 98.50%.

Keywords: 1) Diesel Price 2) Linear Regression 3) Neural Network 4) k-Nearest Neighbors 5) Support Vector Machines

[1]นิสิตปริญญาโท หลักสูตรบัญชีมหาบัณฑิต (Graduate Student, Master of Accountancy Program)
[2,3*] อาจารย์ประจำคณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม E-mail: anupong.s@acc.msu.ac.th

บทนำ

น้ำมันเป็นปัจจัยสำคัญในการขับเคลื่อนเศรษฐกิจทั้งในระดับท้องถิ่นและระดับชาติ เนื่องจากมีบทบาทสำคัญในการดำรงชีวิตประจำวันของประชาชนทุกคน และมีผลต่อการพัฒนาเศรษฐกิจในรูปแบบต่างๆ เมื่อมีการเจริญเติบโตทางเศรษฐกิจสูงขึ้น ความต้องการใช้น้ำมันเชื้อเพลิงในชีวิตประจำวันและในการพัฒนาเศรษฐกิจจะเพิ่มขึ้นตามไปด้วย หากสภาพการผลิตน้ำมันไม่พร้อมทันต่อการต้องการที่เพิ่มขึ้นนี้ จะทำให้ราคาน้ำมันสูงขึ้น ในขณะที่เมื่ออัตราการเจริญเติบโตของเศรษฐกิจอยู่ในระดับต่ำ ซึ่งนำมาซึ่งสภาพการผลิตน้ำมันที่มากกว่าความต้องการของตลาด ราคาน้ำมันอาจลดลง แต่ในปัจจุบันพบว่าราคาน้ำมันในตลาดเศรษฐกิจของประเทศไทยมีความแปรผันไม่คงที่ ซึ่งส่งผลกระทบต่อด้านเศรษฐกิจทั้งในภาคราชการและภาคเอกชน โดยส่งผลให้ประชาชนที่มีรายได้ต่ำต้องปรับเปลี่ยนการดำรงชีวิตใหม่เพื่อให้สามารถรับมือกับความแพงของน้ำมันได้ นอกจากนี้ รัฐบาลยังต้องมีการเตรียมการรับมือกับปัญหาอัตราการเจริญเติบโตของเศรษฐกิจที่ขยายตัวสูงขึ้นด้วยการวางแผนนโยบายที่เหมาะสมในการบรรเทาผลกระทบที่เกิดขึ้นจากความแปรผันของราคาน้ำมันในตลาดและการสนับสนุนประชาชนที่มีรายได้ต่ำในการดำรงชีวิตให้มีความมั่นคงและเหมาะสม (มยุรี เกศากิจ, 2562)

ดังนั้นผู้วิจัยได้สนใจศึกษาการสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย โดยใช้เทคนิคการประมาณค่า 4 เทคนิคมาทำการสร้างตัวแบบการพยากรณ์ที่ได้มาจากการศึกษางานวิจัยที่เกี่ยวข้องกับงานวิจัยในครั้งนี้ ซึ่งได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาท (Neural Network) เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors: k-NN) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) เพื่อหาตัวแบบที่เหมาะสมสำหรับการนำไปใช้ในการสร้างตัวแบบการพยากรณ์ราคาน้ำมันดีเซล เพื่อช่วยให้หน่วยงานที่เกี่ยวข้องที่ยังต้องพึ่งพาการใช้น้ำมันดีเซลในการประกอบธุรกิจพร้อมรับมือกันการผันผวนของราคาน้ำมันดีเซลที่เกิดขึ้นอย่างต่อเนื่อง

การทบทวนวรรณกรรม

การเรียนรู้ของเครื่อง (Machine Learning) เป็นกระบวนการที่ระบบคอมพิวเตอร์สามารถเรียนรู้และปรับปรุงการปฏิบัติตามข้อมูลหรือประสบการณ์ที่มีอยู่ โดยไม่ต้องโปรแกรมใหม่หรือมีการคำนวณทางคณิตศาสตร์โดยตรงจากผู้ใช้หรือผู้พัฒนาต่าง ๆ โดยแบ่งชุดข้อมูลออกเป็นสองส่วนหลัก คือ ชุดข้อมูลการเรียนรู้ (Training Set) ซึ่งเป็นข้อมูลที่ถูกนำไปฝึกเครื่อง และชุดข้อมูลทดสอบ (Test Set) ซึ่งใช้สำหรับการประเมินประสิทธิภาพของระบบในการทำนาย ซึ่งการเรียนรู้ของเครื่องมีหลากหลายแบบ โดยขึ้นอยู่กับลักษณะของข้อมูล โดยเฉพาะการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งเป็นการเรียนรู้แบบการจำแนกประเภทของข้อมูล (Classification) ในการเรียนรู้แบบมีผู้สอน โมเดลจะถูกฝึกด้วยชุดข้อมูลที่มีข้อมูลตัวอย่าง (Examples) และคำตอบที่ถูกต้อง (Labels) เพื่อให้เครื่องเรียนรู้จำแนกประเภทของข้อมูลและสร้างโมเดลที่สามารถทำนายคำตอบสำหรับข้อมูลใหม่ (สมาคมโปรแกรมเมอร์ไทย, 2565)

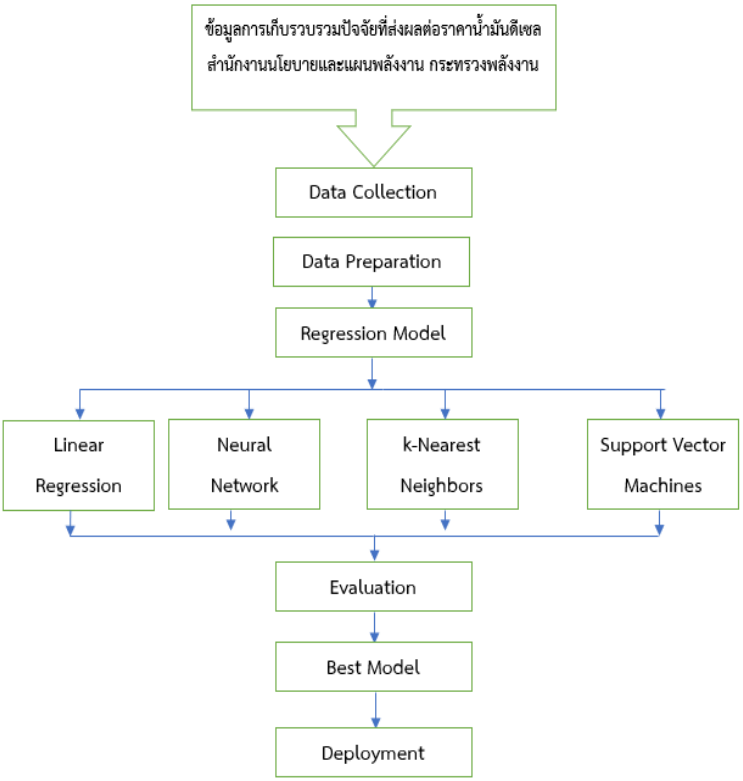
การวิเคราะห์แบบ Linear Regression เป็นหนึ่งในเทคนิคหลักในการเรียนรู้ของเครื่อง (Machine Learning) ที่ใช้ในการวิเคราะห์และทำนายความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent variables) กับตัวแปรตาม (Dependent variable) ที่มีลักษณะเป็นตัวแปรต่อเนื่อง โดยมีความสมมาตรเป็นเส้นตรง (Linear relationship) ระหว่างตัวแปรทั้งสอง ซึ่งมักจะถูกใช้งานในหลายด้านต่างๆ อาทิเช่น การทำนายราคาของสินค้า การวิเคราะห์ข้อมูลทางการเงิน เป็นต้น (สถาบันนวัตกรรมและธรรมาภิบาลข้อมูล, 2565)

มารุต จำลอง และคณะ (2561) ได้ศึกษางานวิจัย เรื่อง การศึกษาปัจจัยที่ส่งผลกระทบต่อราคาน้ำมันสำเร็จรูปดีเซลและการพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย ผลการศึกษาพบว่า อัตราภาษีสรรพสามิต (บาทต่อลิตร) (TAX) กองทุนน้ำมันเชื้อเพลิง (FUND) ดัชนีราคาผู้บริโภค (CPI) และ ราคาน้ำมันดีเซลในตลาดสิงคโปร์ (ดอลลาร์สหรัฐฯ/บาร์เรล) (HSD_S) มีความสัมพันธ์ในทิศทางเดียวกันกับราคาน้ำมันดีเซลในประเทศไทยซึ่งเป็นไปตามสมมติฐานที่กำหนดไว้อย่างมีนัยสำคัญทางสถิติ ณ ระดับความเชื่อมั่นร้อยละ 95

พิศาล สามัง (2563) ได้ศึกษาวิจัยเรื่อง การพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่อง ผลการศึกษาพบว่าตัวแบบผสม SARIMAX-ANN-SVR-RANSAC-REG ตัวแบบผสม SARIMAX-ANN-RANSAC-REG และตัวแบบผสม SARIMAX-SVR มีความแม่นยำสูงและเหมาะสมที่สุดในการพยากรณ์ปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ตามลำดับ และมีค่า MAPE เท่ากับ 2.2785% 1.9966% และ 3.5055% ตามลำดับ จากงานวิจัยที่เกี่ยวข้อง ผู้วิจัยได้ประยุกต์ใช้ตัวแปรที่เกี่ยวข้องในการพยากรณ์ราคาน้ำมันดีเซล คือ ราคา ณ โรงกลั่น (EX-REFIN) ภาษีสรรพสามิต (TAX) กองทุนน้ำมันเชื้อเพลิง (Oil Fund) กองทุนเพื่อส่งเสริมการอนุรักษ์พลังงาน (CONSV. FUND) ค่าการตลาด (WS&VAT) ราคาน้ำมันดิบดูไบ (Dubai Crude Oil) และอัตราแลกเปลี่ยนดอลลาร์ (Rale Ex) เพื่อใช้ในการสร้างตัวแบบสำหรับการพยากรณ์ราคาน้ำมันดีเซล เพื่อเป็นแนวทางในการทำการศึกษาและพยากรณ์ทางด้านราคาน้ำมันดีเซลให้กับหน่วยงานหรือบุคคลที่เกี่ยวข้องทั้งภาครัฐและเอกชนที่ยังต้องพึ่งพาการใช้น้ำมันดีเซลในการประกอบธุรกิจ

วิธีดำเนินการ

งานวิจัยนี้มีกรอบแนวคิดสำคัญเพื่อสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของเทคนิคเหมืองข้อมูลที่ใช้สำหรับการสร้างตัวแบบสำหรับการพยากรณ์ราคาน้ำมันดีเซล เพื่อช่วยให้หน่วยงานที่เกี่ยวข้องเตรียมพร้อมรับมือกันการผันผวนของราคาน้ำมันดีเซล โดยมีขั้นตอนการดำเนินงานเริ่มจากการเก็บรวบรวมข้อมูลจากสำนักงานนโยบายและแผนพลังงาน กระทรวงพลังงาน และนำมาทำความสะอาดข้อมูล เพื่อเตรียมความพร้อมของข้อมูลก่อนที่จะมีการวิเคราะห์ จากนั้นทำการสร้างตัวแบบสำหรับการพยากรณ์ราคาน้ำมันดีเซล โดยใช้เทคนิคการทำเหมืองข้อมูลและเปรียบเทียบประสิทธิภาพของตัวแบบเพื่อหาตัวแบบที่มีความเหมาะสมสำหรับนำไปใช้พยากรณ์ต่อไป ดังแสดงในภาพที่ 1



ภาพที่ 1 ขั้นตอนวิธีการดำเนินงานวิจัย

ข้อมูลที่ผู้วิจัยได้นำชุดข้อมูลมาวิเคราะห์ตามมาตรฐานในการทำเหมืองข้อมูล (Cross Standard Process for Data Mining : CRISP-DM) สำหรับการสร้างตัวแบบ การเพิ่มประสิทธิภาพของตัวแบบและเปรียบเทียบเทคนิคเหมืองข้อมูลสำหรับการพยากรณ์ราคาน้ำมันดีเซล ซึ่งประกอบไปด้วย 6 ขั้นตอนดังนี้

1. การทำความเข้าใจปัญหา (Business Understanding) เนื่องด้วยสถานการณ์เศรษฐกิจของประเทศไทยทำให้เห็นว่าราคาน้ำมันดีเซลนั้นมีความผันผวนอย่างมากทั้งทางตรงและทางอ้อมและอีกหลายเหตุผลที่ส่งผลให้ราคาน้ำมันดีเซลเปลี่ยนแปลง ไม่ว่าจะเป็นการตรึงราคาของรัฐบาล การปรับโครงสร้างราคาน้ำมัน ราคาน้ำมันดิบลดลง หรือตามช่วงเวลาฤดูกาล ล้วนส่งผลต่อการขับเคลื่อนเศรษฐกิจของภาครัฐและการลงทุนของภาคเอกชน และจากความผันผวนของราคาน้ำมันดิบในตลาดโลกที่เป็นปัจจัยในการผลิตน้ำมันสำเร็จรูปที่ประเทศไทยมีความต้องการใช้อย่างมากโดยเฉพาะน้ำมันดีเซลทำให้ราคาน้ำมันดีเซลผันผวนตามกันและส่งผลกระทบต่อกัน จึงทำให้ผู้วิจัยมีความสนใจและต้องการศึกษาถึงปัจจัยที่ส่งผลกระทบต่อราคาน้ำมันสำเร็จรูปดีเซลและการพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย โดยคาดว่าผลการศึกษานี้จะเป็นแนวทางในการทำการศึกษาและพยากรณ์ทางด้านราคาน้ำมันดีเซลให้กับหน่วยงานหรือบุคคลที่เกี่ยวข้องทั้งภาครัฐและเอกชนที่ยังต้องพึ่งพาการใช้น้ำมันดีเซลในการประกอบธุรกิจหรือขับเคลื่อนเศรษฐกิจตลอดจนการรับมือของการผันผวนของราคาน้ำมันดีเซลที่เกิดขึ้นอย่างต่อเนื่อง

2. การทำความเข้าใจเกี่ยวกับข้อมูล (Data Understanding) ผู้วิจัยได้ดำเนินการรวบรวมข้อมูลที่ได้จากสำนักงานนโยบายและแผนพลังงาน กระทรวงพลังงาน (www.dmf.go.th) ที่ถูกจัดเก็บในรูปแบบรายเดือนตั้งแต่วันที่ 1 เดือนมกราคม 2560 ถึงวันที่ 31 เดือนธันวาคม พ.ศ. 2565 จำนวนทั้งสิ้น 72 แถว ที่ถูกจัดเก็บไว้ในรูปแบบไฟล์ Excel สำหรับข้อมูลที่นำใช้ในการวิเคราะห์นี้ คือปัจจัยที่ส่งผลต่อราคาน้ำมันดีเซล จำนวน 8 ตัวแปร ประกอบไปด้วย วัน/เดือน/ปี (DATE), ราคา ณ โรงกลั่น (EX-REFIN) ภาษีสรรพสามิต (TAX) กองทุนน้ำมันเชื้อเพลิง (Oil Fund) กองทุนเพื่อส่งเสริมการอนุรักษ์พลังงาน (CONSV. FUND) ค่าการตลาด (WS&VAT) ราคาน้ำมันดิบดูไบ (Dubai Crude Oil) และอัตราแลกเปลี่ยนดอลลาร์ (Rale Ex) ดังแสดงในตารางที่ 1

ตารางที่ 1 ข้อมูลที่ใช้ในการสร้างตัวแบบพยากรณ์

ลำดับ	ตัวแปร	คำอธิบาย	ประเภทข้อมูล
1	DATE	ด้วย วัน/เดือน/ปี	integer
2	EX-REFIN	ราคา ณ โรงกลั่น	Real
3	TAX	ภาษีสรรพสามิต	Real
4	Oil Fund	กองทุนน้ำมันเชื้อเพลิง	Real
5	CONSV. FUND	กองทุนเพื่อส่งเสริมการอนุรักษ์พลังงาน	Real
6	WS&VAT	ค่าการตลาด	Real
7	Dubai Crude Oil	ราคาน้ำมันดิบดูไบ	Real
8	Rale Ex	อัตราแลกเปลี่ยนดอลลาร์	Real
9	RETAIL	ราคาน้ำมันดีเซล	Real

3. การเตรียมข้อมูล (Data Preparation) ขั้นตอนเตรียมข้อมูลเป็นขั้นตอนที่ทำให้เกิดความเชื่อมั่นในคุณภาพข้อมูลที่นำมาใช้ แสดงถึงความเชื่อมั่นของข้อมูลก่อนจะนำไปสร้างตัวแปร ซึ่งตัวแปรที่นำมาวิจัยครั้งนี้ประยุกต์จากการทบทวนวรรณกรรมที่เกี่ยวข้องที่เคยมีการศึกษาได้แก่ มารุต จำลอง และคณะ (2561), พิศาล สามัง (2563) ผู้วิจัยได้คัดเลือกปัจจัยจำนวน 7 ตัวแปร ประกอบด้วย 1) ราคา ณ โรงกลั่น 2) ภาษีสรรพสามิต 3) กองทุนน้ำมันเชื้อเพลิง 4) กองทุนเพื่อส่งเสริมการอนุรักษ์พลังงาน 5) ค่าการตลาด 6) ราคาน้ำมันดิบดูไบ และ 7) อัตราแลกเปลี่ยนดอลลาร์

3.1 การนำข้อมูลสำหรับการวิเคราะห์โดยรวบรวมข้อมูลทุติยภูมิจากสำนักงานนโยบายและแผนพลังงาน กระทรวงพลังงาน และอัตราแลกเปลี่ยนเงินตรา จากธนาคารแห่งประเทศไทย รูปแบบไฟล์ Excel และนำข้อมูลมาทำการประมวลผลในโปรแกรม RapidMiner Studio

3.2 การทำความสะอาดข้อมูลไม่พบข้อมูลสูญหาย (Missing Data) ของทุกแอตทริบิวต์ ในช่วงเวลาที่รวบรวมในการศึกษาการพยากรณ์

3.3 การกำหนดหน้าที่ให้แต่ละแอตทริบิวต์เพื่อไปใช้ในการวิเคราะห์ให้แต่ละแอตทริบิวต์ โดยกำหนดแอตทริบิวต์วันที่เก็บข้อมูล (Date) ให้ทำหน้าที่เป็นไอดี (ID) และกำหนดแอตทริบิวต์ราคาน้ำมันดีเซลให้ทำหน้าที่เป็นตัวแปรตาม (Label) ที่ใช้สำหรับการพยากรณ์ สำหรับแอตทริบิวต์ที่เหลือจะมีการกำหนดหน้าที่ให้เป็นตัวแปรอิสระ

4. การสร้างแบบจำลอง (Modeling)

ขั้นตอนนี้เป็นการสร้างแบบจำลอง ผู้วิจัยได้ใช้โปรแกรม RapidMiner Studio มาทำการสร้างตัวแบบสำหรับการพยากรณ์ราคาน้ำมันดีเซล ด้วยเทคนิคการทำเหมืองข้อมูล 4 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาท (Neural Network) เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM)

4.1 เทคนิคการถดถอยเชิงเส้น (Linear Regression) เป็นอธิบายการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ 1 ตัว กับตัวแปรตาม 1 ตัว คล้ายกับการวิเคราะห์สหสัมพันธ์ ต่างกันที่การวิเคราะห์สหสัมพันธ์ไม่ได้ระบุว่าตัวแปรตัวใดเป็นตัวแปรต้น และตัวแปรใดที่เป็นตัวแปรตาม นอกจากจะทราบความสัมพันธ์ระหว่างตัวแปรทั้งสองแล้ว ยังสามารถนำค่าของตัวแปรต้นเหตุไปทำนายหรือพยากรณ์ตัวแปรที่เป็นผลได้ โดยการคำนวณจากค่า x และ y ที่มีความสัมพันธ์กันจะใช้สมการของการถดถอยเชิงเส้นเชิงพหุ คือ

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon$$

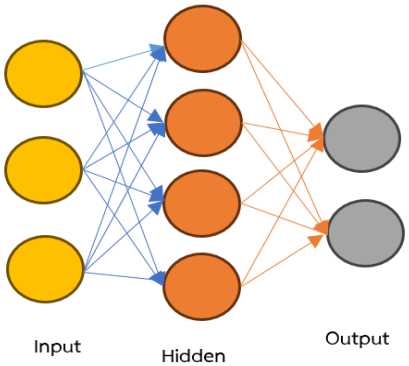
โดยที่ Y คือ ตัวแปรตาม

X คือ ตัวแปรอิสระ

$\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ คือ พารามิเตอร์

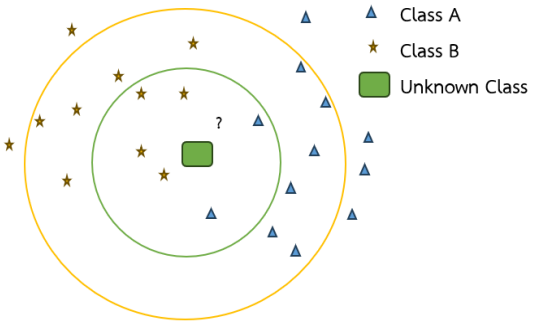
ε คือ ความคลาดเคลื่อน

4.2 เทคนิคโครงข่ายประสาทเทียม (Neural Network) เป็นโมเดลทางคณิตศาสตร์ สำหรับการประมวลผลสารสนเทศ เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้การจดจำแบบรูป (Pattaern Recognition) และการอุปมานความรู้ (Knowledge Deduction) เช่นเดียวกับความสามรถที่มีในสมองมนุษย์ ซึ่งสามารถประยุกต์ใช้กับการพยากรณ์ในหลายรูปแบบที่มีความซับซ้อนได้ดี



ภาพที่ 2 เทคนิคโครงข่ายประสาทเทียม

4.3 เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) เป็นเทคนิคที่ใช้ในการจำแนกประเภทข้อมูล โดยจะใช้หลักการเปรียบเทียบความคล้ายคลึงกันของข้อมูลที่เลือกกับข้อมูลอื่นๆ ว่ามีความคล้ายคลึงหรือใกล้เคียงกับข้อมูลใดมากที่สุด k ตัว จากนั้นจะทำการพยากรณ์ว่าคำตอบของข้อมูลที่เลือกนั้นควีเป็นคำตอบกับข้อมูลที่อยู่ใกล้ที่สุดของ k



ภาพที 3 เทคนิคเพื่อนบ้านใกล้ที่สุด

4.4 เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) เป็นการจำลองการเรียนรู้ของเครื่องแบบผู้สอนที่ใช้ในการแบ่งข้อมูลออกเป็นประเภทต่างๆ ที่ได้กำหนดไว้ซึ่งตัวแบบจำลองนั้นมีรูปแบบการทำงาน คือ สร้างตัวจำแนกประเภท (Hyperplane) ซึ่งการหาระนาบเกินอาจเจอปัญหา Overfitting ได้ ดังนั้น เพื่อที่จะเลือกระนาบเกินที่มีความเหมาะสม (Optimal Hyperplane) ที่ใช้ระยะขอบที่มากที่สุด (Maximized Margin) เป็นเกณฑ์ในการเลือกระนาบเกินให้มีความเหมาะสม

5. การทดสอบประสิทธิภาพ (Evaluation)

หลังจากที่ได้สร้างตัวแบบขึ้นมาได้แล้ว ขั้นตอนถัดมาจะต้องทำการวัดประสิทธิภาพการจำแนกประเภทข้อมูลของตัวแบบที่สร้างได้ ก่อนที่จะนำตัวแบบไปใช้ในการพยากรณ์ สำหรับงานวิจัยนี้ ผู้วิจัยได้ใช้เกณฑ์ในการวัดค่าความแม่นยำของตัวแบบซึ่งมี 5 วิธี ประกอบด้วย

5.1 ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) เป็นการวัดค่าความคลาดเคลื่อนโดยการนำค่าความคลาดเคลื่อนมายกกำลังแล้วนำไปหาค่าเฉลี่ย ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่าที่ได้มีค่าน้อยแสดงว่าตัวแบบที่ได้จะมีความแม่นยำมาก โดยมีสูตรดังนี้

$$MSE= \frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.2 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลัง (Root Mean Squared Error: RMSE) เป็นวิธีการวัดค่าความคลาดเคลื่อนแบบมาตรฐาน ซึ่งนิยมใช้กันอย่างแพร่หลาย โดยมีสมการที่ใช้ในการวัดค่าความแม่นยำจากวิธีการนี้ยิ่งค่า RMSE ที่ได้ค่าน้อยแสดงว่าตัวแบบที่ได้จะมีความแม่นยำมาก โดยมีสูตรดังนี้

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

โดยที่ n คือ จำนวนข้อมูลที่ใช้

y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.3 ค่าคลาดเคลื่อนสัมบูรณ์ (Absolute Error: AE) การนำค่าความคลาดเคลื่อนมาหาค่าสัมบูรณ์ โดยมีสูตรดังนี้

$$AE = \left|(y_t-\hat{y}_t)\right|$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.4 ค่าคลาดเคลื่อนยกกำลังสอง (Square Error: SE) คือการนำค่าความคลาดเคลื่อนมายกกำลังสอง โดยมีสูตรดังนี้

$$SE = \left|(y_t-\hat{y}_t)\right|^2$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

5.5 สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination: R2) คือ สัดส่วนที่ตัวแปรอิสระสามารถอธิบายการเปลี่ยนแปลงของตัวแปรตามได้ ดังนั้นถ้า R2 มีค่ามากแสดงว่าตัวแปรตามและตัวแปรอิสระ มีความสัมพันธ์กันมากหรือแสดงว่าตัวแปรอิสระของสมการความถดถอยนั้นสามารถอธิบายการเปลี่ยนแปลงของตัวแปรตามได้มาก โดยมีสูตรดังนี้

$$R^2 = 1 - \left(\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}\right)$$

โดยที่ n คือ จำนวนข้อมูลที่ใช้

y_t คือ ค่าจริงที่เวลา t ใด ๆ

$\hat{y}$ คือ ค่าที่ได้จากการพยากรณ์ที่เวลา t ใด

6. การนำไปใช้งาน (Deployment) ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลในครั้งนี้จะได้ตัวแบบที่เหมาะสมสำหรับการสร้างตัวแบบการพยากรณ์ราคาน้ำมันดีเซล เพื่อสนับสนุนให้หน่วยงานที่เกี่ยวข้องสามารถนำไปใช้ในการวิเคราะห์ราคาน้ำมัน และเตรียมพร้อมรับมือกับความผันผวนของราคาน้ำมันดีเซลที่จะเกิดขึ้นในอนาคต และนอกจากนี้ยังสามารถนำผลการวิจัยในครั้งนี้ไปพัฒนาต่อยอดในการสร้างระบบสารสนเทศสำหรับ

การพยากรณ์เกี่ยวกับราคาน้ำมัน โดยอาจเพิ่มปัจจัยในส่วนอื่น ที่ยังไม่ได้นำมาศึกษาในครั้งนี้ เพื่อให้มีความแม่นยำสำหรับการพยากรณ์มากยิ่งขึ้น

ผลการศึกษา

ผู้วิจัยได้ทำการสร้างตัวแบบและเปรียบเทียบค่าที่ทดสอบประสิทธิภาพของตัวแบบการพยากรณ์ราคาน้ำมันดีเซลด้วยเทคนิคการทำเหมืองข้อมูลทั้ง 4 เทคนิค ได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาทเทียม (Neural Network) เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) และทำการทดสอบประสิทธิภาพของการประมาณค่าด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error)ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error) ค่าความคลาดเคลื่อนสัมบูรณ์ (Absolute Error) ค่าความคลาดเคลื่อนกำลังสอง (Squared Error) และให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ (R^2) ซึ่งผลการวิเคราะห์ข้อมูลได้ผลแสดงดังตารางที่ 2

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพของตัวแบบการประมาณค่าข้อมูลทั้ง 4 เทคนิค

Regression model	Regression Performance				
	MSE	RMSE	AE	SE	R-Square
Neural Network	0.365	0.604	0.438	0.465	98.40%
Linear Regression	0.307	0.554	0.180	0.460	98.30%
k-Nesrest Neighbor	1.306	1.143	0.870	1.649	89.20%
Support Vector Machine*	0.320	0.566	0.182	0.483	98.50%

* คือเทคนิคที่มีความเหมาะสมสำหรับการนำไปสร้างตัวแบบการพยากรณ์น้ำมันดีเซล

จากตารางที่ 2 พบว่าเทคนิคที่มีความหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ราคาน้ำมันดีเซล คือเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) โดยให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 0.320 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง เท่ากับ 0.566 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 0.182 ค่าความคลาดเคลื่อนกำลังสองเท่ากับ 0.483 ซึ่งมีค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับเทคนิคอื่นๆ และให้ค่าสัมประสิทธิ์การตัดสินใจพหุคูณ สูงที่สุดเมื่อเทียบกับเทคนิคอื่นๆ ซึ่งมีค่าเท่ากับ 98.50% จึงมีความเหมาะสมนำตัวแบบที่ได้นี้ไปสร้างเป็นตัวแบบสำหรับการนำไปสร้างตัวแบบการพยากรณ์น้ำมันดีเซลต่อไป

สรุปและอภิปรายผล

จากการศึกษาในครั้งนี้มีวัตถุประสงค์หลักเพื่อสร้างตัวแบบและเปรียบเทียบประสิทธิภาพของตัวแบบสำหรับการพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย โดยใช้เทคนิคการประมาณค่า 4 เทคนิคมาทำการสร้างตัวแบบการพยากรณ์ที่ได้มาจากการศึกษางานวิจัยที่เกี่ยวข้องกับงานวิจัยในครั้งนี้ ซึ่งได้แก่ เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาท (Neural Network) เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) เพื่อหาตัวแบบที่เหมาะสมสำหรับการนำไปใช้ในการสร้างตัวแบบการพยากรณ์ราคาน้ำมันดีเซล ซึ่งพบว่า เทคนิคที่มีความหมาะสมที่สุดสำหรับการสร้างตัวแบบการพยากรณ์ราคาน้ำมันดีเซล คือเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines) โดยให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย เท่ากับ 0.320 ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง เท่ากับ 0.566 ค่าความคลาดเคลื่อนสัมบูรณ์ เท่ากับ 0.182 ค่าความคลาดเคลื่อนกำลังสองเท่ากับ 0.483 ซึ่งมีค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับเทคนิคอื่นๆ และให้ค่าสัมประสิทธิ์การ

ตัดสินใจพหุคูณ สูงที่สุดเมื่อเทียบกับเทคนิคอื่นๆ ซึ่งมีค่าเท่ากับ 98.50% และเทคนิคที่ใช้ในการสร้างตัวแบบพยากรณ์มีความสอดคล้องกับงานวิจัยของพิศาล สามัง (2563) ที่ได้ศึกษางานวิจัยเรื่อง การพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่อง พบว่า การใช้เทคนิคตัวแบบผสม SARIMAX-ANN-SVRRANSAC-REG ตัวแบบผสม SARIMAX-ANN-RANSAC-REG และตัวแบบผสม SARIMAX-SVR มีความแม่นยำสูงและเหมาะสมที่สุดในการพยากรณ์ปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซินและน้ำมันเครื่องบิน โดยมีค่า MAPE เท่ากับ 2.2785% 1.9966% และ 3.5055% ตามลำดับ เนื่องจากงานวิจัยดังกล่าวได้มุ่งศึกษาโดยเทคนิคตัวแบบผสม SARIMAX-ANN-SVRRANSAC-REG ตัวแบบผสม SARIMAX-ANN-RANSAC-REG และตัวแบบผสม SARIMAX-SVR และทดสอบประสิทธิภาพของการประมาณค่าด้วยค่า MAPE แต่ทางผู้วิจัยเป็นได้ใช้เทคนิคประมาณค่า 4 เทคนิค เทคนิคการถดถอยเชิงเส้น (Linear Regression) เทคนิคโครงข่ายประสาท (Neural Network) เทคนิคเพื่อนบ้านใกล้ที่สุด (k-Nearest Neighbors) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines : SVM) และทดสอบประสิทธิภาพของการประมาณค่าด้วยค่า MSE , RMSE, AE, SE และ R-Square ในการวัดค่าความแม่นยำในการพยากรณ์ ซึ่งพบว่า ตัวแปรอิสระและตัวแบบการพยากรณ์ที่ใช้ศึกษาในครั้งนี้สามารถอธิบายราคาน้ำมันดีเซลในประเทศไทย ได้ถึง 98.50% นอกนั้นเป็นอิทธิพลอื่น ๆ ที่ไม่ได้นำมาศึกษาในครั้งนี้ การศึกษาในครั้งนี้เป็นแนวทางในการทำการศึกษาและพยากรณ์ทางด้านราคาน้ำมันดีเซลให้กับหน่วยงานหรือบุคคลที่เกี่ยวข้องทั้งภาครัฐและเอกชนที่ยังต้องพึ่งพาการใช้น้ำมันดีเซลในการประกอบธุรกิจหรือขับเคลื่อนเศรษฐกิจตลอดจนการรับมือของการผันผวนของราคาน้ำมันดีเซลที่เกิดขึ้นเพื่อเตรียมความพร้อมและมาตรการในการรับมือกับราคามันดีเซลที่อาจเกิดความผันผวนในอนาคตได้

ข้อเสนอแนะ

จากการศึกษาในครั้งนี้ผู้วิจัยได้เลือกเทคนิคในการพยากรณ์ 4 เทคนิคมาเปรียบเทียบทำการพยากรณ์ซึ่งอาจจะไม่ได้เทคนิคที่ดีที่สุด หากมีการทำวิจัยในครั้งถัดไปควรเลือกใช้เทคนิคอื่นๆ มาใช้ในการเปรียบเทียบประสิทธิภาพความแม่นยำให้ได้ตัวแบบพยากรณ์ที่ดีที่สุด และยังสามารถหาปัจจัยอื่นๆที่ส่งผลต่อราคาน้ำมันดีเซล นอกเหนือจากที่ไม่ได้นำมาศึกษาในครั้งมาทำการวิเคราะห์ในครั้งถัดไป

บรรณานุกรม

กฤติกา ทิพย์คำมี, อนุพงศ์ สุขประเสริฐ. สุพัตรา กอผจญ และณัฐกานต์ ชุติมารังสรรค์ (2566). ประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องสำหรับการพยากรณ์ฝุ่นละอองขนาดเล็กในอากาศ. *วารสารวิทยาศาสตร์เทคโนโลยีและนวัตกรรม มหาวิทยาลัยกาฬสินธุ์*. 2(1), 58-74.

กิตติศักดิ์ ขำจิตร, ดารภา ใจคุ้มเก่า, วัชรพงษ์ ภูมิรัง อาทิตยา สัตนาโค และอนุพงศ์ สุขประเสริฐ (2565). ประสิทธิภาพของเทคนิคเหมืองข้อมูลสำหรับพยากรณ์การเกิดโรคหลอดเลือดในสมอง. *วารสารวิทยาการสารสนเทศและเทคโนโลยีประยุกต์*, 4(2), 87-98.

พิศาล สามัง (2563). *การพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่อง*. (วิทยานิพนธ์มหาบัณฑิต). กรุงเทพมหานคร. จุฬาลงกรณ์ มหาวิทยาลัย. สืบค้นจาก https://digital.car.chula.ac.th/chulaetd/3827.

มารุต จำลอง และศักดิ์ชาย นาคนก (2561). การศึกษาปัจจัยที่ส่งผลกระทบต่อราคาน้ำมันสำเร็จรูปดีเซลและการพยากรณ์ราคาน้ำมันดีเซลในประเทศไทย. *รายงานการประชุม Graduate School Conference*, 1(1), 818-828.

มยุรี เกษากิจ นิติมา ลักขณานุรักษ์ และไก้รุ่ง เฮงพระพรม (2562). *การเปรียบเทียบประสิทธิภาพเทคนิคการพยากรณ์สำหรับข้อมูลราคาน้ำมันด้วยเทคนิคเหมืองข้อมูล*. สืบค้น 1 พฤศจิกายน 2566. สืบค้นจาก : https://publication.-npru.ac.th/jspui/handle/123456789/711.

อนุพงศ์ สุขประเสริฐ. (2465). คู่มือการทำเหมืองข้อมูลด้วยโปรแกรม RapidMiner Studio. สาขาวิชาคอมพิวเตอร์ธุรกิจ คณะการบัญชีและการจัดการ มหาวิทยาลัยมหาสารคาม.

เชลล์ ประเทศไทย. (2565). *ปัจจัยการกำหนดราคาน้ำมัน*. สืบค้น 1 พฤศจิกายน 2566. สืบค้นจาก : https://www.shell.co.th/.

สำนักงานนโยบายและแผนพลังงาน กระทรวงพลังงาน. (2565). *ปัจจัย น้ำมันแพง*. สืบค้น 1 พฤศจิกายน 2566. สืบค้นจาก : www.eppo.go.th.

สถาบันนวัตกรรมและธรรมาภิบาล. (2565*). Linear Regression*. สืบค้น 1 พฤศจิกายน 2566. สืบค้นจาก : https://digi.data.go.th/-en/blog/linear-regression-en.

สมาคมโปรแกรมเมอร์ไทย (2565). *อะไรคือ การเรียนรู้ของเครื่อง (Machine Learning)? (ฉบับมือใหม่)*. สืบค้น 1 พฤศจิกายน 2566. สืบค้นจาก : https://www.thaipro-grammer.org.

Laung-Iem, Kornkamol and PrapitaThanarak (2021). Forecasting of biodiesel prices in Thailand using time series decompo- sition method for long term from 2017 to 2036. *International Journal of Energy Economics and Policy* 11 (4), S. 593 - 600. doi:10.32479/ijeep.10666.