

การพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่อง



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

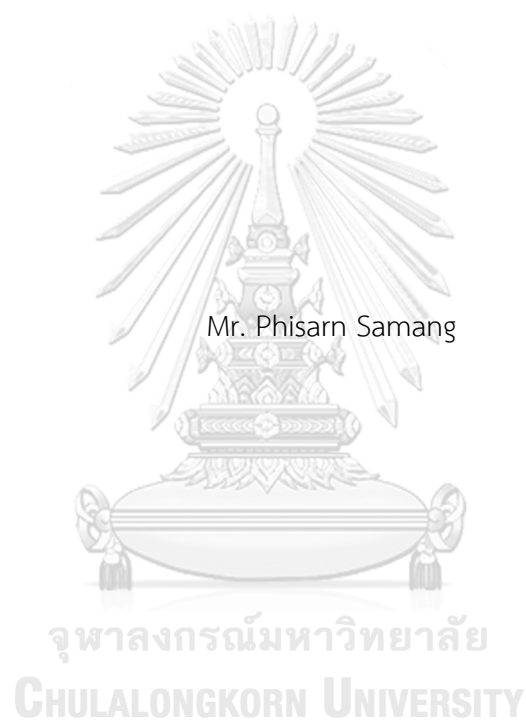
สาขาวิชาวิศวกรรมอุตสาหการ ภาควิชาวิศวกรรมอุตสาหการ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2563

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Forecasting of Thailand Major Petroleum Product Consumption using Machine  
Learning Techniques



A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Engineering in Industrial Engineering

Department of Industrial Engineering

FACULTY OF ENGINEERING

Chulalongkorn University

Academic Year 2020

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่อง
โดย	นายพิศาล สามัง
สาขาวิชา	วิศวกรรมอุตสาหการ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ดร.นันทชัย กานตานันทะ

---

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

.....	คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)	
คณะกรรมการสอบวิทยานิพนธ์	
.....	ประธานกรรมการ
(รองศาสตราจารย์ ดร.ดาริชา สุธีวงศ์)	
.....	อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ดร.นันทชัย กานตานันทะ)	
.....	กรรมการ
(รองศาสตราจารย์ ดร.นระเกณต์ พุ่มชูศรี)	
.....	กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.จันทร์ศิริ สิงห์เถื่อน)	

พิศาล สามัง : การพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยด้วย  
 เทคนิคการเรียนรู้ของเครื่อง. ( Forecasting of Thailand Major Petroleum  
 Product Consumption using Machine Learning Techniques) อ.ที่ปรึกษาหลัก :  
 ดร.นันทชัย กานตานันทะ

กิจกรรมทางเศรษฐกิจของประเทศไทยในปัจจุบันมีการใช้ประโยชน์จากการบริโภคน้ำมันสำเร็จรูปเป็นหลักในภาคการขนส่ง ภาคอุตสาหกรรม และภาคครัวเรือน การพยากรณ์การบริโภคน้ำมันสำเร็จรูปเป็นวิธีหนึ่งที่จะช่วยให้ผู้ที่เกี่ยวข้องใช้ตัดสินใจการวางแผนบริหารจัดการน้ำมันสำเร็จรูปให้เพียงพอต่อความต้องการใช้งานที่จะเกิดขึ้นในอนาคต งานวิจัยนี้นำเสนอและเปรียบเทียบตัวแบบการพยากรณ์เพื่อใช้สำหรับการพยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทย ได้แก่ น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ซึ่งเป็นน้ำมันสำเร็จรูปที่มีปริมาณการบริโภคมากที่สุด 3 อันดับแรก วัตถุประสงค์ของงานวิจัยนี้คือเพื่อศึกษาวิธีการพยากรณ์และเลือกตัวแบบที่เหมาะสมสำหรับการพยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยโดยการเปรียบเทียบผลการพยากรณ์ของตัวแบบอนุกรมเวลาเชิงสถิติ ตัวแบบการพยากรณ์เชิงสาเหตุ ตัวแบบการเรียนรู้ของเครื่อง และตัวแบบผสม ความแม่นยำของการพยากรณ์จะถูกเปรียบเทียบโดยใช้ค่าเฉลี่ยของร้อยละความคลาดเคลื่อนสัมบูรณ์ (MAPE) ข้อมูลที่ใช้ในงานวิจัยนี้เป็นข้อมูลรายไตรมาสในช่วงเวลาตั้งแต่เดือนมกราคม พ.ศ. 2536 ถึงเดือนธันวาคม พ.ศ. 2562 ตัวแบบที่ศึกษาในงานวิจัยนี้ได้แก่ Holt-Winters, SARIMA, SARIMAX, Multiple Linear Regression (MLR), RANSAC Regression, K-nearest Neighbor Algorithm (KNN), Support Vector Regression (SVR), Adaboost (ADA), Artificial Neural Network (ANN) และตัวแบบผสม 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

สาขาวิชา วิศวกรรมอุตสาหการ

ลายมือชื่อนิสิต .....

ปีการศึกษา 2563

ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

# # 6270195621 : MAJOR INDUSTRIAL ENGINEERING

KEYWORD: Forecasting, Petroleum products consumption amount, Diesel, Gasoline,  
Jet fuel, Machine Learning, Hybrid model

Phisarn Samang : Forecasting of Thailand Major Petroleum Product Consumption  
using Machine Learning Techniques. Advisor: NANTACHAI KANTANANTHA, Ph.D.

The economic activities of Thailand in the present utilize mainly petroleum product consumption in the sections of transportation, industry and household. Forecasting of petroleum product consumption amount is a method to assist related parties in making petroleum product management planning decisions to meet the demand in the future. This research presents and compares forecasting models to forecast major petroleum product consumption amounts of Thailand including diesel, gasoline, and jet fuel which are the top three of petroleum product consumption. The objective of this research is to study the forecasting methods and select the appropriate forecasting models by comparing the forecasting results from models including statistical time series forecasting models, causal forecasting models, machine learning models and hybrid models. The forecast accuracies are compared by mean absolute percentage error (MAPE). The data used in this research are the quarterly data from January 1993 to December 2019. The models studied in this research included Holt-Winters, SARIMA, SARIMAX, Multiple Linear Regression (MLR), RANSAC Regression, K-nearest Neighbor Algorithm (KNN), Support Vector Regression (SVR), Adaboost (ADA), Artificial Neural Network (ANN), and hybrid models. According to the results, the hybrid SARIMAX-ANN-SVR-RANSAC-REG model, the hybrid SARIMAX-ANN-RANSAC-REG model, and the hybrid SARIMAX-SVR model provide high accuracy and are most suitable for forecasting of diesel, gasoline, and jet fuel consumption amounts, respectively, with MAPEs of 2.2785%, 1.9966%, and 3.5055%, respectively.

Field of Study: Industrial Engineering

Student's Signature .....

Academic Year: 2020

Advisor's Signature .....

## กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบพระคุณอาจารย์ ดร. นันทชัย กานตานันทะ อาจารย์ที่ปรึกษาวิทยานิพนธ์เป็นอย่างสูงที่กรุณาสละเวลาให้คำแนะนำ ให้ความช่วยเหลือ ให้ความรู้ ตลอดจนช่วยตรวจสอบและแก้ไขจุดบกพร่องด้วยความเอาใจใส่อย่างดียิ่งตลอดการดำเนินงานวิจัย เป็นผลให้วิทยานิพนธ์ฉบับนี้สมบูรณ์และสำเร็จลุล่วงไปได้ด้วยดี

ข้าพเจ้าขอกราบขอบพระคุณรองศาสตราจารย์ ดร. ดาริชา สุธีวงศ์ ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. นระเกณท์ พุ่มชูศรี กรรมการสอบ และรองศาสตราจารย์ ดร. จันทร์ศิริ สิงห์เถื่อน กรรมการสอบภายนอกมหาวิทยาลัย ที่กรุณาชี้แนะและให้คำแนะนำแนวทางการแก้ไขข้อบกพร่องในการดำเนินการวิจัย เป็นผลให้วิทยานิพนธ์นี้ให้สมบูรณ์ยิ่งขึ้น

ข้าพเจ้าขอกราบขอบพระคุณคณะอาจารย์ ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยทุกท่าน ที่ได้ถ่ายทอดวิชาความรู้และคำแนะนำ ซึ่งสามารถช่วยให้ข้าพเจ้าทำงานวิจัยนี้สำเร็จไปได้ด้วยดี

ขอขอบพระคุณครอบครัวของข้าพเจ้าที่ให้การสนับสนุน ให้คำปรึกษา และคอยช่วยเหลือในทุก ๆ ด้าน จนสามารถทำงานวิจัยได้จนสำเร็จการศึกษา และขอขอบคุณเพื่อน ๆ ในภาควิชา ที่ได้ให้คำปรึกษา ให้กำลังใจ ให้ความช่วยเหลือตลอดมา

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

พิศาล สามัง

## สารบัญ

	หน้า
.....	ค
บทคัดย่อภาษาไทย.....	ค
.....	ง
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ฎ
สารบัญรูปภาพ.....	ฐ
บทที่ 1 .....	1
บทนำ.....	1
1.1 ที่มาและความสำคัญของปัญหา .....	1
1.2 วัตถุประสงค์ของงานวิจัย .....	5
1.3 ขอบเขตงานวิจัย .....	5
1.4 ผลที่คาดว่าจะได้รับ.....	7
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	7
บทที่ 2 .....	8
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....	8
2.1 ทฤษฎีที่เกี่ยวข้อง .....	8
2.1.1 ความหมายของการพยากรณ์.....	8
2.1.2 เทคนิคการพยากรณ์.....	8
2.1.3 ขั้นตอนพื้นฐานของการพยากรณ์.....	9

2.1.4	วิธีการพยากรณ์ด้วยวิธีอนุกรมเวลาเชิงสถิติ (Statistical Time Series Forecasting Methods) .....	10
2.1.4.1	วิธีปรับเรียบเอ็กซ์โพเนนเชียลแบบโฮลต์และวินเทอร์ (Holt-Winters Exponential Smoothing) .....	11
2.1.4.1	วิธีบอกซ์-เจนกินส์ (Box-Jenkins : ARIMA Model).....	12
2.1.5	การถดถอยเชิงเส้น (Linear Regression, LR).....	14
2.1.5.1	การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression, SLR).....	15
2.1.5.2	การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression, MLR) .....	15
2.1.6	อัลกอริทึม Random Sample Consensus (RANSAC) .....	16
2.1.7	การเรียนรู้ของเครื่อง (Machine Learning).....	16
2.1.8	ประเภทของการเรียนรู้ของเครื่อง .....	17
2.1.8.1	การเรียนรู้แบบมีผู้ฝึกสอน (Supervised Learning).....	17
2.1.8.2	การเรียนรู้แบบไม่มีผู้ฝึกสอน (Unsupervised Learning) .....	18
2.1.8.3	การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) .....	19
2.1.9	ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้สุด k ตัว (K-nearest Neighbor Algorithm, KNN) .....	19
2.1.10	ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines Algorithm, SVM) .....	20
2.1.11	โครงข่ายประสาทเทียม (Artificial Neural Network, ANN).....	22
2.1.12	Ensemble Machine Learning Based on Regression Trees.....	24
2.2	งานวิจัยที่เกี่ยวข้อง.....	26
2.2.1	การพยากรณ์น้ำมันสำเร็จรูป.....	26
2.2.2	การพยากรณ์ความต้องการด้วยเทคนิคการเรียนรู้ของเครื่อง.....	29
บทที่ 3	.....	37
วิธีการดำเนินงานวิจัย	.....	37
3.1	การเตรียมข้อมูล .....	37



3.2 การพยากรณ์ด้วยตัวแบบอนุกรมเวลาเชิงสถิติ.....	38
3.2.1 ข้อมูลที่ใช้สำหรับการพยากรณ์ด้วยตัวแบบอนุกรมเวลาเชิงสถิติ .....	38
3.2.2 การตรวจสอบและสำรวจข้อมูลเบื้องต้น .....	39
3.2.2.1 การสร้างกราฟอนุกรมเวลา .....	39
3.2.2.2 การตรวจสอบรูปแบบอนุกรมเวลาด้วยการแยกส่วนประกอบของอนุกรมเวลา .....	40
3.2.2.3 การตรวจสอบรูปแบบอนุกรมเวลาด้วยฟังก์ชันอัตสหสัมพันธ์ (ACF).....	43
3.2.3 การสร้างตัวแบบอนุกรมเวลาเชิงสถิติ .....	44
3.2.3.1 การสร้างตัวแบบ Holt-Winters.....	45
3.2.3.2 การสร้างตัวแบบ SARIMA.....	45
3.2.3.3 การสร้างตัวแบบ SARIMAX.....	46
3.3 การพยากรณ์เชิงสาเหตุ.....	46
3.3.1 การพยากรณ์ด้วยตัวแบบการถดถอยเชิงเส้นพหุคูณ (MLR).....	47
3.3.2 การพยากรณ์ด้วยตัวแบบ Random Sample Consensus Regression (RANSAC-REG) .....	48
3.4 การพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง .....	48
3.4.1 การปรับไฮเปอร์พารามิเตอร์ของตัวแบบการเรียนรู้ของเครื่อง .....	48
3.4.2 การสร้างตัวแบบ K-nearest Neighbors .....	51
3.4.3 การสร้างตัวแบบ Support Vector Regression .....	52
3.4.4 การสร้างตัวแบบ Artificial Neural Network .....	52
3.4.5 การสร้างตัวแบบ Adaboost .....	54
3.4.6 การสร้างตัวแบบผสม (Hybrid Model) .....	55
3.4.6.1 การสร้างตัวแบบผสมตามโครงสร้างการผสมแบบขนาน .....	55
3.4.6.2 การสร้างตัวแบบผสมตามโครงสร้างการผสมแบบอนุกรม .....	56

3.5 การวัดประสิทธิภาพของตัวแบบพยากรณ์ .....	57
3.6 การคัดเลือกตัวแบบการพยากรณ์ .....	58
บทที่ 4 .....	59
ผลการดำเนินการ .....	59
4.1 ตัวแบบ Holt-Winters .....	61
4.2 ตัวแบบ SARIMA.....	65
4.3 ตัวแบบการถดถอยเชิงเส้นพหุคูณ (MLR).....	74
4.4 ตัวแบบ Random Sample Consensus Regression.....	87
4.5 ตัวแบบ SARIMA กับตัวแปรอิสระภายนอก (SARIMAX).....	94
4.6 ตัวแบบ K-nearest Neighbors.....	98
4.7 ตัวแบบ Support Vector Regression.....	103
4.8 ตัวแบบ Artificial Neural Network.....	107
4.9 ตัวแบบ Adaboost.....	113
4.10 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ในกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติและ ตัวแบบการถดถอยเชิงเส้นพหุคูณ.....	118
4.11 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ในกลุ่มตัวแบบการเรียนรู้ของเครื่อง ..	122
4.12 ตัวแบบผสม (Hybrid Model).....	124
4.12.1 ตัวแบบ SARIMAX-ANN .....	124
4.12.2 ตัวแบบ SARIMAX-SVR .....	129
4.12.3 ตัวแบบ ANN-SARIMAX .....	131
4.12.4 ตัวแบบ SVR-SARIMAX .....	134
4.12.5 ตัวแบบ SARIMAX-ANN-RANSAC-REG .....	136
4.12.6 ตัวแบบ SARIMAX-SVR-RANSAC-REG .....	139
4.12.7 ตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG .....	140

4.13 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ในกลุ่มตัวแบบผสม .....	143
4.14 การคัดเลือกตัวแบบการพยากรณ์ .....	145
บทที่ 5 .....	149
สรุปผลการดำเนินการและข้อเสนอแนะ .....	149
5.1 สรุปผลการดำเนินการ.....	149
5.2 ข้อเสนอแนะ .....	154
บรรณานุกรม.....	156
ภาคผนวก.....	160
ประวัติผู้เขียน.....	170



## สารบัญตาราง

	หน้า
ตารางที่ 1 สรุปการทบทวนวรรณกรรมที่เกี่ยวข้องการพยากรณ์น้ำมันสำเร็จรูป.....	35
ตารางที่ 2 สรุปการทบทวนวรรณกรรมที่เกี่ยวข้องการพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง ...	36
ตารางที่ 3 ตัวแปรอิสระที่มีอิทธิพลต่อการบริโภคน้ำมันสำเร็จรูปจากการทบทวนวรรณกรรม.....	37
ตารางที่ 4 ค่าไฮเปอร์พารามิเตอร์ที่ทดลองใช้ในการสร้างตัวแบบ KNN.....	51
ตารางที่ 5 ค่าไฮเปอร์พารามิเตอร์ที่ทดลองใช้ในตัวแบบ Support Vector Regression .....	52
ตารางที่ 6 ค่าไฮเปอร์พารามิเตอร์ที่ทดลองใช้ในตัวแบบ Artificial Neural Network .....	53
ตารางที่ 7 ค่า Hyperparameter ที่กำหนดใช้ในตัวแบบ Adaboost .....	54
ตารางที่ 8 ค่าพารามิเตอร์ของตัวแบบ Holt-Winters ที่ได้จากการสร้างตัวแบบด้วยชุดข้อมูลฝึกสอน สำหรับการพยากรณ์ระยะกลาง.....	61
ตารางที่ 9 ค่าพารามิเตอร์ของตัวแบบ Holt-Winters ที่ได้จากการสร้างตัวแบบด้วยชุดข้อมูลฝึกสอน สำหรับการพยากรณ์ระยะสั้น จำนวน 5 ช่วงเวลา.....	63
ตารางที่ 10 ตัวแบบ SARIMA ของน้ำมันสำเร็จรูปที่ได้จากการสร้างตัวแบบด้วยชุดข้อมูลฝึกสอน สำหรับการพยากรณ์ระยะสั้น จำนวน 5 ช่วงเวลา.....	73
ตารางที่ 11 ตัวแบบ SARIMAX ของน้ำมันสำเร็จรูปที่ได้จากการสร้างตัวแบบด้วยชุดข้อมูลฝึกสอน สำหรับการพยากรณ์ระยะสั้น จำนวน 5 ช่วงเวลา.....	97
ตารางที่ 12 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ KNN .....	100
ตารางที่ 13 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ SVR.....	105
ตารางที่ 14 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ ANN.....	110
ตารางที่ 15 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ ADA .....	115
ตารางที่ 16 ค่า MAPE ของแต่ละตัวแบบการพยากรณ์ระยะสั้นและระยะกลางของน้ำมันดีเซลสำหรับ การตรวจสอบตัวแบบในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557 .....	119
ตารางที่ 17 ค่า MAPE ของแต่ละตัวแบบการพยากรณ์ระยะสั้นและระยะกลางของน้ำมันเบนซิน สำหรับการตรวจสอบตัวแบบในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557.....	120

ตารางที่ 18 ค่า MAPE ของแต่ละตัวแบบการพยากรณ์ระยะสั้นและระยะกลางของน้ำมันเครื่องบิน สำหรับการตรวจสอบตัวแบบในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557..... 121

ตารางที่ 19 ค่า MAPE ของแต่ละตัวแบบการเรียนรู้ของเครื่องสำหรับการตรวจสอบตัวแบบของ น้ำมันสำเร็จรูปกลุ่มหลัก ในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557..... 123

ตารางที่ 20 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ ANN ที่ใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอนและใช้ค่าส่วนเหลือของตัวแบบ SARIMAX เป็นตัวแปรตาม..... 127

ตารางที่ 21 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ SVR ที่ใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอนและใช้ค่าส่วนเหลือของตัวแบบ SARIMAX เป็นตัวแปรตาม..... 129

ตารางที่ 22 ค่า MAPE ของแต่ละตัวแบบการเรียนรู้ของเครื่องสำหรับการตรวจสอบตัวแบบของ น้ำมันดีเซล ในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557..... 143

ตารางที่ 23 ค่า MAPE ของแต่ละตัวแบบการเรียนรู้ของเครื่องสำหรับการตรวจสอบตัวแบบของ น้ำมันเบนซิน ในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557..... 144

ตารางที่ 24 ค่า MAPE ของแต่ละตัวแบบการเรียนรู้ของเครื่องสำหรับการตรวจสอบตัวแบบของ น้ำมันเครื่องบิน ในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557 ..... 144

ตารางที่ 25 ค่า Validation MAPE ของแต่ละตัวแบบการพยากรณ์ระยะกลางของน้ำมันดีเซล .... 146

ตารางที่ 26 ค่า Validation MAPE ของแต่ละตัวแบบการพยากรณ์ระยะกลางของน้ำมันเบนซิน . 147

ตารางที่ 27 ค่า Validation MAPE ของแต่ละตัวแบบการพยากรณ์ระยะกลางของน้ำมันเครื่องบิน ..... 148

ตารางที่ 28 ค่า Validation MAPE และ Testing MAPE ของตัวแบบที่ถูกคัดเลือก ..... 148

## สารบัญรูปร่างภาพ

	หน้า
รูปที่ 1 อัตราการเติบโตทางเศรษฐกิจและการใช้พลังงานของประเทศไทย .....	1
รูปที่ 2 ปริมาณการใช้พลังงานเชิงพาณิชย์ขั้นสุดท้ายของประเทศไทย.....	2
รูปที่ 3 ปริมาณการบริโภคน้ำมันสำเร็จรูปของประเทศไทยจำแนกตามประเภท .....	4
รูปที่ 4 ชุดข้อมูลฝึกสอนที่มีป้ายกำกับสำหรับการเรียนรู้แบบมีผู้ฝึกสอน (Géron, 2019).....	17
รูปที่ 5 ขอบเขตการตัดสินใจของการจำแนกด้วย SVM (Géron, 2019).....	20
รูปที่ 6 การเพิ่มคุณลักษณะเพื่อให้สามารถแยกกลุ่มข้อมูลด้วย SVM แบบเชิงเส้นได้ (Géron, 2019) .....	21
รูปที่ 7 Support Vector Regression (Géron, 2019).....	22
รูปที่ 8 การจำลองโครงข่ายประสาทเทียมจากแบบจำลองของเซลล์ประสาทหนึ่งเซลล์ (Géron, 2019).....	22
รูปที่ 9 โครงข่ายประสาทเทียมแบบหลายชั้น (นันทชัย, 2555) .....	23
รูปที่ 10 โครงสร้างการผสมแบบขนาน (Hajirahimi และ Khashei, 2019) .....	32
รูปที่ 11 โครงสร้างการผสมแบบอนุกรม (Hajirahimi และ Khashei, 2019).....	33
รูปที่ 12 กราฟอนุกรมเวลาของปริมาณการบริโภคน้ำมันดีเซล .....	39
รูปที่ 13 กราฟอนุกรมเวลาของปริมาณการบริโภคน้ำมันเบนซิน .....	40
รูปที่ 14 กราฟอนุกรมเวลาของปริมาณการบริโภคน้ำมันเครื่องบิน .....	40
รูปที่ 15 กราฟส่วนประกอบของอนุกรมเวลาของน้ำมันดีเซล .....	41
รูปที่ 16 กราฟส่วนประกอบของอนุกรมเวลาของน้ำมันเบนซิน .....	42
รูปที่ 17 กราฟส่วนประกอบของอนุกรมเวลาของน้ำมันเครื่องบิน .....	42
รูปที่ 18 กราฟฟังก์ชันอัตโนมัติสัมพันธ์ (ACF) ของน้ำมันดีเซล .....	43
รูปที่ 19 กราฟฟังก์ชันอัตโนมัติสัมพันธ์ (ACF) ของน้ำมันเบนซิน .....	44
รูปที่ 20 กราฟฟังก์ชันอัตโนมัติสัมพันธ์ (APF) ของน้ำมันเครื่องบิน.....	44

รูปที่ 21 การแบ่งชุดข้อมูลแบบ Time Series Split สำหรับชุดข้อมูลปริมาณน้ำมันสำเร็จรูป.....	50
รูปที่ 22 กระบวนการพยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลัก.....	59
รูปที่ 23 กระบวนการสร้างตัวแบบอนุกรมเวลาเชิงสถิติและตัวแบบการพยากรณ์เชิงสาเหตุ.....	60
รูปที่ 24 การพยากรณ์ระยะกลางด้วยตัวแบบ Holt-Winters ของน้ำมันดีเซล.....	61
รูปที่ 25 การพยากรณ์ระยะกลางด้วยตัวแบบ Holt-Winters ของน้ำมันเบนซิน.....	62
รูปที่ 26 การพยากรณ์ระยะกลางด้วยตัวแบบ Holt-Winters ของน้ำมันเครื่องบิน.....	62
รูปที่ 27 การพยากรณ์ระยะสั้นด้วยตัวแบบ Holt-Winters ของน้ำมันดีเซล.....	64
รูปที่ 28 การพยากรณ์ระยะสั้นด้วยตัวแบบ Holt-Winters ของน้ำมันเบนซิน.....	64
รูปที่ 29 การพยากรณ์ระยะสั้นด้วยตัวแบบ Holt-Winters ของน้ำมันเครื่องบิน.....	64
รูปที่ 30 การหาค่าเหมาะสมของพารามิเตอร์ในตัวแบบ SARIMA(p,d,q)(P,D,Q) <sub>s</sub> ของน้ำมันดีเซล.....	65
รูปที่ 31 การหาค่าเหมาะสมของพารามิเตอร์ในตัวแบบ SARIMA(p,d,q)(P,D,Q) <sub>s</sub> ของน้ำมันเบนซิน.....	66
รูปที่ 32 การหาค่าเหมาะสมของพารามิเตอร์ในตัวแบบ SARIMA(p,d,q)(P,D,Q) <sub>s</sub> ของน้ำมันเครื่องบิน.....	66
รูปที่ 33 ผลลัพธ์การสร้างตัวแบบ SARIMA ของน้ำมันดีเซล.....	67
รูปที่ 34 ผลลัพธ์การสร้างตัวแบบ SARIMA ของน้ำมันเบนซิน.....	67
รูปที่ 35 ผลลัพธ์การสร้างตัวแบบ SARIMA ของน้ำมันเครื่องบิน.....	68
รูปที่ 36 การตรวจสอบรูปแบบ (Diagnostic Checking) ของตัวแบบ SARIMA ของน้ำมันดีเซล.....	69
รูปที่ 37 การตรวจสอบรูปแบบ (Diagnostic Checking) ของตัวแบบ SARIMA ของน้ำมันเบนซิน.....	69
รูปที่ 38 การตรวจสอบรูปแบบ (Diagnostic Checking) ของตัวแบบ SARIMA ของน้ำมันเครื่องบิน.....	70
รูปที่ 39 การพยากรณ์ระยะกลางด้วยตัวแบบ SARIMA ของน้ำมันดีเซล.....	70
รูปที่ 40 การพยากรณ์ระยะกลางด้วยตัวแบบ SARIMA ของน้ำมันเบนซิน.....	71
รูปที่ 41 การพยากรณ์ระยะกลางด้วยตัวแบบ SARIMA ของน้ำมันเครื่องบิน.....	71
รูปที่ 42 การพยากรณ์ระยะสั้นด้วยตัวแบบ SARIMA ของน้ำมันดีเซล.....	72

รูปที่ 43 การพยากรณ์ระยะสั้นด้วยตัวแบบ SARIMA ของน้ำมันเบนซิน.....	72
รูปที่ 44 การพยากรณ์ระยะสั้นด้วยตัวแบบ SARIMA ของน้ำมันเครื่องบิน.....	72
รูปที่ 45 ผลลัพธ์การวิเคราะห์การถดถอยของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง.....	74
รูปที่ 46 ผลลัพธ์การวิเคราะห์การถดถอยของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง.....	76
รูปที่ 47 ผลลัพธ์การวิเคราะห์การถดถอยของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะกลาง.....	78
รูปที่ 48 ผลลัพธ์การทดสอบสมมติฐานการถดถอยของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง	79
รูปที่ 49 ผลลัพธ์การทดสอบสมมติฐานการถดถอยของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง .....	80
รูปที่ 50 ผลลัพธ์การทดสอบสมมติฐานการถดถอยของน้ำมันเครื่องบิน.....	80
รูปที่ 51 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง.....	81
รูปที่ 52 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง.....	81
รูปที่ 53 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะกลาง.....	81
รูปที่ 54 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะสั้น.....	86
รูปที่ 55 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้น.....	86
รูปที่ 56 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะสั้น.....	87
รูปที่ 57 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง	88
รูปที่ 58 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง .....	89
รูปที่ 59 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันเครื่องบิน สำหรับการพยากรณ์ระยะ กลาง.....	89
รูปที่ 60 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะสั้น ...	93
รูปที่ 61 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้น	94
รูปที่ 62 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะสั้น .....	94
รูปที่ 63 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง.....	95



รูปที่ 64 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง....	95
รูปที่ 65 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง....	95
รูปที่ 66 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะสั้น .....	96
รูปที่ 67 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้น .....	96
รูปที่ 68 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะสั้น ...	97
รูปที่ 69 กระบวนการสร้างตัวแบบ KNN.....	99
รูปที่ 70 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ KNN ของน้ำมันดีเซล .....	101
รูปที่ 71 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ KNN ของน้ำมันเบนซิน .....	101
รูปที่ 72 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ KNN ของน้ำมันเครื่องบิน ..	101
รูปที่ 73 การพยากรณ์ด้วยตัวแบบ KNN ของน้ำมันดีเซล .....	102
รูปที่ 74 การพยากรณ์ด้วยตัวแบบ KNN ของน้ำมันเบนซิน .....	102
รูปที่ 75 การพยากรณ์ด้วยตัวแบบ KNN ของน้ำมันเครื่องบิน .....	102
รูปที่ 76 กระบวนการสร้างตัวแบบ SVR.....	104
รูปที่ 77 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ SVR ของน้ำมันดีเซล .....	105
รูปที่ 78 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ SVR ของน้ำมันเบนซิน .....	106
รูปที่ 79 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ SVR ของน้ำมันเครื่องบิน ...	106
รูปที่ 80 การพยากรณ์ด้วยตัวแบบ SVR ของน้ำมันดีเซล.....	106
รูปที่ 81 การพยากรณ์ด้วยตัวแบบ SVR ของน้ำมันเบนซิน.....	107
รูปที่ 82 การพยากรณ์ด้วยตัวแบบ SVR ของน้ำมันเครื่องบิน .....	107
รูปที่ 83 กระบวนการสร้างตัวแบบ ANN.....	109
รูปที่ 84 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ ANN ของน้ำมันดีเซล .....	111
รูปที่ 85 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ ANN ของน้ำมันเบนซิน .....	111
รูปที่ 86 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ ANN ของน้ำมันเครื่องบิน ..	111
รูปที่ 87 การพยากรณ์ด้วยตัวแบบ ANN ของน้ำมันดีเซล.....	112

รูปที่ 88 การพยากรณ์ด้วยตัวแบบ ANN ของน้ำมันเบนซิน.....	112
รูปที่ 89 การพยากรณ์ด้วยตัวแบบ ANN ของน้ำมันเครื่องบิน.....	112
รูปที่ 90 กระบวนการสร้างตัวแบบ ADA.....	114
รูปที่ 91 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ ADA ของน้ำมันดีเซล.....	116
รูปที่ 92 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ ADA ของน้ำมันเบนซิน.....	116
รูปที่ 93 ค่า Mean_CV_MSE และ Mean_Train_MSE ของตัวแบบ ADA ของน้ำมันเครื่องบิน.....	116
รูปที่ 94 การพยากรณ์ด้วยตัวแบบ ADA ของน้ำมันดีเซล.....	117
รูปที่ 95 การพยากรณ์ด้วยตัวแบบ ADA ของน้ำมันเบนซิน.....	117
รูปที่ 96 การพยากรณ์ด้วยตัวแบบ ADA ของน้ำมันเครื่องบิน.....	117
รูปที่ 97 กระบวนการสร้างตัวแบบ SARIMAX-ANN และตัวแบบ SARIMAX-SVR.....	126
รูปที่ 98 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN ของน้ำมันดีเซล.....	128
รูปที่ 99 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN ของน้ำมันเบนซิน.....	128
รูปที่ 100 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN ของน้ำมันเครื่องบิน.....	128
รูปที่ 101 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR ของน้ำมันดีเซล.....	130
รูปที่ 102 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR ของน้ำมันเบนซิน.....	130
รูปที่ 103 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR ของน้ำมันเครื่องบิน.....	130
รูปที่ 104 กระบวนการสร้างตัวแบบ ANN-SARIMAX และตัวแบบ SVR-SARIMAX.....	132
รูปที่ 105 การพยากรณ์ด้วยตัวแบบ ANN-SARIMAX ของน้ำมันดีเซล.....	133
รูปที่ 106 การพยากรณ์ด้วยตัวแบบ ANN-SARIMAX ของน้ำมันเบนซิน.....	133
รูปที่ 107 การพยากรณ์ด้วยตัวแบบ ANN-SARIMAX ของน้ำมันเครื่องบิน.....	133
รูปที่ 108 การพยากรณ์ด้วยตัวแบบ SVR-SARIMAX ของน้ำมันดีเซล.....	134
รูปที่ 109 การพยากรณ์ด้วยตัวแบบ SVR-SARIMAX ของน้ำมันเบนซิน.....	135
รูปที่ 110 การพยากรณ์ด้วยตัวแบบ SVR-SARIMAX ของน้ำมันเครื่องบิน.....	135

รูปที่ 111 กระบวนการสร้างตัวแบบ SARIMAX-ANN- RANSAC-REG และตัวแบบ SARIMAX-SVR- RANSAC-REG ..... 137

รูปที่ 112 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันดีเซล ..... 138

รูปที่ 113 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันเบนซิน ..... 138

รูปที่ 114 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันเครื่องบิน..... 138

รูปที่ 115 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR-RANSAC-REG ของน้ำมันดีเซล ..... 139

รูปที่ 116 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR-RANSAC-REG ของน้ำมันเบนซิน ..... 139

รูปที่ 117 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR-RANSAC-REG ของน้ำมันเครื่องบิน..... 140

รูปที่ 118 กระบวนการสร้างตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ..... 141

รูปที่ 119 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันดีเซล ..... 142

รูปที่ 120 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันเบนซิน ..... 142

รูปที่ 121 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันเครื่องบิน ... 142

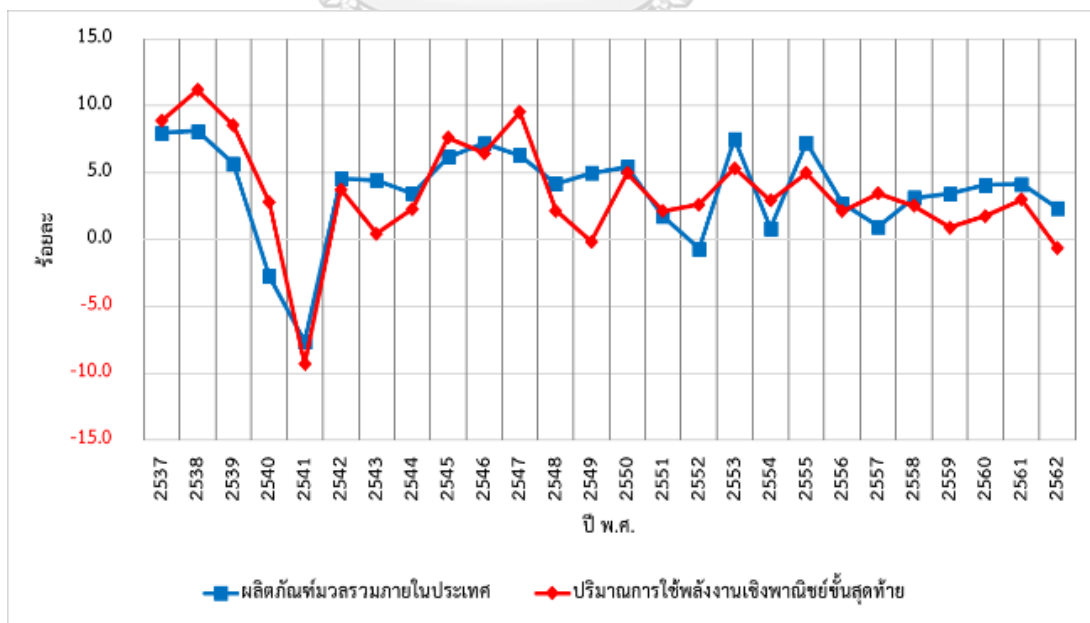


# บทที่ 1

## บทนำ

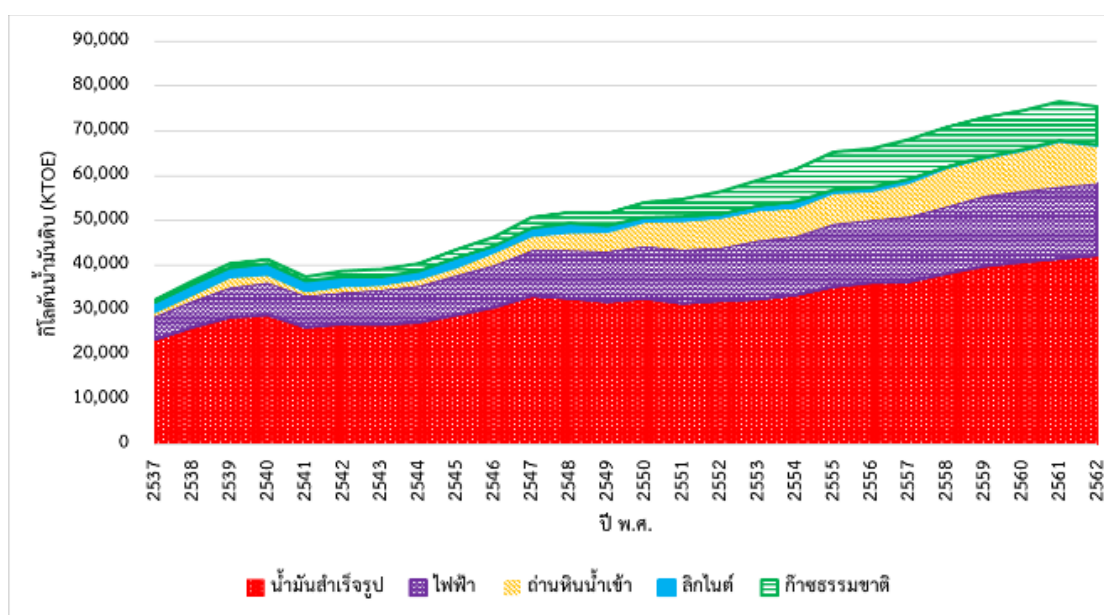
### 1.1 ที่มาและความสำคัญของปัญหา

ประเทศไทยในปัจจุบันมีการเปลี่ยนแปลงโครงสร้างทางเศรษฐกิจจากที่เน้นการผลิตในภาคการเกษตรและเป็นสังคมเกษตรกรรมเป็นหลัก ไปสู่เศรษฐกิจที่เน้นการผลิตในภาคอุตสาหกรรมและบริการ ซึ่งการเพิ่มขึ้นของกิจกรรมทางเศรษฐกิจส่งผลทำให้การใช้พลังงานในประเทศไทยเพิ่มมากขึ้น จากรูปที่ 1 จะเห็นได้ว่าอัตราการเติบโตของเศรษฐกิจไทยซึ่งใช้ตัวชี้วัดคือผลิตภัณฑ์มวลรวมภายในประเทศกับปริมาณการใช้พลังงานเชิงพาณิชย์ขั้นสุดท้ายมีแนวโน้มอัตราการเติบโตไปในทิศทางเดียวกันตั้งแต่ปี พ.ศ. 2537 จนถึงปี พ.ศ. 2562 โดยมีอัตราการเติบโตของเศรษฐกิจและการใช้พลังงานเฉลี่ยอยู่ในระดับร้อยละ 3.7 และร้อยละ 3.5 ตามลำดับ และในปี พ.ศ. 2540 เกิดวิกฤติทางเศรษฐกิจขึ้นกับประเทศไทย ซึ่งต้องเผชิญกับปัญหาวิกฤติทางการเงิน การประกาศลอยตัวค่าเงินบาทในวันที่ 2 กรกฎาคม พ.ศ. 2540 มีผลทำให้ค่าเงินบาทอ่อนตัวลงเป็นอย่างมาก เศรษฐกิจไทยชะลอตัวลงมาก โดยปี พ.ศ. 2540 และปี พ.ศ. 2541 อัตราการเติบโตของเศรษฐกิจลดลงต่ำจนมีค่าติดลบคือร้อยละ -2.8 และร้อยละ -7.6 ส่งผลให้อัตราการเติบโตของการใช้พลังงานเชิงพาณิชย์ขั้นสุดท้ายลดลงจนมีค่าติดลบที่ร้อยละ -9.3 ในปี พ.ศ. 2541 ต่อมาสถานะทางเศรษฐกิจของไทยได้เริ่มฟื้นตัวตั้งแต่ปี พ.ศ. 2542 ส่งผลให้ปริมาณการใช้พลังงานเชิงพาณิชย์ขั้นสุดท้ายเริ่มเพิ่มสูงขึ้นเรื่อยมา



รูปที่ 1 อัตราการเติบโตทางเศรษฐกิจและการใช้พลังงานของประเทศไทย

จากข้อมูลการใช้พลังงานของประเทศไทยที่รวบรวมจากสำนักงานนโยบายและแผนพลังงาน (สนพ.) กระทรวงพลังงาน ระหว่างปี พ.ศ. 2537 ถึงปี พ.ศ. 2562 พบว่าปริมาณการใช้พลังงานเชิงพาณิชย์ขั้นสุดท้ายมีแนวโน้มเพิ่มสูงขึ้นดังแสดงในรูปที่ 2 โดยสามารถแบ่งประเภทของพลังงานเชิงพาณิชย์ขั้นสุดท้ายได้เป็น น้ำมันสำเร็จรูป ไฟฟ้า ถ่านหินนำเข้า ลิกไนต์ และก๊าซธรรมชาติ (สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ, 2563)



รูปที่ 2 ปริมาณการใช้พลังงานเชิงพาณิชย์ขั้นสุดท้ายของประเทศไทย

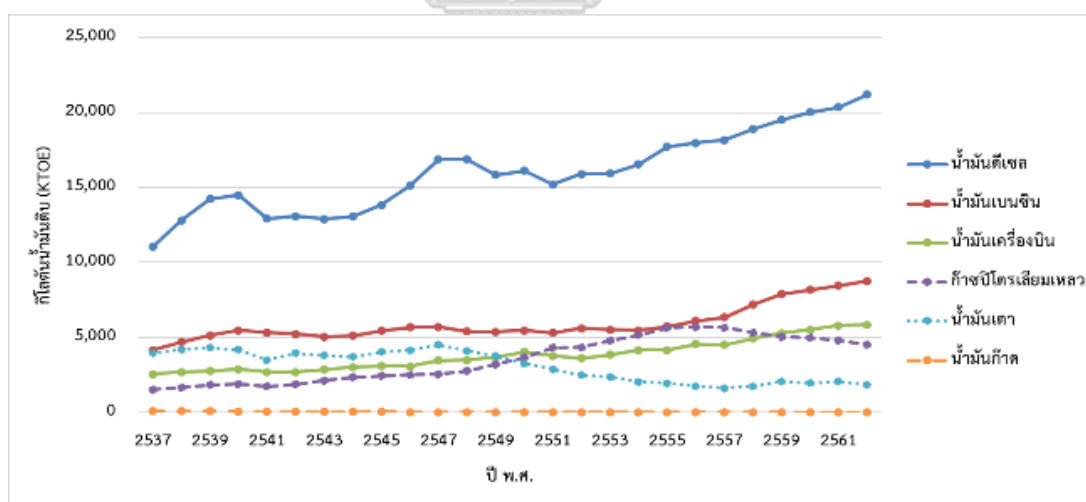
จากรูปที่ 2 จะเห็นได้ว่าสัดส่วนการใช้พลังงานเชิงพาณิชย์ขั้นสุดท้ายประเภทที่มากที่สุดคือน้ำมันสำเร็จรูป และมีแนวโน้มการใช้้ำมันสำเร็จรูปสูงขึ้นเรื่อยมาตั้งแต่ปี พ.ศ. 2537 ถึงปี พ.ศ. 2562 ซึ่งการใช้พลังงานเชิงพาณิชย์ขั้นสุดท้ายในปี พ.ศ. 2562 มีปริมาณ 89,167.1 กิโลตันน้ำมันดิบ โดยมีการบริโภคน้ำมันสำเร็จรูปมากที่สุดคิดเป็นร้อยละ 55.93 รองลงมาคือ ไฟฟ้าร้อยละ 21.43 ถ่านหินนำเข้าและลิกไนต์ร้อยละ 11.44 และสุดท้ายคือ ก๊าซธรรมชาตร้อยละ 11.21 จากปริมาณการบริโภคพลังงานประเภทน้ำมันสำเร็จรูปซึ่งมีสัดส่วนการบริโภคสูงสุดเมื่อเทียบกับพลังงานเชิงพาณิชย์ขั้นสุดท้ายประเภทอื่น ดังนั้นการเติบโตทางเศรษฐกิจของประเทศไทยและการบริโภคน้ำมันสำเร็จรูปจึงเป็นไปในทิศทางเดียวกัน โดยน้ำมันสำเร็จรูปเกิดจากการนำน้ำมันดิบมาเข้าสู่กระบวนการกลั่นโดยน้ำมันดิบได้มาจากการนำเข้าจากต่างประเทศและการผลิตขึ้นเองในประเทศจากแหล่งน้ำมันดิบต่าง ๆ ของประเทศไทย ในปี พ.ศ. 2562 ประเทศไทยมีการนำเข้าพลังงาน คิดเป็นมูลค่า 1,069.8 พันล้านบาท โดยมีการนำเข้าน้ำมันดิบมากที่สุด โดยนำเข้าน้ำมันดิบคิดเป็นร้อยละ 79 ของปริมาณน้ำมันดิบทั้งหมด ในขณะที่ผลิตขึ้นเองในประเทศร้อยละ 21 เนื่องจากลักษณะทางภูมิศาสตร์ของ

ประเทศไทยไม่ใช่แหล่งผลิตน้ำมันดิบ จึงต้องพึ่งพาการนำเข้าน้ำมันดิบจากต่างประเทศเป็นหลัก อีกทั้งราคาน้ำมันดิบมีความผันผวนมาก ไม่ว่าจะมาจากสถานการณ์ความขัดแย้งทางการค้าระหว่างประเทศ ผลกระทบของเศรษฐกิจโลก ปัญหาทางการเมืองของผู้ผลิตน้ำมันรายใหญ่ของโลก ซึ่งความผันผวนของราคาน้ำมันดิบส่งผลกระทบต่อประเทศไทยเนื่องจากการนำเข้าน้ำมันดิบจากต่างประเทศในสัดส่วนที่สูง ราคาน้ำมันดิบที่เพิ่มสูงขึ้นส่งผลให้รายจ่ายของแต่ละภาคส่วนเพิ่มขึ้น ในภาคประชาชน ค่าใช้จ่ายสำหรับน้ำมันเพิ่มขึ้นทั้งในด้านการขนส่งแบบรถยนต์ส่วนตัวและรถโดยสารสาธารณะ ต้นทุนในภาคอุตสาหกรรม ภาคบริการ หรือภาคธุรกิจที่ต้องใช้การขนส่งสินค้าหรือน้ำมันสำเร็จรูปเป็นวัตถุดิบล้วนแต่เพิ่มขึ้น

การบริโภคน้ำมันสำเร็จรูปแต่ละประเภทจะมีการนำไปใช้งานในภาคส่วนต่าง ๆ แตกต่างกันไป น้ำมันดีเซลถูกนำไปใช้ในภาคขนส่งทางบก การบริโภคน้ำมันดีเซลเพิ่มขึ้นหรือลดลงนั้นเกี่ยวข้องกับ การขยายตัวหรือชะลอตัวของภาคการส่งออก การบริการ การท่องเที่ยว รวมถึงการขนส่งสินค้า เกษตร ส่วนการบริโภคน้ำมันเบนซินรวมถึงแก๊สโซฮอล์นั้นมีความเกี่ยวข้องกับความต้องการเดินทาง ด้วยรถยนต์ หากใช้รถยนต์ในการเดินทางเพิ่มมากขึ้น ปริมาณการบริโภคน้ำมันเบนซินจะเพิ่มมากขึ้น เช่นกัน สำหรับน้ำมันเครื่องบินนั้นจะถูกนำไปใช้ในภาคขนส่งทางอากาศซึ่งมีความสัมพันธ์กับภาคการ ท่องเที่ยวและความต้องการเดินทางด้วยเครื่องบิน ในภาวะที่มีจำนวนนักท่องเที่ยวต่างชาติเดินทางเข้ามาในประเทศไทยเพิ่มขึ้นการบริโภคน้ำมันเครื่องบินจะเพิ่มขึ้นตาม ในส่วนของการบริโภคก๊าซปิโตรเลียมเหลวจะเกี่ยวข้องกับการนำไปใช้เป็นวัตถุดิบในอุตสาหกรรมปิโตรเคมี ภาคครัวเรือนใช้เป็น เชื้อเพลิงในการหุงต้ม และใช้เป็นเชื้อเพลิงในภาคขนส่ง นอกจากนี้ในภาคอุตสาหกรรมจะใช้น้ำมันเตา และก๊าซปิโตรเลียมเป็นเชื้อเพลิงสำหรับหม้อไอน้ำ เตเผาหรือเตาหลอม ส่วนน้ำมันก๊าดซึ่งมีปริมาณ การบริโภคน้อยที่สุดเมื่อเทียบกับน้ำมันสำเร็จรูปประเภทอื่นจะถูกนำไปใช้งานเป็นเชื้อเพลิงใน ภาคอุตสาหกรรมและครัวเรือน ซึ่งการบริโภคน้ำมันสำเร็จรูปในปี พ.ศ. 2562 มีปริมาณ 42,168.0 กิโลตันน้ำมันดิบ โดยแบ่งเป็นปริมาณการบริโภคน้ำมันดีเซลร้อยละ 50.26 น้ำมันเบนซินร้อยละ 20.76 น้ำมันเครื่องบินร้อยละ 13.87 ก๊าซปิโตรเลียมเหลวร้อยละ 10.70 น้ำมันเตาร้อยละ 4.40 และ น้ำมันก๊าดร้อยละ 0.01 จะเห็นได้ว่าปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน รวมคิดเป็นร้อยละ 84.89 ของปริมาณการบริโภคน้ำมันสำเร็จรูปทั้งหมด โดยน้ำมันสำเร็จรูปทั้ง 3 ชนิดดังกล่าวมีแนวโน้มการบริโภคเพิ่มขึ้น แต่น้ำมันเตามีแนวโน้มลดลงระหว่างปี พ.ศ.2537 ถึงปี พ.ศ. 2562 ดังแสดงในรูปที่ 3 (สำนักงานนโยบายและแผนพลังงาน, 2563)

จากที่กล่าวมาจะเห็นได้ว่ากิจกรรมทางเศรษฐกิจต่าง ๆ นั้นจะต้องใช้พลังงานในการขับเคลื่อน โดยเฉพาะกิจกรรมทางเศรษฐกิจที่มีการบริโภคพลังงานในภาคการขนส่ง ภาคอุตสาหกรรม และภาคครัวเรือน ซึ่งต้องใช้ประโยชน์จากการบริโภคน้ำมันสำเร็จรูปเป็นหลัก ดังนั้นการขาดแคลน น้ำมันสำเร็จรูปจะทำให้เกิดผลกระทบต่อระบบเศรษฐกิจและชีวิตความเป็นอยู่เป็นอย่างมาก จึงมี

ความจำเป็นอย่างยิ่งที่ภาครัฐจะต้องวางแผนการนำเข้าหรือการผลิตน้ำมันสำเร็จรูปให้เพียงพอต่อความต้องการที่เกิดขึ้น การพยากรณ์การบริโภคน้ำมันสำเร็จรูปเป็นวิธีหนึ่งที่ช่วยให้ผู้ที่เกี่ยวข้องใช้ตัดสินใจในการวางแผนบริหารจัดการน้ำมันสำเร็จรูปให้เพียงพอต่อความต้องการใช้งานที่จะเกิดขึ้นในอนาคต ซึ่งข้อมูลการพยากรณ์ดังกล่าวจะช่วยให้ภาครัฐสามารถวางแผนนโยบายได้อย่างเหมาะสม หากค่าการพยากรณ์น้ำมันสำเร็จรูปมีความแม่นยำต่ำจะส่งผลให้ภาครัฐวางแผนการการจัดหาน้ำมันสำเร็จรูปหรือวางแผนนโยบายด้านพลังงานผิดพลาด อาจทำให้เกิดภาวะขาดแคลนหรือมีการสำรองน้ำมันสำเร็จรูปมากเกินไป ซึ่งจะส่งผลกระทบต่อระบบเศรษฐกิจของประเทศไทยได้ ในปัจจุบันได้มีการพัฒนาเทคนิคการพยากรณ์ให้มีความแม่นยำมากขึ้น ด้วยการพยากรณ์ด้วยการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งเป็นส่วนหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) เป็นระบบที่มีความสามารถในการเลียนแบบความสามารถของมนุษย์ที่ซับซ้อนและมีความสามารถในการเรียนรู้แบบมนุษย์ได้ และได้มีการนำมาประยุกต์ใช้งานในหลาย ๆ ด้าน เช่น การแพทย์ การเงิน อุตสาหกรรมและการผลิตเกษตรกรรม ธุรกิจ รวมถึงด้านพลังงานด้วย อีกทั้งในปัจจุบันการประมวลผลด้วยคอมพิวเตอร์มีประสิทธิภาพมากขึ้น สามารถประมวลผลข้อมูลที่มีความซับซ้อนได้อย่างรวดเร็วมากขึ้น งานวิจัยนี้มุ่งศึกษาการพยากรณ์การบริโภคน้ำมันสำเร็จรูปในประเทศไทยโดยนำเทคนิคการเรียนรู้ของเครื่องเข้ามาประยุกต์ใช้เปรียบเทียบกับเทคนิคการพยากรณ์แบบดั้งเดิม



รูปที่ 3 ปริมาณการบริโภคน้ำมันสำเร็จรูปของประเทศไทยจำแนกตามประเภท

## 1.2 วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการพยากรณ์และเลือกตัวแบบที่เหมาะสมสำหรับการพยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยโดยการเปรียบเทียบผลการพยากรณ์ของตัวแบบอนุกรมเวลาเชิงสถิติ ตัวแบบการพยากรณ์เชิงสาเหตุ และตัวแบบการเรียนรู้ของเครื่อง

## 1.3 ขอบเขตงานวิจัย

1. น้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยที่เลือกนำมาศึกษาในงานวิจัยนี้คือ กลุ่มน้ำมันดีเซล กลุ่มน้ำมันเบนซิน และกลุ่มน้ำมันเครื่องบิน ซึ่งเป็นกลุ่มมีแนวโน้มการบริโภคเพิ่มขึ้นและมีปริมาณการบริโภคในปี พ.ศ.2562 คิดเป็นร้อยละ 84.89 ของน้ำมันสำเร็จรูปทั้งหมด จากการใช้หลักการพาเรโตจึงเลือกพยากรณ์น้ำมันทั้ง 3 กลุ่มดังกล่าว

2. ข้อมูลและตัวแปรที่ใช้ในการศึกษาเป็นข้อมูลทุติยภูมิอ้างอิงมาจากหน่วยงานภาครัฐดังนี้ ปริมาณการบริโภค ปริมาณการผลิต และราคาเฉลี่ยรายไตรมาสของน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทย (น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน) อ้างอิงจากสำนักงานนโยบายและแผนพลังงาน ผลิตภัณฑ์มวลรวมรายไตรมาสของประเทศไทยอ้างอิงจากสำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ จำนวนรถใหม่จดทะเบียนตามกฎหมายว่าด้วยรถยนต์และกฎหมายว่าด้วยการขนส่งทางบกของประเทศไทย อ้างอิงจากกองแผนงาน กรมการขนส่งทางบก เป็นต้น

3. ผู้วิจัยเลือกใช้ระยะเวลาของการพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยเป็นแบบการพยากรณ์ระยะสั้นและระยะกลางแบบรายไตรมาส ดังนี้

3.1 การพยากรณ์ระยะสั้นจะเป็นการพยากรณ์ล่วงหน้าเป็นระยะเวลา 1 ปี หรือจำนวน 4 ไตรมาส ซึ่งสำนักงานนโยบายและแผนพลังงาน (สนพ.) หรือหน่วยงานอื่น ๆ ที่เกี่ยวข้องสามารถนำไปใช้ประโยชน์ในการวางแผนบริหารจัดการน้ำมันสำเร็จรูปในส่วนของความต้องการให้สอดคล้องกับความสามารถในการจัดหา โดยข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักจะถูกแบ่งออกเป็น 3 ช่วงข้อมูลแบบเลื่อนตำแหน่งช่วงเวลาไปครั้งละ 1 ปี หรือจำนวน 4 ไตรมาสจนครบ 5 ตำแหน่งช่วงเวลา ดังนี้

ช่วงที่ 1 ข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสในปี พ.ศ. 2536 ถึงปี พ.ศ. 2552 จะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน (Training Data) ส่วนในปี พ.ศ. 2553 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ (Cross Validation Data) และในปี พ.ศ. 2554 จะถูกใช้เป็นชุดข้อมูลสำหรับทดสอบ (Testing Data)



ช่วงที่ 2 ข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสในปี พ.ศ. 2537 ถึงปี พ.ศ. 2553 จะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน ส่วนในปี พ.ศ. 2554 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ และในปี พ.ศ. 2555 จะถูกใช้เป็นชุดข้อมูลสำหรับทดสอบ

ช่วงที่ 3 ข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสในปี พ.ศ. 2538 ถึงปี พ.ศ. 2554 จะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน ส่วนในปี พ.ศ. 2555 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ และในปี พ.ศ. 2556 จะถูกใช้เป็นชุดข้อมูลสำหรับทดสอบ

ช่วงที่ 4 ข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสในปี พ.ศ. 2539 ถึงปี พ.ศ. 2555 จะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน ส่วนในปี พ.ศ. 2556 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ และในปี พ.ศ. 2557 จะถูกใช้เป็นชุดข้อมูลสำหรับทดสอบ

ช่วงที่ 5 ข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสในปี พ.ศ. 2540 ถึงปี พ.ศ. 2556 จะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน ส่วนในปี พ.ศ. 2557 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ และในปี พ.ศ. 2558 จะถูกใช้เป็นชุดข้อมูลสำหรับทดสอบ

จากนั้นจะนำผลการพยากรณ์ช่วงที่ใช้ชุดข้อมูลสำหรับตรวจสอบที่ครบทั้ง 5 ปีและชุดข้อมูลสำหรับทดสอบที่ครบทั้ง 5 ปีไปเปรียบเทียบความแม่นยำกับการพยากรณ์แบบระยะกลางที่จะใช้การแบ่งช่วงข้อมูลดังในข้อ 3.2

3.2 การพยากรณ์ระยะกลางจะเป็นการพยากรณ์ล่วงหน้าเป็นระยะเวลา 5 ปี หรือจำนวน 20 ไตรมาส ซึ่งสำนักงานนโยบายและแผนพลังงาน (สนพ.) หรือหน่วยงานอื่น ๆ ที่เกี่ยวข้องสามารถนำไปใช้ประโยชน์เป็นแนวทางในการวางแผนขยายกำลังการผลิตในอนาคต (Capacity Expansion Planning) และวางแผนการลงทุนโครงการได้ โดยข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักจะถูกแบ่งออกเป็น 3 ช่วงข้อมูลแบบตำแหน่งเวลาที่ดังนี้

- ข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสในปี พ.ศ. 2536 ถึงปี พ.ศ. 2552 จะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน
- ข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสในปี พ.ศ. 2553 ถึงปี พ.ศ. 2557 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ
- ข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสในปี พ.ศ. 2558 ถึงปี พ.ศ. 2562 จะถูกใช้เป็นชุดข้อมูลสำหรับทดสอบ

4. ตัวแบบพยากรณ์ในงานวิจัยนี้จะมุ่งศึกษาทั้งแบบวิธีอนุกรมเวลา ได้แก่ ตัวแบบ Holt-Winters ตัวแบบ SARIMA และตัวแบบ SARIMAX วิธีพยากรณ์เชิงสาเหตุ ได้แก่ ตัวแบบ Multiple Linear Regression และตัวแบบ RANSAC Regression เทคนิคการเรียนรู้ของเครื่อง ได้แก่ ตัวแบบ K-nearest Neighbors Algorithm ตัวแบบ Support Vector Regression ตัวแบบ Adaboost ตัวแบบ Artificial Neural Network และตัวแบบ Hybrid

5. การวัดประสิทธิผลของตัวแบบพยากรณ์เพื่อหาตัวแบบที่เหมาะสมจะใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error, MAPE)

#### 1.4 ผลที่คาดว่าจะได้รับ

1. ตัวแบบการพยากรณ์ที่เหมาะสมสำหรับใช้พยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทย

2. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยทั้งแบบวิธีอนุกรมเวลา ได้แก่ ตัวแบบ Holt-Winters ตัวแบบ SARIMA และตัวแบบ SARIMAX วิธีพยากรณ์เชิงสาเหตุ ได้แก่ ตัวแบบ Multiple Linear Regression และตัวแบบ RANSAC Regression เทคนิคการเรียนรู้ของเครื่อง เช่น ตัวแบบ K-nearest Neighbors Algorithm ตัวแบบ Support Vector Regression ตัวแบบ Adaboost ตัวแบบ Artificial Neural Network และตัวแบบ Hybrid

3. ปัจจัยที่ส่งผลต่อการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทย

#### 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ผลการพยากรณ์การบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยที่แม่นยำ ซึ่งสามารถนำไปใช้ตัดสินใจวางแผนในการบริหารจัดการน้ำมันสำเร็จรูปให้เพียงพอต่อความต้องการบริโภคในประเทศไทยได้

2. เข้าใจปัจจัยที่ส่งผลต่อการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทย ซึ่งสามารถให้ข้อมูลเชิงลึก (Data Insight) ที่เกี่ยวกับพฤติกรรมกรรมการบริโภคน้ำมันสำเร็จรูปในแต่ละกลุ่มหลักได้

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการพยากรณ์น้ำมันสำเร็จรูปในปัจจุบันมีอยู่หลากหลาย โดยการพยากรณ์นั้นจะแบ่งรูปแบบของการพยากรณ์ได้ 2 ประเภทคือการพยากรณ์ด้วยวิธีอนุกรมเวลา (Time Series Methods) และการพยากรณ์ด้วยวิธีเชิงสาเหตุ (Causal Methods)

#### 2.1 ทฤษฎีที่เกี่ยวข้อง

##### 2.1.1 ความหมายของการพยากรณ์

การพยากรณ์เป็นเรื่องที่ถูกนำไปใช้งานในสถานการณ์ที่หลากหลาย เช่น การตัดสินใจในการสร้างโรงไฟฟ้าในอีก 5 ปีข้างหน้าจะใช้ข้อมูลการพยากรณ์ความต้องการใช้ไฟฟ้าในอนาคต การจัดการตารางงานของพนักงานรับสายโทรศัพท์สำหรับการทำงานในสัปดาห์ถัดไปจะใช้ข้อมูลการพยากรณ์ปริมาณสายโทรศัพท์เข้าของสัปดาห์ถัดไป หรือการพยากรณ์ปริมาณพัสดุคงคลังเพื่อจัดการกับสต็อกพัสดุคงคลัง เป็นต้น ซึ่งการพยากรณ์เป็นการทำนายอนาคตอย่างแม่นยำที่สุดเท่าที่จะเป็นไปได้ โดยใช้สารสนเทศที่มีอยู่ ซึ่งประกอบด้วยข้อมูลในอดีตและองค์ความรู้เกี่ยวกับเหตุการณ์ในอนาคตที่อาจส่งผลกระทบต่อพยากรณ์ (Hyndman และ Athanasopoulos, 2018)

##### 2.1.2 เทคนิคการพยากรณ์

วิธีการพยากรณ์ที่เหมาะสมขึ้นอยู่กับข้อมูลที่มีอยู่ โดยจะสามารถจำแนกวิธีหรือเทคนิคการพยากรณ์ได้ 2 ประเภท ถ้าไม่มีข้อมูลหรือข้อมูลที่มีอยู่ไม่มีความสัมพันธ์กับการพยากรณ์จะใช้เทคนิคการพยากรณ์เชิงคุณภาพ (Qualitative Forecasting Techniques) เป็นเทคนิคที่ใช้ดุลพินิจและขึ้นกับประสบการณ์ของผู้เชี่ยวชาญ เช่น การพยากรณ์ความต้องการของผลิตภัณฑ์ใหม่ และอีกประเภทหนึ่งคือเทคนิคการพยากรณ์เชิงปริมาณ (Quantitative Forecasting Techniques) เป็นเทคนิคที่ใช้ข้อมูลในอดีตมาสร้างตัวแบบทางคณิตศาสตร์เพื่อใช้ในการพยากรณ์ โดยเทคนิคในกลุ่มนี้สามารถแบ่งได้เป็น 2 ประเภทใหญ่ ๆ ได้แก่

2.1.2.1 วิธีอนุกรมเวลา (Time Series Methods) เป็นวิธีที่ใช้เฉพาะข้อมูลในอดีตเพื่อพยากรณ์ค่าในอนาคต จากการวิเคราะห์แนวโน้ม (Trend) ฤดูกาล (Seasonal) วัฏจักร (Cyclical) และความผิดปกติ (Irregular) โดยมีตัวแปรอิสระคือเวลาเพียงอย่างเดียว

2.1.2.2 วิธีการพยากรณ์เชิงสาเหตุ (Causal Methods) เป็นวิธีที่ศึกษาถึงความสัมพันธ์ในเชิงเหตุและผลของตัวแปรที่สนใจกับตัวแปรอื่นที่เป็นสาเหตุหรือมีอิทธิพลต่อตัวแปรที่สนใจ เมื่อประมาณค่าความสัมพันธ์ระหว่างตัวแปรได้แล้วก็สามารถพยากรณ์ค่าของตัวแปรที่สนใจได้โดยอาศัยค่าของตัวแปรที่เป็นสาเหตุ (นันทชัย, 2555)

### 2.1.3 ขั้นตอนพื้นฐานของการพยากรณ์

โดยทั่วไปการพยากรณ์มีขั้นตอนพื้นฐาน 5 ขั้นตอนดังนี้

1. การระบุปัญหา (Problem Definition) เป็นส่วนที่ยากที่สุดของการพยากรณ์ ในการระบุปัญหาต้องทำความเข้าใจว่าจะพยากรณ์อะไร จะใช้วิธีการใดในการพยากรณ์ ใครต้องการใช้การพยากรณ์นี้ และตัวแบบการพยากรณ์นี้จะสอดคล้องกับความต้องการขององค์กรหรือผู้ใช้งานอย่างไร ผู้พยากรณ์จำเป็นต้องใช้เวลาในการสื่อสารทำความเข้าใจกับทุกคนที่เกี่ยวข้องกับการเก็บรวบรวมข้อมูล การรักษาฐานข้อมูล และการใช้ค่าพยากรณ์สำหรับการวางแผนอนาคต
2. การรวบรวมสารสนเทศ (Gathering Information) ส่วนใหญ่มีความต้องการสารสนเทศอย่างน้อย 2 ชนิดคือ (ก) ข้อมูลเชิงสถิติ และ (ข) ผู้ที่มีความเชี่ยวชาญในการรวบรวมข้อมูลและการพยากรณ์ ในหลาย ๆ ครั้งการได้มาของข้อมูลในอดีตอาจเพียงพอต่อการสร้างตัวแบบทางสถิตินั้นเป็นเรื่องที่ทำไต่ยาก ในกรณีนี้จะใช้เทคนิคการพยากรณ์เชิงวิจลนญาณหรือเรียกว่าเทคนิคการพยากรณ์เชิงคุณภาพ ซึ่งเป็นวิธีที่อาศัยความชำนาญในการประมาณการและความคิดเห็นของผู้เชี่ยวชาญเป็นหลัก ในบางครั้งข้อมูลในอดีตที่เก่าอาจนำไปใช้ประโยชน์ได้น้อยกว่าข้อมูลในอดีตที่ใกล้ปัจจุบันกว่า เนื่องจากระบบที่จะพยากรณ์อาจมีการเปลี่ยนแปลงโครงสร้าง แต่ตัวแบบทางสถิติที่ดีอาจจัดการกับการเปลี่ยนแปลงของสร้างของระบบได้ ถ้าไม่จำเป็นจึงไม่ควรที่จะทิ้งข้อมูลนั้น
3. การตรวจสอบและสำรวจข้อมูลเบื้องต้น (Exploratory Data Analysis) จะเริ่มต้นด้วยการสร้างกราฟของข้อมูลเพื่อตรวจสอบว่าข้อมูลนั้นมีรูปแบบเป็นลักษณะใด มีรูปแบบแนวโน้ม (Trend) มีรูปแบบฤดูกาล (Seasonal) มีหลักฐานการปรากฏขึ้นของวัฏจักรทางธุรกิจ มีค่าผิดปกติ (Outlier) ที่จำเป็นต้องอธิบายโดยผู้เชี่ยวชาญหรือไม่ หรือความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องเป็นอย่างไร
4. การเลือกและสร้างตัวแบบการพยากรณ์ที่เหมาะสม (Choosing and Fitting Models) ตัวแบบการพยากรณ์ที่เหมาะสมที่จะนำมาใช้งานขึ้นอยู่กับข้อมูลในอดีตที่มีอยู่ ความสัมพันธ์ระหว่างตัวแปรพยากรณ์ (Forecast Variable) กับตัวแปรอธิบาย (Explanatory Variables) และวิธีที่ใช้ในการพยากรณ์ โดยปกติจะเปรียบเทียบตัวแบบการพยากรณ์ที่มีศักยภาพ 2 ถึง 3 ตัวแบบ ซึ่งแต่ละตัวแบบจะมีรูปแบบที่ตั้งอยู่บนสมมติฐานที่เกี่ยวข้องกับตัวแบบนั้น และมักจะประกอบด้วยพารามิเตอร์ที่จะถูกประมาณค่าโดยใช้ข้อมูลในอดีตที่เกี่ยวข้อง
5. การใช้งานและการประเมินผลตัวแบบการพยากรณ์ (Using and Evaluating a Forecasting Model) เมื่อทำการเลือกตัวแบบและค่าพารามิเตอร์ของตัวแบบถูกประมาณค่า

แล้วก็จะนำตัวแบบนั้นไปใช้งาน จะสามารถประเมินประสิทธิภาพของตัวแบบได้หลังจากที่มีข้อมูลจริงที่เกิดขึ้นซึ่งจะใช้เปรียบเทียบกับข้อมูลที่เป็นผลลัพธ์ของตัวแบบพยากรณ์ในช่วงเวลานั้น การประเมินประสิทธิภาพของตัวแบบมีหลายวิธีที่ใช้ในการประเมินความแม่นยำของการพยากรณ์ (Hyndman และ Athanasopoulos, 2018)

#### 2.1.4 วิธีการพยากรณ์ด้วยวิธีอนุกรมเวลาเชิงสถิติ (Statistical Time Series Forecasting Methods)

อนุกรมเวลา คือ ชุดข้อมูลที่ถูกเก็บบันทึกตั้งแต่อดีตจนถึงปัจจุบันโดยระยะห่างของข้อมูลที่มีช่วงเวลาเท่ากัน เช่น รายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือรายปี โดยสมมติฐานสำคัญของการพยากรณ์คือ การพยากรณ์ข้อมูลในอนาคตจะขึ้นอยู่กับข้อมูลในอดีตและปัจจุบันในกรณีที่ไม่ได้มีการเปลี่ยนแปลงใด ๆ ที่ส่งผลกระทบต่อระบบที่จะพยากรณ์ อิทธิพลใด ๆ ที่ส่งผลกระทบต่อข้อมูลในอดีตและปัจจุบันจะยังส่งผลกระทบต่อข้อมูลเช่นเดิมในอนาคต และรูปแบบข้อมูลของอนุกรมเวลาอาจแบ่งลักษณะของข้อมูลได้เป็น

- (1) รูปแบบคงที่ (Stationary) คือรูปแบบข้อมูลอนุกรมเวลาที่มีลักษณะแกว่งอยู่รอบ ๆ ระดับคงที่หรือค่าเฉลี่ย
- (2) รูปแบบแนวโน้ม (Trend) คือรูปแบบข้อมูลอนุกรมเวลาที่มีทิศทางเพิ่มขึ้นหรือลดลงของข้อมูลในระยะยาว
- (3) รูปแบบฤดูกาล (Seasonal) คือรูปแบบอนุกรมเวลาที่มีการเคลื่อนไหวขึ้นลงซ้ำกันตามรอบของเวลา ซึ่งอาจเป็นรอบการซ้ำแบบรายสัปดาห์ รายเดือน รายไตรมาส หรือรายปี
- (4) รูปแบบวัฏจักร (Cycle) คือรูปแบบอนุกรมเวลาที่มีการเคลื่อนไหวของข้อมูลขึ้นลงซ้ำกันในระยะยาวที่มีช่วงเวลามากกว่า 1 ปี ส่วนใหญ่เกิดจากผลกระทบทางเศรษฐกิจ โดยระยะเวลาในการครบหนึ่งรอบอาจอยู่ระหว่าง 2 – 20 ปี
- (5) รูปแบบผิดปกติ (Irregular) คือรูปแบบอนุกรมเวลาที่มีการเคลื่อนไหวขึ้นลงที่ไม่มีการเรียงแบบแผน ไม่ทราบสาเหตุที่แน่ชัด และเกิดในช่วงเวลาสั้น ๆ ไม่เกิดซ้ำ (เอกจิตต์, 2560)

ซึ่งปัจจุบันมีตัวแบบให้เลือกมากมายตามลักษณะของข้อมูล อย่างไรก็ตามในงานวิจัยนี้จะขอกล่าวถึงวิธีดังต่อไปนี้

### 2.1.4.1 วิธีปรับเรียบเอ็กซ์โพเนนเชียลแบบโฮลต์และวินเทอร์ (Holt-Winters Exponential Smoothing)

วิธีนี้เหมาะกับการพยากรณ์ข้อมูลอนุกรมเวลาที่มีอิทธิพลของแนวโน้มและฤดูกาลเข้ามาเกี่ยวข้อง โดยค่าพยากรณ์ที่ได้จะคิดจากค่าเฉลี่ยถ่วงน้ำหนักของข้อมูลในอดีตที่ไม่เท่ากัน ซึ่งข้อมูลล่าสุดจะถูกให้น้ำหนักมากกว่าและน้ำหนักข้อมูลที่ส่งผลต่อค่าพยากรณ์จะลดลงแบบเอกซ์โพเนนเชียล (Exponential) รวมทั้งจะพิจารณาค่าแนวโน้มและฤดูกาลเข้าไปในตัวแบบ โดยตัวแบบ Holt-Winters มี 2 รูปแบบดังนี้ (Hanke และ Wichern, 2009)

#### 1) รูปแบบการบวก (Additive)

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.1)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (2.2)$$

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-1} \quad (2.3)$$

$$\hat{Y}_{t+p} = L_t + pT_t + S_{t-s+p} \quad (2.4)$$

#### 2) รูปแบบการคูณ (Multiplicative)

$$L_t = \alpha Y_t / S_{t-s} + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.5)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (2.6)$$

$$S_t = \gamma Y_t / L_t + (1 - \gamma)S_{t-s} \quad (2.7)$$

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \quad (2.8)$$

โดย  $L_t$  คือ ค่าของระดับที่เวลา  $t$

$T_t$  คือ ค่าของแนวโน้มที่เวลา  $t$

$S_t$  คือ ค่าของฤดูกาลที่เวลา  $t$

$Y_t$	คือ ค่าข้อมูลจริงที่เวลา $t$
$\hat{Y}_{t+p}$	คือ ค่าพยากรณ์ที่เวลา $t + p$
$\alpha$	คือ ค่าคงที่ปรับเรียบของค่าระดับ ( $0 < \alpha < 1$ )
$\beta$	คือ ค่าคงที่ปรับเรียบของแนวโน้ม ( $0 < \beta < 1$ )
$\gamma$	คือ ค่าคงที่ปรับเรียบของฤดูกาล ( $0 < \gamma < 1$ )
$p$	คือ จำนวนช่วงเวลาที่ต้องการจะพยากรณ์
$s$	คือ จำนวนช่วงเวลาในรอบฤดูกาล

#### 2.1.4.1 วิธีบ็อกซ์-เจนกินส์ (Box-Jenkins : ARIMA Model)

วิธีการพยากรณ์แบบ Autoregressive Integrated Moving Average (ARIMA) เป็นเทคนิคที่ใช้รูปแบบของข้อมูลในอดีตและปัจจุบันเป็นข้อมูลตั้งต้นแล้วนำมาวิเคราะห์ค่าผิดพลาดของผลการพยากรณ์ล่าสุดเพื่อที่จะทำการแก้ไขค่าพยากรณ์และนำไปใช้ในการพยากรณ์ต่อไปในอนาคต ในการพัฒนาตัวแบบ Box Jenkins: ARIMA Model นั้นจะเริ่มจากข้อมูลอนุกรมเวลาที่สังเกตได้ คัดเลือกตัวแบบที่เหมาะสม ทดสอบความน่าเชื่อถือของตัวแบบ ซึ่งถ้าตัวแบบที่เลือกมีความน่าเชื่อถือถูกต้องแล้ว ค่าความผิดพลาดของการพยากรณ์สมควรที่จะมีคุณสมบัติที่สำคัญคือ ไม่มีความสัมพันธ์กันเองในแต่ละข้อมูลและข้อมูลในอดีตไม่สามารถช่วยให้พยากรณ์ข้อมูลในอนาคตได้ ซึ่ง ARIMA Model มีตัวแบบหลัก 3 ตัวแบบ คือ ตัวแบบ Moving Average (MA) ตัวแบบ Autoregressive (AR) และตัวแบบ Mixed Autoregressive Moving Average (ARMA) โดยตัวแบบทั้ง 3 สามารถแตกออกเป็นตัวย่อย ๆ ได้อีกมากมาย การพิจารณาว่าตัวแบบใดใน 3 ตัวแบบนี้เหมาะสมกับข้อมูลอนุกรมเวลาในเบื้องต้นดูได้จากกราฟฟังก์ชันอัตสหสัมพันธ์ (Autocorrelation Function, ACF) ที่แสดงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อมูลที่คาบเวลาปัจจุบันกับข้อมูลที่คาบเวลาย้อนหลังไปในอดีต และกราฟฟังก์ชันอัตสหสัมพันธ์บางส่วน (Partial Autocorrelation Function, PACF) ที่แสดงค่าสัมประสิทธิ์ระหว่างข้อมูลที่เวลา  $t$  กับข้อมูลย้อนหลังที่เวลาย้อนไป  $t-k$  คาบ (ที่  $k$  Time lag)

สำหรับการพยากรณ์ข้อมูลอนุกรมเวลาที่มีแนวโน้มหรือไม่คงที่ (Non Stationary Data) จะมีการแปลงข้อมูลโดยทำการกำจัดอิทธิพลของแนวโน้มออกก่อนที่จะพิจารณาตัวแบบที่เหมาะสมว่าควรเป็น AR(p) MA(q) หรือ ARMA(p,q) โดยการปรับค่าแนวโน้มด้วย

วิธีการหาผลต่างลำดับที่ 1 (1st Differencing) ซึ่งเป็นการหาความแตกต่างระหว่างข้อมูลที่คาบเวลา  $t$  กับคาบเวลา  $t - 1$  เมื่อทำการหาผลต่างลำดับที่ 1 แล้วยังไม่สามารถกำจัดแนวโน้มให้หมดไปได้ จะต้องทำการปรับค่าแนวโน้มครั้งที่สองโดยทำการหาผลต่างลำดับที่ 2 (2nd Differencing) โดยจำนวนครั้งที่ทำการปรับค่าแนวโน้มจะถูกใส่เข้าไปในตัวแบบด้วย โดยเขียนเป็น ARIMA(p,d,q) ในกรณีที่ข้อมูลมีฤดูกาลมาเกี่ยวข้องจะมีการนำฤดูกาลเข้ามาพิจารณาในตัวแบบ Box-Jenkins: ARIMA ด้วย (เอกจิตต์, 2560) ซึ่งตัวแบบจะมีความซับซ้อนมากขึ้น โดยตัวแบบจะเขียนอย่างง่ายเป็น SARIMA(p,d,q)(P,D,Q)<sub>s</sub> มีชื่อเรียกว่า Seasonal Autoregressive Integrated Moving Average Model ต่อไปจะเรียกว่าตัวแบบ SARIMA ซึ่ง Box and Jenkins (1970) ได้อธิบายแบบจำลองการพยากรณ์ข้อมูลอนุกรมเวลาด้วยตัวแบบ SARIMA ซึ่งเป็นหนึ่งในแบบจำลองที่ได้รับความนิยมในการนำมาใช้เป็นมาตรฐานในการพยากรณ์ข้อมูลอนุกรมเวลา สามารถเขียนรูปแบบสมการได้ดังนี้

$$\Phi(B^S)\phi(B)\Delta^d\Delta_s^d y_t = c + \Theta(B^S)\theta(B)\varepsilon_t \quad (2.9)$$

โดยมีส่วนประกอบของตัวแบบดังนี้  
ส่วนประกอบที่ไม่มีฤดูกาลได้แก่

$$\text{ตัวแบบ MA: } \theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (2.10)$$

$$\text{ตัวแบบ AR: } \phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (2.11)$$

$$\text{การปรับค่าแนวโน้ม: } \Delta^d = (1 - B)^d \quad (2.12)$$

ส่วนประกอบที่มีฤดูกาลได้แก่

$$\text{ตัวแบบ MA: } \Theta(B^S) = 1 + \Theta_1 B^S + \dots + \Theta_Q B^{QS} \quad (2.13)$$

$$\text{ตัวแบบ AR: } \Phi(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS} \quad (2.14)$$

$$\text{การปรับค่าแนวโน้ม: } \Delta_s^d = (1 - B^S)^D \quad (2.15)$$



โดยที่  $y_t$  คือ ค่าพยากรณ์ที่คาบเวลา  $t$

$\phi$  และ  $\Phi$  คือ พารามิเตอร์ของตัวแบบ Autoregressive ที่ไม่มีฤดูกาล และมีฤดูกาลตามลำดับ

$d$  และ  $D$  คือ ค่าของลำดับการหาผลต่าง (Differencing) เพื่อกำจัดอิทธิพลของแนวโน้มออกที่ไม่มีฤดูกาลและมีฤดูกาลตามลำดับ

$\theta$  และ  $\Theta$  คือ พารามิเตอร์ของตัวแบบ Moving Average ที่ไม่มีฤดูกาล และมีฤดูกาลตามลำดับ

$p$  และ  $P$  คือ อันดับของตัวแบบ Autoregressive ที่ไม่มีฤดูกาลและมีฤดูกาลตามลำดับ

$q$  และ  $Q$  คือ อันดับของตัวแบบ Moving Average ที่ไม่มีฤดูกาลและมีฤดูกาลตามลำดับ

$s$  คือ จำนวนคาบของฤดูกาล

$c$  คือ ค่าคงที่ของตัวแบบ SARIMA

$\varepsilon_t$  คือ White Noise และ

$B$  คือ Backward Shift Operator ( $B_{y_t} = y_{t-1}$ )

ตัวแบบ Seasonal Autoregressive Integrated Moving Average with Exogenous Variables ต่อไปจะเรียกอย่างย่อว่า ตัวแบบ SARIMAX เป็นตัวแบบที่เป็นส่วนขยายของตัวแบบ SARIMA โดยการปรับปรุงความสามารถในการรวมตัวแปรจากภายนอก (Exogenous Variables) หรือตัวแปรอิสระจากภายนอกที่ไม่ใช่ตัวแปรที่เกี่ยวข้องกับอนุกรมเวลา เพื่อเพิ่มประสิทธิภาพในการพยากรณ์ สามารถเขียนรูปแบบสมการได้ดังนี้

$$\Phi(B^S)\phi(B)\Delta^d\Delta_s^d y_t = c + \Theta(B^S)\theta(B)\varepsilon_t + \sum_{k=1}^n \beta_k x_{k,t} \quad (2.16)$$

โดยที่  $x_{k,t}$  คือ ตัวแปรอิสระภายนอกตัวที่  $k^{th}$  ที่คาบเวลา  $t$

$\beta_k$  คือ ค่าสัมประสิทธิ์ของตัวแปรอิสระภายนอกตัวที่  $k^{th}$

### 2.1.5 การถดถอยเชิงเส้น (Linear Regression, LR)

การถดถอยเชิงเส้นเป็นเครื่องมือที่ใช้ในการพยากรณ์โดยอาศัยเทคนิคทางสถิติในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variables) และตัวแปรตาม (Dependent Variable) ซึ่งเป็นค่าที่ของตัวแปรที่เราต้องการจะพยากรณ์ โดยความสัมพันธ์ระหว่างตัวแปรอิสระ

และตัวแปรตามจะเป็นแบบเชิงเส้น จุดเริ่มต้นของการพัฒนาวิธีการถดถอยเชิงเส้นมาจากการพัฒนาเทคนิคกำลังสองที่น้อยสุด (Method of Least Squares) จนกระทั่งตัวแบบถดถอยได้ถูกพัฒนาแตกแขนงไปอย่างแพร่หลาย ซึ่งการถดถอยเชิงเส้นสามารถแบ่งได้เป็น 2 ประเภทดังนี้

#### 2.1.5.1 การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression, SLR)

การถดถอยเชิงเส้นอย่างง่ายเป็นการวิเคราะห์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระจำนวนหนึ่งตัวกับตัวแปรตามจำนวนหนึ่งตัว มีสมการความสัมพันธ์ดังนี้

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2.17)$$

โดยที่  $y$  คือ ตัวแปรตาม  
 $x$  คือ ตัวแปรอิสระ  
 $\beta_0$  และ  $\beta_1$  คือ สัมประสิทธิ์ของสมการการถดถอย  
 $\varepsilon$  คือ ค่าความผิดพลาด

#### 2.1.5.2 การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression, MLR)

การถดถอยเชิงเส้นพหุคูณเป็นการวิเคราะห์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระจำนวนมากว่าหนึ่งตัวกับตัวแปรตามจำนวนหนึ่งตัวมีสมการความสัมพันธ์ดังนี้

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (2.18)$$

โดยที่  $y$  คือ ตัวแปรตาม  
 $x_1, x_2, \dots, x_n$  คือ ตัวแปรอิสระ  
 $\beta_0, \beta_1, \dots, \beta_n$  คือ สัมประสิทธิ์ของสมการการถดถอย  
 $\varepsilon$  คือ ค่าความผิดพลาด

ในการประมาณค่าสัมประสิทธิ์ของสมการถดถอยนั้นมีหลายวิธีได้แก่ วิธีกำลังสองน้อยที่สุด (Least Squares Method) การเคลื่อนลงตามความชัน (Gradient Descent Algorithm) วิธีการถดถอยแบบบริดจ์ (Ridge Regression) การถดถอยลาสโซ (Lasso Regression) และการถดถอยอิลาสติกเน็ต (Elastic Net Regression) เป็นต้น ซึ่งวิธีกำลังสองน้อยที่สุดเป็นวิธีที่นิยมใช้มากที่สุด

การวิเคราะห์การถดถอยมีข้อสมมติฐานพื้นฐาน 4 ข้อคือ ที่ค่าใด ๆ ของตัวแปรอิสระ (1) ประชากรของค่าความผิดพลาดมีค่าเฉลี่ยเท่ากับศูนย์ (2) ประชากรของค่าความผิดพลาดมีความแปรปรวนคงที่ (3) ประชากรของค่าความผิดพลาดมีการแจกแจงแบบปกติ และ (4) ค่าความผิดพลาดมีความเป็นอิสระต่อกัน (นันทชัย, 2555)

#### 2.1.6 อัลกอริทึม Random Sample Consensus (RANSAC)

อัลกอริทึม Random Sample Consensus ต่อไปจะเรียกว่า RANSAC ถูกนำเสนอโดย Fischler และ Bolles (1981) ซึ่งถูกใช้ในการสร้างตัวแบบของข้อมูลทดลองและมีการใช้งานในการวิเคราะห์สถานที่และการทำแผนที่แบบอัตโนมัติ RANSAC เป็นวิธีการวนซ้ำ (Iterative Method) สำหรับการประมาณค่าพารามิเตอร์ของแบบจำลองทางคณิตศาสตร์จากชุดข้อมูลป้อนเข้าที่มีค่าผิดปกติ (Outlier) โดยในงานวิจัยนี้ผู้วิจัยใช้แบบจำลองทางคณิตศาสตร์เป็นแบบการถดถอยเชิงเส้นพหุคูณสำหรับการประมาณค่าพารามิเตอร์ของตัวแบบ ซึ่งจะเรียกว่า RANSAC Regressor มีขั้นตอนในการประมาณค่าดังนี้

- 1) เลือกสุ่มจำนวนจุดข้อมูลขั้นต่ำที่ต้องการใช้ในการประมาณค่าพารามิเตอร์
- 2) คำนวณค่าพารามิเตอร์ของตัวแบบด้วยวิธีการประมาณค่าพารามิเตอร์ของแบบจำลองทางคณิตศาสตร์ที่เลือก (ในงานวิจัยนี้ใช้แบบจำลองคณิตศาสตร์เป็นการถดถอยเชิงเส้นพหุคูณ)
- 3) กำหนดจำนวนจุดข้อมูลจากชุดของจุดข้อมูลทั้งหมดที่พอดีกับค่าความคลาดเคลื่อนที่กำหนดไว้ล่วงหน้า
- 4) ถ้าสัดส่วนของจำนวนจุดข้อมูลที่มีค่าปกติ (Inliers) มากกว่าจำนวนทั้งหมดของจุดข้อมูลในกลุ่มที่เกินเกณฑ์ที่กำหนดไว้ล่วงหน้า (Predefined Threshold) ให้ประมาณค่าพารามิเตอร์แบบจำลองใหม่โดยใช้ Inliers ที่ระบุทั้งหมดและสิ้นสุด
- 5) ถ้าไม่ใช่ดังในขั้นตอนที่ 4 ให้ทำซ้ำในขั้นตอนที่ 1 ถึง 4 จนครบรอบที่ได้กำหนดไว้ (Derpanis, 2010)

#### 2.1.7 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่องเป็นวิธีที่ได้รับการพิสูจน์ว่าเป็นแนวทางที่นำไปใช้ในทางปฏิบัติการทำเหมืองข้อมูล (Data Mining) เพื่อวิเคราะห์เชิงลึกจากข้อมูลหลาย ๆ ด้าน เช่น การแพทย์ การเงิน อุตสาหกรรมและการผลิต เกษตรกรรม ธุรกิจ รวมถึงด้านพลังงานด้วย นอกจากนี้ยังสามารถนำมาใช้ในการพยากรณ์เพื่อคาดการณ์อนาคตได้ และการเรียนรู้ของเครื่องยังเป็นส่วนหนึ่งของปัญญาประดิษฐ์หรือที่เรียกกันว่า AI (Artificial Intelligent) ซึ่งความหมายของการเรียนรู้ของเครื่อง

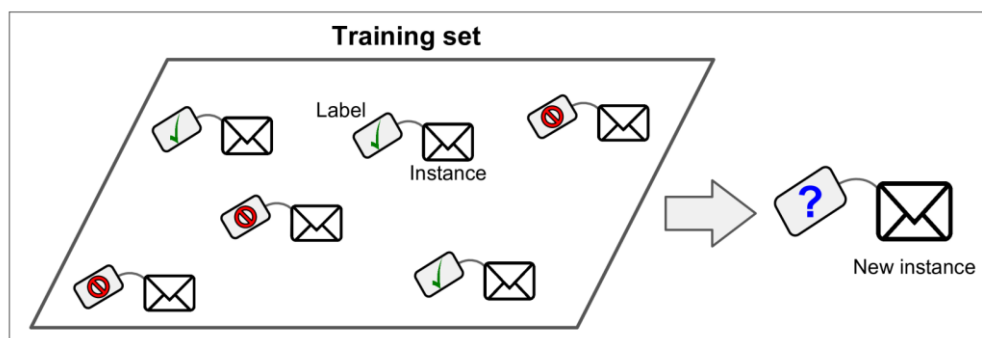
นั้นมีหลากหลายนิยามและหนึ่งในนั้น Arthur Samuel (1959) ได้ให้นิยามว่า ศาสตร์การเรียนรู้ที่ทำให้คอมพิวเตอร์สามารถเรียนรู้ได้ด้วยตัวเอง โดยปราศจากการป้อนโปรแกรมไว้อย่างชัดเจน ซึ่งคอมพิวเตอร์สามารถเรียนรู้ที่จะหาผลลัพธ์ได้ด้วยตัวเองจากข้อมูลที่ป้อนให้โดยไม่ต้องป้อนโปรแกรมให้คอมพิวเตอร์ทำงานตามคำสั่งที่กำหนดไว้ชัดเจน และข้อดีของการใช้การเรียนรู้ของเครื่องคือเหมาะสมสำหรับปัญหาที่มีความซับซ้อนและมีข้อมูลเป็นจำนวนมาก

### 2.1.8 ประเภทของการเรียนรู้ของเครื่อง

ในการแบ่งประเภทของการเรียนรู้ของเครื่องนั้นมีหลากหลายวิธี แต่โดยทั่วไปสามารถแบ่งได้เป็น 3 ประเภทจำแนกตามการใช้นุ้ษย์มนุษย์ในการฝึกสอนหรือไม่ ได้แก่ การเรียนรู้แบบมีผู้ฝึกสอน (Supervised Learning) การเรียนรู้แบบไม่มีผู้ฝึกสอน (Unsupervised Learning) และการเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) โดยมีลักษณะการนำไปใช้งานที่แตกต่างกันขึ้นอยู่กับวัตถุประสงค์ที่ต้องการ ซึ่งมีรูปแบบในการเรียนรู้ดังต่อไปนี้

#### 2.1.8.1 การเรียนรู้แบบมีผู้ฝึกสอน (Supervised Learning)

การเรียนรู้แบบมีผู้ฝึกสอนจะมีการฝึกตัวแบบโดยใช้ข้อมูลที่มีคำตอบหรือป้ายกำกับ (Label) สำหรับการสังเกตแต่ละครั้งในชุดข้อมูล ป้ายกำกับนี้อาจเป็นตัวแปรแบบจัดประเภท (Categorical Variable) เช่น เพศ อาชีพ ชนิดของโรค เป็นต้น หรือเป็นตัวแปรแบบต่อเนื่อง (Continuous Variable) คือตัวแปรที่สามารถวัดค่าได้ต่อเนื่อง เช่น น้ำหนัก ส่วนสูง รายได้ เป็นต้น ในการเรียนรู้แบบมีผู้ฝึกสอน ข้อมูลฝึกสอน (Training Data) จะถูกป้อนเข้าสู่อัลกอริทึมที่ประกอบด้วยคำตอบที่ต้องการหรือเรียกว่าป้ายกำกับ (Label) ดังแสดงในรูปที่ 4



รูปที่ 4 ชุดข้อมูลฝึกสอนที่มีป้ายกำกับสำหรับการเรียนรู้แบบมีผู้ฝึกสอน

(Géron, 2019)

โดยทั่วไปงานของการเรียนรู้แบบมีผู้ฝึกสอนเป็นแบบการจำแนก (Classification) ซึ่งจะใช้ตัวอย่างในการอธิบายการจำแนกคือ การกรองจดหมายอิเล็กทรอนิกส์หรืออีเมลที่ไม่พึงประสงค์ (Spam Filter) ซึ่งใช้การฝึกสอนด้วยตัวอย่างของอีเมลจำนวนมากที่มีทั้งอีเมลที่เป็นอีเมลที่ไม่พึงประสงค์ (Spam) กับอีเมลประเภทอื่น ๆ ที่ไม่ใช่อีเมลที่ไม่พึงประสงค์โดยการจำแนกตามกลุ่ม (Class) และจะต้องเรียนรู้วิธีที่จะจำแนกอีเมลที่เป็นชุดใหม่ด้วย

นอกจากนี้ยังมีการเรียนรู้แบบมีผู้ฝึกสอนอีกแบบที่เป็นงานในการทำนายค่าตัวเลขเป้าหมาย ตัวอย่างเช่น การทำนายราคาของรถยนต์ โดยกำหนดให้เซตของคุณลักษณะ (Features) จำพวก เลขไมล์ อายุ ยี่ห้อ เป็นต้น ซึ่งจะเรียกคุณลักษณะเหล่านี้ว่าตัวทำนาย (Predictors) กระบวนการในการทำนายงานในลักษณะนี้เรียกว่าการถดถอย (Regression) ในการทำนายราคาของรถยนต์จำเป็นต้องใช้ข้อมูลตัวอย่างรถยนต์จำนวนมากซึ่งประกอบไปด้วยตัวทำนายและป้ายกำกับ (ในตัวอย่างนี้คือราคาของรถยนต์)

ในอัลกอริทึมของการถดถอยบางอย่างสามารถใช้กับปัญหาแบบการจำแนกได้เป็นอย่างดี เช่น Logistic Regression ซึ่งเป็นการถดถอยแบบที่ใช้สำหรับการจำแนกที่สามารถให้ผลลัพธ์ที่มีค่าสอดคล้องกับความน่าจะเป็นของกลุ่มที่กำหนด (เช่น มีโอกาส 20% ที่จะเป็นอีเมลที่ไม่พึงประสงค์)

สำหรับในงานวิจัยนี้ซึ่งเป็นการพยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูป อัลกอริทึมการเรียนรู้ของเครื่องที่เหมาะสมที่จะนำมาใช้พยากรณ์เป็นการเรียนรู้แบบมีผู้ฝึกสอนที่มีลักษณะงานแบบการถดถอย โดยสามารถประยุกต์ใช้อัลกอริทึมเหล่านี้ได้ เช่น k-Nearest Neighbors (KNN) Multiple Linear Regression (MLR) Support Vector Regression (SVR) Artificial Neural Network (ANN) และ Adaboost (ADA) เป็นต้น ซึ่งจะกล่าวถึงในหัวข้อต่อ ๆ ไป

### 2.1.8.2 การเรียนรู้แบบไม่มีผู้ฝึกสอน (Unsupervised Learning)

เป้าหมายของการเรียนรู้แบบไม่มีผู้ฝึกสอนคือการค้นหาความสัมพันธ์และรูปแบบระหว่างข้อมูลที่ป้อนเข้ามาในอัลกอริทึม และการเรียนรู้แบบไม่มีผู้ฝึกสอนสามารถใช้ลดขนาดมิติข้อมูล (Dimensionality Reduction) สำหรับข้อมูลที่มีมิติขนาดใหญ่หรือมีจำนวนข้อมูลและคุณลักษณะ (Features) เป็นจำนวนมากได้ ซึ่งมีความแตกต่างจากการเรียนรู้แบบมีผู้ฝึกสอนคือไม่ต้องเตรียมข้อมูลป้อนเข้าที่เป็นป้ายกำกับ ทำให้ประหยัดเวลาและแรงงานในการดำเนินงานลงได้ โดยงานที่ทำโดยส่วนใหญ่ของการเรียนรู้แบบไม่มีผู้ฝึกสอนได้แก่ การจัดกลุ่มของข้อมูล (Clustering) การหารูปแบบของข้อมูลที่เกิดขึ้นร่วมกันบ่อย ๆ (Association Mining) และการลดขนาดมิติข้อมูล (Dimensionality Reduction) ในการใช้วิธีการเหล่านี้

รูปแบบของข้อมูลที่ไม่รู้จักมาก่อนจะถูกค้นพบระหว่างจุดข้อมูลต่าง ๆ สำหรับอัลกอริทึมหรือตัวแบบการเรียนรู้แบบไม่มีผู้ฝึกสอนจะไม่สามารถใช้กับการพยากรณ์ได้โดยตรงเนื่องจากไม่มีผลลัพธ์หรือคำตอบที่เป็นค่าตัวเลขของผลการพยากรณ์

### 2.1.8.3 การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning)

การเรียนรู้แบบเสริมกำลังเป็นการเรียนรู้จากปฏิสัมพันธ์ (Interaction) ระหว่างผู้เรียนรู้ (Agent) กับสิ่งแวดล้อม (Environment) ซึ่งผู้เรียนรู้จะสังเกตสิ่งแวดล้อม ทำการเลือกและปฏิบัติโดยการลองผิดลองถูกเพื่อให้ได้ผลลัพธ์หรือสิ่งตอบแทน (Reward) การเรียนรู้ด้วยตัวเองว่าวิธีการใดดีที่สุดเรียกว่าวิธิต่าง (Policy) เพื่อให้ได้รับสิ่งตอบแทนมากที่สุดตลอดระยะเวลา นั้น วิธิต่างนิยามได้ว่าเป็นการกระทำที่ผู้เรียนรู้ควรจะเลือกทำเมื่อกำหนดสถานการณ์ให้ ตัวอย่างของการเรียนรู้แบบเสริมกำลังเช่น การสร้างหุ่นยนต์จำนวนมากมีการนำอัลกอริทึมการเรียนรู้แบบเสริมกำลังมาใช้งานในการเรียนรู้วิธีการเดิน โปรแกรมอัลฟาโก (DeepMind's AlphaGo Program) เป็นอีกหนึ่งตัวอย่างของการเรียนรู้แบบเสริมกำลังที่เป็นโปรแกรมบอทที่สามารถเอาชนะผู้เล่นระดับชนะเลิศของโลกได้โดยการเรียนรู้วิธีการชนะด้วยการวิเคราะห์เกมส์การเล่นกว่าหลายล้านเกมส์เพื่อทำการประเมินว่าควรเดินหมากรอบถัดไปอย่างไร โดย AlphaGo จะทำการพิจารณาหลาย ๆ ตำแหน่งที่ควรเดินด้วยจนสามารถเอาชนะผู้เล่นระดับชนะเลิศของโลกได้ (Géron, 2019)

### 2.1.9 ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้สุด k ตัว (K-nearest Neighbor Algorithm, KNN)

ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้สุด k ตัว หรือเรียกแบบย่อว่า KNN เป็นอัลกอริทึมการเรียนรู้ของเครื่องแบบหนึ่งที่ใช้ต้นทุนการคำนวณต่ำและนำไปใช้งานได้ง่าย ซึ่งรองรับการนำไปใช้งานทั้งปัญหาแบบการจำแนก (Classification) และแบบการถดถอย (Regression) เมื่อทำการทำนายจะมีการเก็บชุดข้อมูลฝึกสอนทั้งหมดและค้นหาตำแหน่งของจุดข้อมูลจำนวน k จุดในชุดข้อมูลฝึกสอนที่ใกล้เคียงจุดข้อมูลที่ต้องการจำแนกมากที่สุด อย่างไรก็ตามไม่มีตัวแบบอื่นนอกจากที่มาจากชุดข้อมูลฝึกสอนและทำการคำนวณระยะทางระหว่างจุดคือการค้นหาจากชุดข้อมูลฝึกสอนเท่านั้น

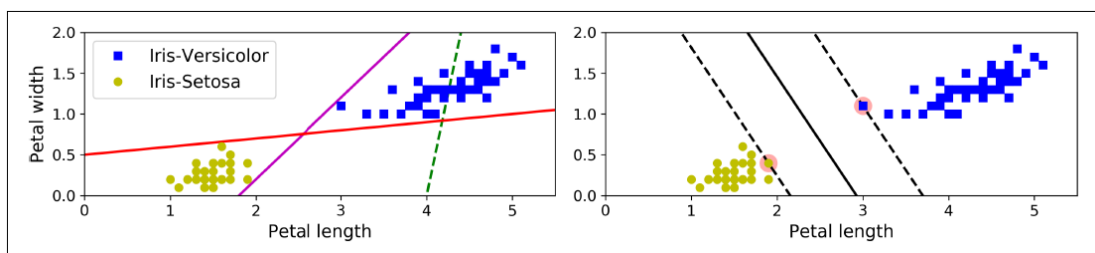
ในการใช้วิธี KNN ในปัญหาการถดถอยเพื่อทำการพยากรณ์ ค่าของตัวแปรตามจะถูกคำนวณจากผลรวมถ่วงน้ำหนักของค่าตัวแปรตามของเพื่อนบ้านใกล้เคียง k ทั้งหมด โดยที่น้ำหนักจะแปรผกผันกับระยะห่างระหว่างจุดข้อมูลป้อนเข้า (Input Record) และระยะห่างจะใช้การวัดระยะทางแบบยูคลิด (Euclidean) ฟังก์ชันระยะทางแบบยูคลิดเขียนเป็นสมการได้ดังนี้ (Al-Dosary et al., 2019)

$$E(x, p) = \sqrt{\sum_{i=1}^n (x_i - p_i)^2} \quad (2.19)$$

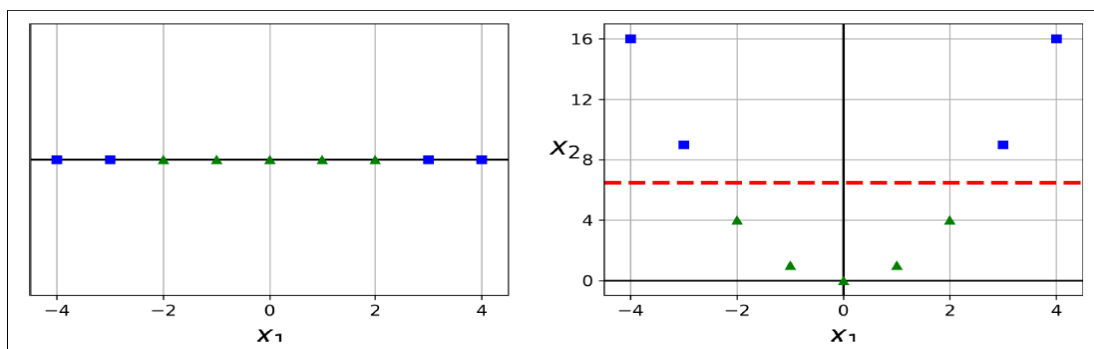
โดยที่  $x$  และ  $p$  คือ จุดที่จะใช้หาระยะห่างจากกัน  
 $n$  คือ จำนวนข้อมูลของตัวแปรป้อนเข้า (Attributes)

#### 2.1.10 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machines Algorithm, SVM)

ซัพพอร์ตเวกเตอร์แมชชีนเป็นอัลกอริทึมการเรียนรู้ของเครื่องที่มีประสิทธิภาพและมีความอ่อนน้อมประสงค์สามารถที่จะใช้งานได้กับปัญหาการจำแนก (Classification) การถดถอย (Regression) ทั้งแบบเชิงเส้นหรือไม่เชิงเส้น และยังสามารถใช้ตรวจสอบค่าผิดปกติ (Outlier) ของข้อมูลได้อีกด้วย สำหรับซัพพอร์ตเวกเตอร์แมชชีนแบบเชิงเส้นมีหลักการเบื้องหลังซึ่งอธิบายได้จากรูปที่ 5 โดยแสดงถึงชุดข้อมูลที่ต้องการจะจำแนกประเภทของดอกไอริส (Iris) ที่มี 2 ประเภทคือ Versicolor และ Sentosa ซึ่งทั้ง 2 กลุ่ม (Class) ประเภทของดอกไอริสจะถูกแบ่งแยกได้ง่ายด้วยเส้นตรง จากกราฟทางด้านซ้ายแสดงถึงขอบเขตการตัดสินใจที่เป็นไปได้ 3 ลักษณะของการจำแนกแบบเชิงเส้น ตัวแบบที่ใช้ขอบเขตการตัดสินใจที่ใช้เส้นประสีเขียวแสดงในกราฟเป็นลักษณะที่ไม่ดีเพราะไม่สามารถแบ่งกลุ่มประเภทของดอกไอริสออกจากกันได้และเส้นตัดผ่านกลุ่มจุดข้อมูลอีกด้วย เมื่อเปรียบเทียบกับเส้นตรงสีแดงและสีม่วงซึ่งเป็นขอบเขตการตัดสินใจของการจำแนกด้วย SVM โดยสามารถแบ่งกลุ่มประเภทของดอกไอริสออกจากกันได้และเส้นไม่มีการตัดผ่านกลุ่มหรือจุดของข้อมูลแต่ยังมีช่วงที่ระยะห่างระหว่างเส้นตรงกับจุดข้อมูลในกลุ่มที่ใกล้ที่สุดมีระยะที่ใกล้กันมาก การจำแนกด้วย SVM ดังในกราฟด้านขวาจะเห็นว่าเส้นตรงนั้นเป็นขอบเขตการตัดสินใจที่สามารถแบ่งกลุ่มประเภทของดอกไอริสได้และยังมีระยะห่างที่มากที่สุดระหว่างเส้นตรงกับจุดข้อมูลในกลุ่มที่ใกล้ที่สุด โดยมีเส้นประแสดงถึงเส้นประคู่ขนานกับเส้นตรง ระยะห่างระหว่างเส้นประคู่ขนานทั้งสองเส้นกับเส้นตรงที่กว้างที่สุดที่เป็นไปได้คือระยะที่เส้นประคู่ขนานมีระยะห่างกับจุดในกลุ่มน้อยที่สุดซึ่งเรียกว่า Large Margin



รูปที่ 5 ขอบเขตการตัดสินใจของการจำแนกด้วย SVM (Géron, 2019)



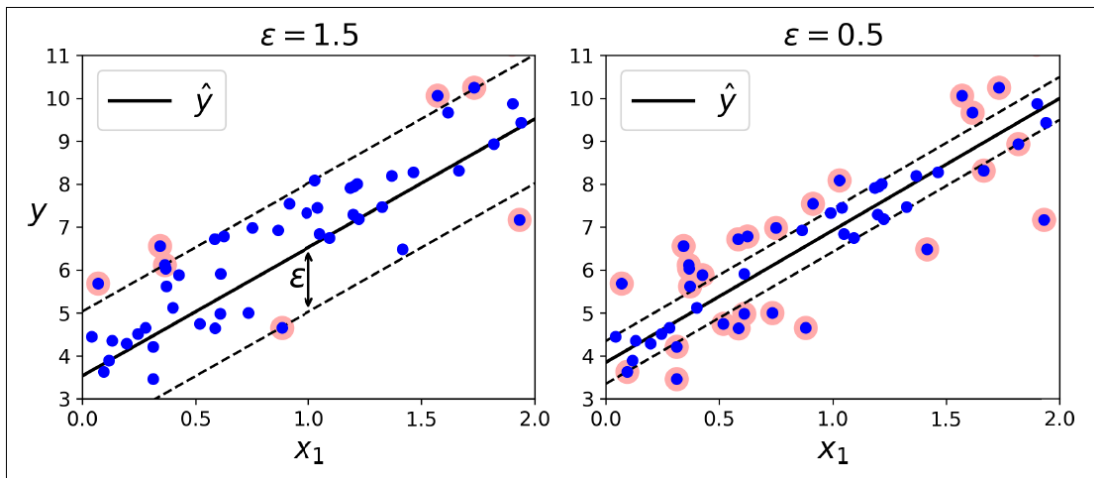
รูปที่ 6 การเพิ่มคุณลักษณะเพื่อให้สามารถแยกกลุ่มข้อมูลด้วย SVM แบบเชิงเส้นได้  
(Géron, 2019)

สำหรับชุดข้อมูลที่ไม่สามารถใช้ SVM แบบเชิงเส้นได้ หนึ่งในแนวทางที่ใช้แก้ปัญหาได้คือการเพิ่มลักษณะ (Features) เช่นเพิ่มคุณลักษณะแบบพหุนาม กราฟด้านซ้ายของรูปที่ 6 แสดงถึงชุดข้อมูลที่มีเพียงหนึ่งคุณลักษณะ ซึ่งไม่สามารถใช้การแบ่งกลุ่มในลักษณะเชิงเส้นได้ แต่เมื่อเพิ่มคุณลักษณะที่สองเข้ามา  $x_2 = x_1^2$  ผลลัพธ์ที่ได้คือชุดข้อมูล 2 มิติที่สามารถใช้ SVM เชิงเส้นได้

การเพิ่มคุณลักษณะพหุนามเข้าไปในตัวแบบนั้นถ้ามีดีกรีพหุนามต่ำจะไม่สามารถใช้งานกับชุดข้อมูลที่มีความซับซ้อนได้ หากมีดีกรีพหุนามสูงจะสร้างคุณลักษณะจำนวนมากทำให้ตัวแบบดำเนินการได้ช้าลงมาก แต่ในการใช้งาน SVM สามารถที่จะนำเทคนิคทางคณิตศาสตร์ที่เรียกว่า Kernel Trick ที่จะให้ผลลัพธ์เช่นเดียวกับการเพิ่มคุณลักษณะพหุนามที่มีดีกรีสูงโดยไม่ต้องเพิ่มคุณลักษณะเช่นนั้นเข้าไปในตัวแบบ

ในปัญหาการถดถอย มีการนำ SVM ไปประยุกต์ใช้เพื่อทำนายค่าได้ทั้งแบบการถดถอยเชิงเส้นและไม่เชิงเส้น เรียกว่าซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression, SVR) กลวิธีที่ใช้จะคล้ายกับ SVM แต่มีวัตถุประสงค์ต่างกันคือ SVR มีแถบเส้นขอบเขตที่พยายามครอบคลุมจุดข้อมูลในชุดข้อมูลฝึกสอนให้มากที่สุด โดยความกว้างของแถบเส้นขอบเขตจะถูกควบคุมด้วยด้วยไฮเปอร์พารามิเตอร์ (Hyperparameter)  $\epsilon$  รูปที่ 7 แสดงถึงแถบเส้นขอบเขตของ SVR แบบเชิงเส้นในตัวแบบที่ถูกฝึกสอนจากชุดข้อมูลเชิงเส้น กราฟด้านซ้ายมี Margin ขนาดใหญ่ ( $\epsilon = 1.5$ ) และกราฟด้านขวามี Margin ขนาดเล็ก ( $\epsilon = 0.5$ ) (Géron, 2019)

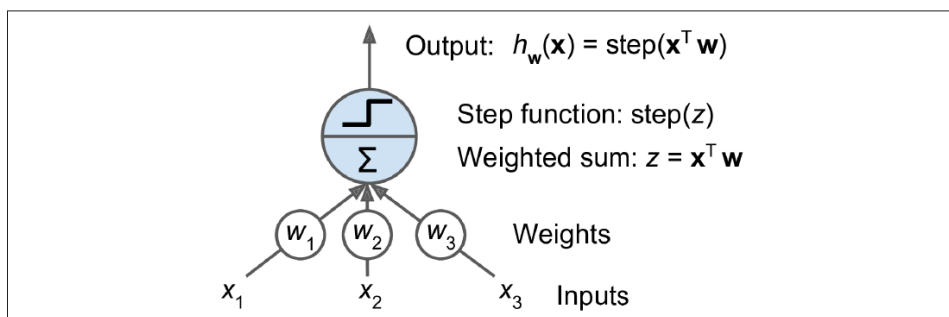




รูปที่ 7 Support Vector Regression (Géron, 2019)

### 2.1.11 โครงข่ายประสาทเทียม (Artificial Neural Network, ANN)

โครงข่ายประสาทเทียมเป็นตัวแทนของการคำนวณที่ได้รับแรงบันดาลใจจากสมองมนุษย์ ที่มีเซลล์ประสาท (Neurons) จำนวนมากเชื่อมต่อกันเป็นโครงข่ายการสื่อสารที่ซับซ้อน ซึ่งสามารถทำการคำนวณที่มีความซับซ้อนสูงได้ โครงข่ายประสาทเทียมมีโครงสร้างการคำนวณที่จำลองตามรูปแบบการคำนวณลักษณะนี้ ซึ่งสามารถอธิบายได้ว่ากราฟระบุทิศทาง (Directed Graph) ที่มีโหนดเช่นเดียวกับ Neurons และมีเส้นเชื่อม (Edges) ที่เชื่อมต่อระหว่างโหนด แต่ละ Neurons จะได้รับข้อมูลนำเข้า (Input) ส่งเข้ามาคูณกับค่าน้ำหนัก (Weight) ซึ่งแทนความสำคัญของข้อมูลนำเข้าแต่ละตัว ผลรวมถ่วงน้ำหนัก (Weighted Sum) ที่เกิดจากผลคูณระหว่างข้อมูลนำเข้าและค่าน้ำหนักจะถูกนำไปวิเคราะห์ด้วยฟังก์ชันกระตุ้น (Activation Function หรือ Step Function) (Ben-David และ Shalev-Shwartz, 2014) สุดท้ายเกิดเป็นผลลัพธ์ดังรูปที่ 8

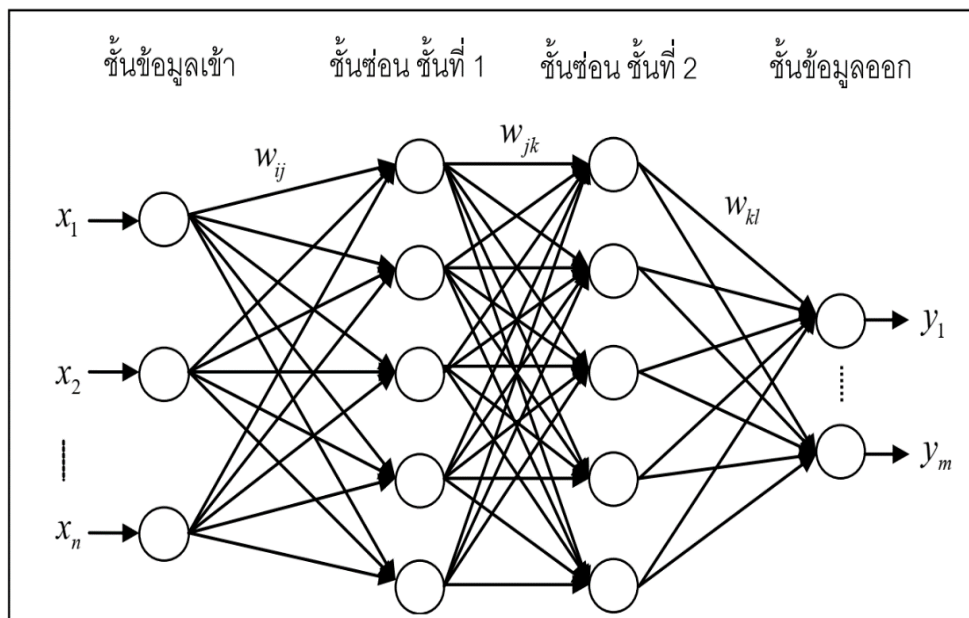


รูปที่ 8 การจำลองโครงข่ายประสาทเทียมจากแบบจำลองของเซลล์ประสาทหนึ่งเซลล์

(Géron, 2019)

นักวิจัยได้เสนอรูปแบบโครงสร้างหรือสถาปัตยกรรมของ ANN หลายรูปแบบเพื่อนำไปใช้ในการทำงานที่แตกต่างกัน เช่น แบบป้อนไปข้างหน้า (Feedforward) แบบย้อนกลับ (Feedback หรือ Recurrent) และแบบแข่งขัน (Competitive) นอกจากนี้ ANN ยังมีการเรียนรู้ซึ่งเป็นการปรับค่าน้ำหนักเพื่อให้โครงข่ายได้เรียนรู้และแสดงพฤติกรรมที่ต้องการ โดยการเรียนรู้แบ่งได้เป็น 2 ประเภท คือ การเรียนรู้แบบมีการสอน (Supervised Learning) และ การเรียนรู้แบบไม่มีการสอน (Unsupervised Learning)

ในการนำ ANN มาใช้ในการพยากรณ์นั้น โครงสร้างที่นิยมใช้คือโครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron, MLP) ซึ่งเป็นโครงสร้างแบบป้อนไปข้างหน้า โครงสร้างนี้จะประกอบไปด้วยชั้น (Layer) ต่าง ๆ 3 ชั้น ได้แก่ ชั้นข้อมูลเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นข้อมูลออก (Output Layer) โดยที่ชั้นซ่อนอาจมีมากกว่าหนึ่งชั้น ในแต่ละชั้นจะประกอบด้วยโหนดหรือเซลล์ประสาทหนึ่งเซลล์หรือมากกว่า และข้อมูลจะถูกส่งจากชั้นข้อมูลเข้าไปถึงชั้นข้อมูลออกโดยไม่มีการส่งย้อนกลับ แม้อโหนดในชั้นเดียวกันก็ไม่มีการเชื่อมต่อกัน ดังแสดงในรูปที่ 9 และใช้เทคนิคการเรียนรู้แบบมีการสอนชนิดแพร่กลับ (Back Propagation) ซึ่งเป็นวิธีที่ปรับค่าน้ำหนักที่เชื่อมต่อนระหว่างโหนดให้เหมาะสมโดยพิจารณาจากผลต่างระหว่างผลลัพธ์ที่ได้กับผลลัพธ์ที่ต้องการ



รูปที่ 9 โครงข่ายประสาทเทียมแบบหลายชั้น (นันทชัย, 2555)

### 2.1.12 Ensemble Machine Learning Based on Regression Trees

วิธีการ Ensemble เป็นวิธีที่รวมเอาตัวแบบหลาย ๆ ตัวแบบเพื่อให้ได้ประสิทธิภาพในการคาดการณ์ที่ดีขึ้นแทนที่จะใช้ตัวแบบเดียว โดยจะรวมอัลกอริทึมการเรียนรู้แบบมีผู้ฝึกสอนเข้าด้วยกันเป็น Supermodel วัตถุประสงค์ของการใช้ Ensemble คือการปรับปรุงจุดด้อยของแต่ละตัวแบบเดี่ยว (Individual Model) ให้ได้ประสิทธิภาพที่ดีที่สุดของตัวแบบรวม (Combined Model) เพื่อให้ได้ความแม่นยำของการพยากรณ์ที่ดีขึ้น ซึ่งประกอบด้วยอัลกอริทึมการเรียนรู้ของเครื่องพื้นฐาน เช่น ต้นไม้การตัดสินใจ (Decision Tree) โครงข่ายประสาทเทียม (ANN) หรืออัลกอริทึมประเภทอื่น ๆ (Lee et al., 2020)

ต้นไม้การตัดสินใจ (Decision Tree) เป็นวิธีการเรียนรู้ของเครื่องที่ใช้กับชุดข้อมูลที่ไม่เป็นเชิงเส้นได้ สำหรับการพยากรณ์ตัวแปรตามหรือตัวแปรเป้าหมาย (Target Variable) ที่มีค่าต่อเนื่อง  $y_{t+h}$  ด้วยวิธีต้นไม้การตัดสินใจจะเรียกว่าต้นไม้การถดถอย (Regression Tree) ซึ่งสามารถใช้ในการตรวจสอบลักษณะเฉพาะ (Characteristics) ต่าง ๆ ที่มีความคล้ายกันของแต่ละกลุ่มค่าสังเกต (Observations) ได้ แต่ละโหนดภายในต้นไม้จะถูกสร้างขึ้นอย่างเป็นขั้นตอน ในขั้นตอนแรกค่าสังเกตของตัวแปรเป้าหมาย  $y_{t+h}$  จะถูกจัดเรียงเป็นหลายโหนดโดยใช้หนึ่งตัวแปรจากแต่ละตัวแปรทำนาย (Predictor Variables)  $X_t = (x_{1t}, x_{2t}, \dots, x_{Nt})$  บางโหนดสามารถรับค่าโดยใช้ค่าเฉลี่ยตัวอย่าง (Sample Average) ของ  $y_{t+h}$  ตามเงื่อนไขในช่วงของตัวทำนายที่เลือก โหนดที่ไม่มีค่าจะถูกแบ่งอีกครั้งโดยใช้ช่วงของตัวทำนายที่เหลือ กระบวนการนี้จะดำเนินต่อไปจนกว่าโหนดทั้งหมดจะได้รับค่า โดยใช้บางกฎการหยุด (Stopping Rule) สมการของต้นไม้การถดถอยกับ Terminal Node (M) สามารถเขียนได้ดังนี้

$$y_{t+h} = \sum_{m=1}^M \theta_m 1_{\{X_t \in R_m\}} + \varepsilon_{t+h} \quad (2.20)$$

เมื่อ  $1_{\{\cdot\}}$  คือ ฟังก์ชันชี้บ่ง (Indicator Function)

$R_m$  คือ การแบ่งพื้นที่ (Partition of the Space) ของ  $X_t$

$\theta_m$  คือ ค่าเฉลี่ยกลุ่มตัวอย่าง (Sample Average) ของ  $y_{t+h}$  ตามเงื่อนไข

$$\text{บน } X_t \in R_m$$

การประมาณค่าของต้นไม้การถดถอยเป็นการค้นหาโครงสร้างต้นไม้ที่ดีที่สุดที่ทำให้ค่ากำลังสองของ  $\sum_{t=1}^T \varepsilon_{t+h}^2$  น้อยที่สุด เพื่อที่จะวาดภาพต้นไม้ดังกล่าวในเชิงตัวเลขให้เลือกตัวแปรที่จัดเรียงจาก  $X_t$  และชุดของค่าที่แยกกันของแต่ละโหนด ในการค้นหาค่าที่ดีที่สุดเฉพาะของการจัดเรียงตัวแปร

(Locally Optimal Sorting Variable) และค่าที่แยกกันของแต่ละโหนดจะใช้อัลกอริทึมของต้นไม้การถดถอยแบบไบนารี (Binary Regression Tree) ที่เสนอโดย Breiman et al. (1984)

ต้นไม้การถดถอยมีประสิทธิภาพในการจัดการกับความไม่เป็นเชิงเส้นและการมีปฏิสัมพันธ์กันระหว่างตัวแปรต่าง ๆ นอกจากนี้การตีความปัจจัยของตัวแปรเป้าหมายนั้นค่อนข้างง่ายและโครงสร้างของตัวแบบสามารถทำให้เข้าใจได้อย่างชัดเจน แต่ในขณะเดียวกันต้นไม้การตัดสินใจจะมีความอ่อนไหวต่อการเปลี่ยนแปลงเล็กน้อยของข้อมูล ดังนั้นการพยากรณ์ที่ใช้ข้อมูลนอกเหนือจากกลุ่มตัวอย่างสามารถให้ผลที่ไม่ดีได้ เพื่อหลีกเลี่ยงปัญหาการเข้ากันเกินไป (Overfitting) ต้นไม้การถดถอยจะใช้วิธีการ Ensemble Machine Learning (Maehashi และ Shintani, 2020)

Bagging คือคำย่อของ Bootstrap Aggregation เป็นรูปแบบหนึ่งของวิธีการ Ensemble ซึ่งเป็นวิธีการสุ่มข้อมูลย่อยจากชุดข้อมูลฝึกสอนแบบสุ่มซ้ำได้ (Boostraps Replacement) จำนวนหลายชุดย่อย นำมาสร้างตัวจำแนกที่แตกต่างกันแล้วจึงนำผลการทำนายของแต่ละตัวจำแนกย่อยมาพิจารณาร่วมกัน (Breiman, 1996) เรียกวิธีการนี้ว่า Bootstraps Sampling ซึ่ง Shmueli et al. (2016) แนะนำว่า Bagging ช่วยปรับปรุงเสถียรภาพในการทำงานของตัวแบบและช่วยหลีกเลี่ยงปัญหาการเข้ากันเกินไป ด้วยการแยกกันการสร้างแบบจำลองข้อมูลกลุ่มตัวอย่างที่แตกต่างกัน จากนั้นจึงนำผลลัพธ์มารวมกัน จึงมีประโยชน์อย่างยิ่งกับอัลกอริทึมประเภทต้นไม้การถดถอยและโครงข่ายประสาทเทียม

Random Forest เป็นหนึ่งในอนุพันธ์ของ Bagging ดังนั้นจึงมีอัลกอริทึมที่มีความสัมพันธ์กันอย่างใกล้ชิด (Breiman, 2001) เพื่อรักษาเสถียรภาพของการพยากรณ์จาก Bagging การทำ Bootstraps Sampling ของต้นไม้การถดถอยไม่ควรมีความสัมพันธ์กันมาก นอกจากนี้การพยากรณ์ด้วยต้นไม้การถดถอยของแต่ละ Bootstraps Sampling ทำงานคล้ายกันและเฉลี่ยเท่า ๆ กัน อาจไม่มีประสิทธิภาพในการลดความแปรปรวนของการพยากรณ์ ใน Random Forest จะใช้วิธีการออกกลางคัน (Dropout Procedure) ในการทำให้ต้นไม้การถดถอยของ Bootstraps Sampling ไม่มีสหสัมพันธ์กัน

Boosting เป็นอัลกอริทึมการเรียนรู้ของเครื่องแบบ Ensemble อีกหนึ่งอย่างที่ถูกใช้ในการแก้ปัญหาการเข้ากันเกินไป นำเสนอโดย Schapire (1990) และ Freund (1995) ในปัญหาการจำแนก (Classification Problems) Boosting เป็นวิธีที่แตกต่างเล็กน้อยในการสร้าง Ensemble โดยการปรับปรุงโดยตรงในบริเวณข้อมูลที่ตัวแบบเกิดข้อผิดพลาดในการบังคับตัวแบบให้ความสำคัญกับชุดข้อมูลที่เก็บ (Records) มากขึ้น (Zhou, 2012) ขั้นตอนในการ Boosting ประกอบด้วยสร้างตัวแบบจากข้อมูล สุ่มตัวอย่างจากข้อมูล ดังนั้นค่าสังเกตการณ์ที่มีความผิดพลาดของการพยากรณ์สูงจะมีโอกาสที่จะถูกเลือกและนำไปสร้างตัวแบบสำหรับตัวกลุ่มตัวอย่างใหม่ และให้ทำซ้ำหลาย ๆ ครั้ง ในการสุ่มตัวอย่างจากข้อมูลและสร้างตัวแบบสำหรับข้อมูลกลุ่มตัวอย่างใหม่ (Lee et al., 2020)

## 2.2 งานวิจัยที่เกี่ยวข้อง

### 2.2.1 การพยากรณ์น้ำมันสำเร็จรูป

Melikoglu (2014) ได้ศึกษาการพยากรณ์ความต้องการใช้น้ำมันเชื้อเพลิงสำหรับการขนส่งทางบกของประเทศตุรกีในปี ค.ศ. 2013 ถึง ค.ศ. 2023 ซึ่งประกอบด้วยน้ำมันเบนซิน น้ำมันดีเซล ก๊าซปิโตรเลียมเหลว น้ำมันไบโอเอทานอล และน้ำมันไบโอดีเซล โดยได้พยากรณ์ความต้องการน้ำมันเบนซินด้วยตัวแบบ Gaussian, Modified Gaussian และ Lorentzian ซึ่งเป็นตัวแบบกึ่งเชิงประจักษ์ (Semi-Empirical Model) สำหรับน้ำมันดีเซลและก๊าซปิโตรเลียมเหลวใช้ตัวแบบกึ่งเชิงประจักษ์แบบ Linear, Quadratic และ Exponential ผลปรากฏว่าในการวัดความแม่นยำของการพยากรณ์ด้วย MAPE ของการพยากรณ์ความต้องการน้ำมันเบนซินด้วยตัวแบบ Modified Gaussian ให้ความแม่นยำมากที่สุดที่ 0.075% ส่วนของน้ำมันดีเซลนั้นตัวแบบ Linear ให้ความแม่นยำมากที่สุดที่ 11.821% และตัวแบบสำหรับก๊าซปิโตรเลียมเหลวที่ให้ความแม่นยำมากที่สุดคือตัวแบบ Quadratic ที่ 19.560% แต่ผู้วิจัยเห็นว่างานวิจัยนี้ซึ่งใช้ตัวแบบกึ่งเชิงประจักษ์ในการพยากรณ์ ยังขาดการศึกษาปัจจัยที่ส่งผลต่อความต้องการน้ำมันสำเร็จรูป

Liddle และ Huntington (2020) ได้ศึกษาความสัมพันธ์ระหว่างปัจจัยด้านความเจริญเติบโตทางเศรษฐกิจ ราคาน้ำมันเชื้อเพลิง กับความต้องการน้ำมันเบนซินและน้ำมันดีเซลในภาคการขนส่งด้วยการรวบรวมชุดข้อมูลที่เรียงตามเวลา (Wide Panel Dataset) ของปริมาณการใช้น้ำมันเชื้อเพลิงและราคาสำหรับกลุ่มประเทศที่เป็นสมาชิกองค์การเพื่อความร่วมมือและการพัฒนาทางเศรษฐกิจ (Organisation for Economic Co-operation and Development, OECD) จำนวน 35 ประเทศ และกลุ่มประเทศที่ไม่ใช่ OECD จำนวน 83 ประเทศ มีช่วงเวลาของข้อมูลแบบรายปีตั้งแต่ ค.ศ. 1978 ถึง ค.ศ. 2016 ซึ่งแต่ละประเทศมีช่วงเวลาไม่เท่ากัน โดยใช้การประมาณค่าอุปสงค์หรือความต้องการน้ำมันเบนซินและน้ำมันดีเซลด้วยตัวแบบ Dynamic Adjustment Model ที่ไม่ใช่เชิงเส้น ซึ่งตัวแปรตาม (การบริโภคน้ำมันเชื้อเพลิงต่อหัวประชากร) ที่ข้อมูลย้อนหลังที่เวลาย้อนไป  $t-1$  คาบ (1 Time lag) อยู่ด้านขวาของสมการตัวแบบพร้อมกับตัวแปรอิสระดังนี้ ผลิตภัณฑ์มวลรวมของประเทศต่อหัวประชากร (GDP per capita) ราคาน้ำมันเชื้อเพลิง และค่าข้อมูลย้อนหลังที่เวลาย้อนไป  $t-1$  คาบของแต่ละตัวแปรอิสระ ผลลัพธ์ที่ได้คือน้ำมันเบนซินมีความยืดหยุ่น (Elasticity) ของราคาเชื้อเพลิงของกลุ่มประเทศ OECD มีค่าประมาณ -0.7 สูงกว่ากลุ่มที่ไม่ใช่ OECD สามเท่า น้ำมันดีเซลมีความยืดหยุ่นของราคาเชื้อเพลิงของกลุ่มประเทศ OECD สูงกว่ากลุ่มที่ไม่ใช่ OECD ไม่มาก (-0.3 และ -0.2 ตามลำดับ) สำหรับความยืดหยุ่นของ GDP น้ำมันเบนซิน กลุ่มที่ไม่ใช่ OECD มีค่าประมาณ 1.0 หรือสูงกว่ากลุ่ม OECD ประมาณสองเท่า สำหรับกลุ่ม OECD ความยืดหยุ่น GDP ของน้ำมันดีเซลมีค่าเป็น

สามเท่าของความยืดหยุ่น GDP ของน้ำมันเบนซิน แต่อย่างไรก็ตามสำหรับกลุ่มที่ไม่ใช่ OECD ความยืดหยุ่น GDP ของทั้งน้ำมันเบนซินและน้ำมันดีเซลมีค่าเท่ากัน

Karakurt (2021) ศึกษาการพยากรณ์การบริโภคน้ำมันของกลุ่มประเทศ BRICS-T ซึ่งเป็นกลุ่มประเทศเศรษฐกิจเกิดใหม่และกำลังเติบโตโดยใช้ตัวแบบการถดถอยเชิงเส้นและใช้ข้อมูลอนุกรมเวลา ตั้งแต่ปี ค.ศ. 1965 ถึง ค.ศ. 2019 ซึ่งตัวแบบอิสระที่เลือกใช้ได้แก่ จำนวนประชากรทั้งหมด จำนวนประชากรในเมือง และผลิตภัณฑ์มวลรวมของประเทศต่อหัวประชากร จากนั้นตัวแบบที่ได้จะถูกยืนยันด้วยวิธีการทางสถิติที่ประกอบด้วยการกำหนดค่าสัมประสิทธิ์ การทดสอบ F และ t และข้อมูลที่คาดการณ์กับข้อมูลจริง ในการวิเคราะห์ความคลาดเคลื่อนของตัวแบบจะใช้วิธี MAPE RMSE MAD และ PER นอกจากนี้ยังนำเสนอปริมาณการบริโภคน้ำมันในอนาคตจำนวน 5 10 15 และ 20 ปี ต่อไปจากตัวแบบที่ได้

Azadeh et al. (2009) ได้ศึกษาการพยากรณ์การบริโภคน้ำมันโดยใช้อัลกอริทึม Flexible Fuzzy Regression ที่ขึ้นอยู่กับตัวชี้วัดทางเศรษฐกิจมาตรฐาน ซึ่งได้แก่ จำนวนประชากรรายปี ค่าใช้จ่ายในการนำเข้าน้ำมันดิบ ผลิตภัณฑ์มวลรวมในประเทศ (GDP) และปริมาณการผลิตน้ำมัน ในคาบเวลาล่าสุดในตอนนั้น อัลกอริทึมที่เสนอใช้การวิเคราะห์ความแปรปรวน (ANOVA) ในการคัดเลือกระหว่าง Fuzzy Regression หรือ Conventional Regression สำหรับการประมาณค่าอุปสงค์ในอนาคต นัยสำคัญของอัลกอริทึมที่เสนอมี 3 อย่าง 1) มีความยืดหยุ่นและระบุตัวแบบที่ดีที่สุดที่ขึ้นอยู่กับผลลัพธ์ของ ANOVA และค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) อย่างไรก็ตามในการศึกษาที่ผ่านมามีการพิจารณาตัวแบบที่ดีที่สุดของ Fuzzy Regression ขึ้นอยู่กับ MAPE หรือผลลัพธ์ของค่าความผิดพลาดสัมพัทธ์แบบอื่น 2) ตัวแบบที่เสนออาจจะระบุให้ Conventional Regression เป็นตัวแบบที่ดีที่สุดสำหรับการพยากรณ์การบริโภคน้ำมันในอนาคตเนื่องจากมีโครงสร้างแบบพลวัต (Dynamic Structure) ในกรณีที่การศึกษาที่ผ่านมากำหนดสมมติฐานว่า Fuzzy Regression ให้คำตอบและประมาณค่าได้ดีที่สุดเสมอ 3) มีการใช้ประโยชน์จากตัวแปรอิสระมาตรฐานเป็นส่วนใหญ่สำหรับตัวแบบการถดถอย (Regression Model) ในตัวแบบ Flexible Fuzzy Regression ได้ใช้ข้อมูลสำหรับการบริโภคน้ำมันในประเทศแคนาดา สหรัฐอเมริกา ญี่ปุ่น และออสเตรเลีย จาก ค.ศ. 1990 ถึง ค.ศ. 2005 ผลลัพธ์ที่ได้แสดงให้เห็นว่าตัวแบบ Flexible Fuzzy Regression ให้คำตอบที่แม่นยำสำหรับปัญหาการประมาณค่าการบริโภคน้ำมัน ซึ่งตัวแบบนี้อาจจะถูกใช้โดยผู้กำหนดนโยบายเพื่อคาดการณ์พฤติกรรมการใช้น้ำมันในภูมิภาคต่าง ๆ อย่างแม่นยำ

Al-Fattah (2020) ศึกษาและพัฒนาตัวแบบปัญญาประดิษฐ์ที่ใช้คาดการณ์ด้วยการขับเคลื่อนด้วยข้อมูลในการวิเคราะห์และพยากรณ์ความต้องการน้ำมันเบนซินของประเทศซาอุดีอาระเบีย ตัวแบบปัญญาประดิษฐ์นี้มีพื้นฐานจากขั้นตอนวิธีทางพันธุกรรม (Genetic Algorithm, GA) โครงข่ายประสาทเทียม (Artificial Neural Network, ANN) และการทำเหมืองข้อมูล (Data Mining, DM)

สำหรับการวิเคราะห์อนุกรมเวลา ตัวแบบนี้เรียกว่า GANNATS ซึ่งตัวแบบคาดการณ์ GANNATS ถูกออกแบบ สร้าง ตรวจสอบและทดสอบโดยใช้ข้อมูลเศรษฐกิจและพลังงานในอดีตรายปีจากปี ค.ศ. 1970 ถึง ค.ศ. 2016 ประกอบด้วยข้อมูลดังนี้ ความต้องการน้ำมันเบนซิน ราคาน้ำมันเบนซินและน้ำมันดีเซล จำนวนประชากร ผลิตภัณฑ์มวลรวมในประเทศ จำนวนรถยนต์ขนาดเล็กและขนาดใหญ่ (Light-duty and Heavy-duty Vehicles) ในงานวิจัยนี้พัฒนาโครงร่าง (Framework) โดยออกแบบระเบียบวิธีซึ่งประกอบด้วยขั้นตอนดังต่อไปนี้ 1) การเตรียมและการจัดเก็บข้อมูล (Data Warehousing and Preparation) 2) การทำเหมืองข้อมูลและกระบวนการก่อนการวิเคราะห์ข้อมูล (Data Mining and Preprocessing) 3) ตัวแปรที่ส่งผลกระทบต่อและการวิเคราะห์การคัดเลือก (Variables Impact and Selection Analysis) 4) การออกแบบตัวแบบ (Model Design) 5) การประเมินสมรรถนะของตัวแบบ (Model Performance Evaluation) 6) การปรับแต่งและการหาค่าที่เหมาะสมของตัวแบบ (Model Optimization and Fine-tuning) และ 7) กระบวนการหลังการวิเคราะห์ข้อมูลและการทำให้เห็นผลลัพธ์ของตัวแบบในรูปแบบที่เห็นได้ (Postprocessing and Visualization of Model Results) ผลลัพธ์แสดงให้เห็นถึงตัวแบบที่ให้การพยากรณ์ที่แม่นยำกับตัวชี้วัดสมรรถนะหลักที่แข็งแกร่ง (Robust Key Performance Indicators) เช่น มีค่า MAPE อยู่ที่ 0.0792% โดยความต้องการน้ำมันเบนซินของซาอุดีอาระเบียที่ลดลงล่าสุดมีสาเหตุหลักมาจากการปรับปรุงประสิทธิภาพของยานพาหนะ การเพิ่มการอุดหนุนราคาน้ำมัน การลดลงของอัตราการการเพิ่มขึ้นของประชากรและการเปลี่ยนแปลงพฤติกรรมของผู้บริโภค ซึ่งงานวิจัยนี้ได้นำเสนอวิธีการปรับปรุงแบบจำลองเศรษฐกิจแบบดั้งเดิมและเพิ่มความสามารถในการพยากรณ์และความแม่นยำของการคาดการณ์ความต้องการน้ำมันเบนซิน

Sagheer และ Kotb (2019) ได้เสนอแนวทางการพยากรณ์การผลิตผลิตภัณฑ์ปิโตรเลียมรายเดือนด้วยตัวแบบการเรียนรู้เชิงลึก (Deep Learning) ที่สามารถระบุข้อจำกัดของแนวทางการพยากรณ์แบบเดิมและแสดงการคาดการณ์ที่แม่นยำ แนวทางที่เสนอคือสถาปัตยกรรมหน่วยความจำระยะสั้นระยะยาว (Deep Long-Short Term Memory, DLSTM) ซึ่งเป็นส่วนขยายของเครือข่ายประสาทที่เกิดขึ้นซ้ำแบบดั้งเดิม (Recurrent Neural Network, RNN) ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm, GA) ถูกนำไปใช้เพื่อกำหนดค่าสถาปัตยกรรมที่เหมาะสมที่สุดของ DLSTM นอกจากนี้ยังใช้วิธีการพยากรณ์ด้วยอนุกรมเวลาด้วยตัวแบบอื่น ๆ เพื่อการเปรียบเทียบผลการพยากรณ์ เช่น ตัวแบบ ARIMA ตัวแบบ RNN ตัวแบบ Deep Gated Recurrent Unit (DGRU) และตัวแบบ Nonlinear Extension for Linear Arps Decline (NEA) งานวิจัยนี้วัดประสิทธิภาพของตัวแบบด้วยวิธี Mean Square Error (MSE) Root Mean Square Error (RMSE) และ Mean Absolute Percentage Error (MAPE) ผลการวิจัยแสดงให้เห็นว่าตัวแบบ DLSTM ที่เสนอมีประสิทธิภาพเหนือกว่าตัวแบบอื่น ๆ ที่ถูกนำไปเปรียบเทียบ

## 2.2.2 การพยากรณ์ความต้องการด้วยเทคนิคการเรียนรู้ของเครื่อง

Xiao et al. (2018) นำเสนอกลุ่มของวิธีการของเทคนิคการจัดการข้อมูลสำหรับการพยากรณ์การบริโภคพลังงานในประเทศจีน และสร้างตัวแบบการพยากรณ์แบบผสม (Hybrid Forecasting Model) ที่มีพื้นฐานมาจากกลุ่มของวิธีการ Ensemble ที่เลือก โดยมุ่งสนใจที่การพยากรณ์การบริโภคพลังงานที่มีการเปลี่ยนแปลงแบบไม่เชิงเส้น ตัวแบบจะคาดการณ์แนวโน้มเชิงเส้นของอนุกรมเวลาการบริโภคพลังงานเป็นอย่างแรกผ่านกลุ่มวิธีการจัดการข้อมูลที่ขึ้นอยู่กับตัวแบบ Autoregressive และหลังจากนั้นจะได้รับชุดข้อมูลค่าความผิดพลาด (Residual) ของการบริโภคพลังงาน ในการพิจารณาคุณลักษณะแบบไม่เชิงเส้นของค่าความผิดพลาดในงานวิจัยนี้เสนอให้ใช้เทคนิค AdaBoost Ensemble ในการปรับปรุงสมรรถนะการพยากรณ์ของตัวแบบการพยากรณ์ไม่เชิงเส้นเชิงเดี่ยว (Single Nonlinear Prediction Model) ได้แก่ Back Propagation Neural Network, Support Vector Regression Machine, Genetic Programming และ Radical Basis Function Neural Network ตามลำดับเพื่อให้ได้ตัวแบบ Ensemble แบบไม่เชิงเส้นจำนวน 4 กลุ่ม ผลลัพธ์ของการพยากรณ์ของตัวแบบ AdaBoost Ensemble ทั้ง 4 ถูกใช้เป็นค่าป้อนเข้าเริ่มต้น (Initial Input) และจะได้ผลการคาดการณ์แบบผสมผสานที่ถูกเลือกสำหรับกลุ่มไม่เชิงเส้นจากการใช้กลุ่มวิธีการของการจัดการข้อมูล สุดท้ายคือการรวมทั้งสองส่วนเข้าด้วยกันเพื่อให้ได้ผลการพยากรณ์ขั้นสุดท้ายเรียกตัวแบบนี้ว่า HFGSE ในการวิเคราะห์เชิงประจักษ์ (Empirical Analysis) ของการใช้พลังงานทั้งหมดและปริมาณการใช้น้ำมันทั้งหมดในจีนแสดงให้เห็นว่าสมรรถนะการพยากรณ์ของตัวแบบ HFGSE ที่นำเสนอที่ดีกว่ากลุ่มวิธีการจัดการข้อมูลที่ขึ้นอยู่กับตัวแบบ Autoregressive และตัวแบบผสมอื่น ๆ ทั้ง 7 ตัวแบบ

Lu et al. (2019) ได้นำตัวแบบผสมที่รวมตัวแบบ Fruit Fly Optimization Algorithm (FFOA) Simulated Annealing Algorithm (SA) Cross Factor (CF) และ Support Vector Machine (SVM) เข้าด้วยกันเพื่อพยากรณ์โหลดการใช้งานระยะสั้นของก๊าซในเมือง (Urban Gas) ในตัวแบบนั้น SA และ CF ถูกใช้ในการหาค่าเหมาะสมให้กับอัลกอริทึม FFOA งานวิจัยนี้ใช้ระบบก๊าซในเมืองคุนหมิง ประเทศจีนเป็นกรณีศึกษาและใช้อัลกอริทึม CF-SA-FFOA-SVM ในการพยากรณ์การบริโภคก๊าซและเปรียบเทียบผลกับตัวแบบการพยากรณ์อื่น 4 ตัวแบบ ได้แก่ Back-Propagation Neural Network (BPNN) Autoregressive Integrated Moving Average Model (ARIMA) Particle Swarm Optimization-Support Vector Machine (PSO-SVM) และ Genetic Algorithm (GA) นอกจากนี้ยังวิเคราะห์อิทธิพลของอุณหภูมิประเภทต่าง ๆ (อุณหภูมิสูงสุดรายวันและอุณหภูมิต่ำสุดรายวัน) ในตัวแบบด้วย ในผลการพยากรณ์สรุปได้ว่า 1) เมื่อเปรียบเทียบกับตัวแบบอื่น ตัวแบบ CF-SA-FFOA-SVM ให้ผลการพยากรณ์โหลดการใช้งานก๊าซแม่นยำมากที่สุด 2) สำหรับเมืองคุนหมิง



ถ้าอุณหภูมิสูงสุดรายวันถูกใช้เป็นตัวแปรป้อนเข้าในตัวแบบ ผลการพยากรณ์จะมีความแม่นยำมากขึ้น

3) การจัดกลุ่มข้อมูลตามลักษณะวันหยุดหรือประเภทก๊าซสามารถปรับปรุงความแม่นยำของการพยากรณ์ไหลลดการใช้ก๊าซได้อย่างมีประสิทธิภาพ

Maehashi และ Shintani (2020) ได้ทำการวิเคราะห์เปรียบเทียบอย่างละเอียดเกี่ยวกับแบบจำลองปัจจัยและการเรียนรู้ของเครื่องเพื่อพยากรณ์อนุกรมเวลาเศรษฐกิจมหภาคของประเทศญี่ปุ่น ผลลัพธ์หลักที่ได้สามารถสรุปได้ดังนี้ 1) ในหลาย ๆ กรณี ตัวแบบปัจจัย (Factor Model) และตัวแบบการเรียนรู้ของเครื่องใช้พยากรณ์ได้ดีกว่าตัวแบบ Univariate Autoregressive 2) การพยากรณ์ด้วยวิธีการเรียนรู้ของเครื่องทำงานได้ดีโดยเฉพาะอย่างยิ่งสำหรับการพยากรณ์ระยะกลางถึงระยะยาว 3) ความสำเร็จของการเรียนรู้ของเครื่องส่วนใหญ่มาจากความไม่เป็นเชิงเส้นและปฏิสัมพันธ์ของตัวแปรซึ่งแสดงให้เห็นถึงความสำคัญของโครงสร้างที่ไม่เป็นเชิงเส้นในการพยากรณ์อนุกรมเวลาเศรษฐกิจมหภาคของญี่ปุ่น 4) การพยากรณ์แบบผสมของตัวแบบปัจจัยและตัวแบบการเรียนรู้ของเครื่องทำงานได้ดีกว่าตัวแบบปัจจัยหรือตัวแบบการเรียนรู้ของเครื่องเพียงอย่างเดียว 5) วิธีการเรียนรู้ของเครื่องถูกใช้ในการหาปัจจัยทั่วไปที่มีประโยชน์ในการพยากรณ์แบบผสม และงานวิจัยนี้ใช้การวัดประสิทธิภาพการพยากรณ์ด้วย Mean Square Forecast Errors (MSFEs)

Al-Dosary et al. (2019) ศึกษาการกำหนดความสามารถของอัลกอริทึม K-nearest Neighbor Algorithm (KNN) ในการพยากรณ์การใช้เชื้อเพลิงของระบบแทรกเตอร์ไถสีว มีพารามิเตอร์ป้อนเข้า ได้แก่ กำลังของแทรกเตอร์ ความกว้างการไถ ความลึกและความเร็วการไถ ร้อยละปริมาณของทราย ตะกอน และโคลนในดิน ความชื้นเริ่มต้นของดิน และความหนาแน่นเริ่มต้นของดิน ประสิทธิภาพของการพยากรณ์ด้วยวิธี KNN จะถูกเปรียบเทียบกับวิธี Multiple Linear Regression (MLR) และวัดประสิทธิภาพด้วยวิธี MAE และ RMSE ผลลัพธ์ที่ได้พบว่าการพยากรณ์ด้วยวิธี KNN ให้ประสิทธิภาพที่ดีกว่าวิธี MLR

Kristjanpoller et al. (2020) นำเสนอวิธีการรวมกลุ่ม (Ensemble Methodology) ซึ่งประกอบด้วยตัวแบบ ARIMA ตัวแบบ ANN ตัวแบบ Fuzzy Inference System ตัวแบบ Adaptive ตัวแบบ SVR ตัวแบบ Extreme Machine Learning และตัวแบบ GA เพื่อพยากรณ์ผลรวมความต้องการพลังงานในระยะยาว และใช้การวัดประสิทธิภาพของตัวแบบด้วยวิธี MSE และ MAPE จากนั้นทดสอบความเชื่อมั่นด้วยวิธีการทางสถิติพบว่าวิธีการรวมกลุ่มที่นำเสนอช่วยเพิ่มความแม่นยำในการคาดการณ์มากกว่าวิธีการก่อนที่จะนำมารวมกลุ่มกัน แนวทางที่พัฒนาขึ้นในงานวิจัยนี้ให้ข้อมูลเพิ่มเติมเกี่ยวกับวิธีการผสมตัวแบบและการทำให้การพยากรณ์มีความแม่นยำมากขึ้น

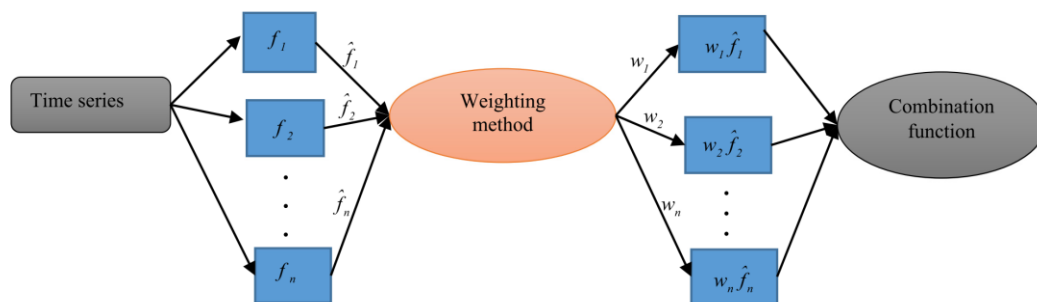
Hajirahimi และ Khashei (2019) ได้รวบรวมและวิเคราะห์โครงสร้างตัวแบบผสมโดยการสำรวจงานวิจัยมากกว่า 150 งานวิจัยที่ใช้ตัวแบบผสมหลายตัวแบบในการสร้างตัวแบบอนุกรมเวลา และตัวแบบที่เกี่ยวข้องกับการพยากรณ์ โดยข้อดีหลักของตัวแบบผสมที่ระบุไว้ในงานวิจัยที่เกี่ยวข้อง

สรุปได้เป็นสามประเด็นสำคัญ ได้แก่ 1) ปรับปรุงความแม่นยำของการพยากรณ์ให้ดีขึ้นเนื่องจากการตรวจจํารูปแบบข้อมูลและการสร้างตัวแบบที่ครอบคลุม 2) ลดความเสี่ยงจากการใช้แบบจําลองที่ไม่เหมาะสมเนื่องจากการใช้การพยากรณ์แบบรวมกัน (Combination of Forecasts) 3) ลดความซับซ้อนของขั้นตอนการคัดเลือกตัวแบบเนื่องจากการใช้ส่วนประกอบที่แตกต่างกัน ในงานวิจัยนี้ได้แบ่งกลุ่มวิธีผสมตัวแบบออกเป็น 4 กลุ่มหลัก ได้แก่ 1) ตัวแบบผสมที่ใช้การประมวลผลข้อมูลล่วงหน้า (Data Preprocessing Based Hybrid Models) 2) ตัวแบบผสมที่ใช้การเพิ่มประสิทธิภาพพารามิเตอร์ของตัวแบบ (Parameters Optimization Based Hybrid Models) 3) ตัวแบบผสมที่มีการพื้นฐานจากการผสมผสานส่วนประกอบของตัวแบบ (Component Combination Based Hybrid Models) 4) ตัวแบบผสมที่มาจากการประมวลผลทีหลัง (Postprocessing Based Hybrid Models) โดยที่ตัวแบบผสมที่ใช้การประมวลผลข้อมูลล่วงหน้าและตัวแบบผสมที่ใช้การเพิ่มประสิทธิภาพพารามิเตอร์ของตัวแบบเป็นตัวแบบผสมที่ได้รับความนิยมและใช้กันอย่างแพร่หลายตามลำดับ และตัวแบบผสมที่มาจากการประมวลผลทีหลังมีการใช้งานน้อยที่สุด ในงานวิจัยนี้ผู้วิจัยเลือกที่จะศึกษาตัวแบบผสม (Hybrid Model) แบบที่ใช้ตัวแบบผสมที่มีการพื้นฐานจากการผสมผสานส่วนประกอบของตัวแบบ เนื่องจากต้องการปรับปรุงความแม่นยำของการพยากรณ์ให้ดีขึ้นและยังสามารถช่วยลดความเสี่ยงจากการใช้แบบจําลองที่ไม่เหมาะสม อีกทั้งในการวิจัยนี้มีการสร้างตัวแบบการเรียนรู้ของเครื่องหลายตัวแบบ ซึ่งสามารถที่จะนำตัวแบบที่ให้ผลลัพธ์การพยากรณ์ที่ดีมาผสมผสานส่วนประกอบของแต่ละตัวแบบได้ ในงานวิจัยของ Hajirahimi และ Khashei (2019) สามารถแบ่งประเภทของตัวแบบผสมที่มีพื้นฐานจากการผสมผสานส่วนประกอบของตัวแบบได้เป็น 3 โครงสร้างดังนี้

1) โครงสร้างการผสมแบบขนาน (Parallel Hybrid Structure) เป็นโครงสร้างที่เป็นการนำผลลัพธ์ของค่าพยากรณ์ที่ได้จากแต่ละตัวแบบมารวมกันแบบถ่วงน้ำหนักโดยมีฟังก์ชันเชิงเส้นและฟังก์ชันไม่เชิงเส้นที่ใช้ถ่วงน้ำหนักที่หลากหลาย เช่น ใช้ค่าเฉลี่ย ใช้การวิเคราะห์การถดถอยเชิงเส้นหรือใช้วิธีการ Var-Cov เป็นต้น ดังรูปที่ 10 โดยมีรูปแบบสมการของโครงสร้างการผสมแบบขนานคือ

$$f_{combined,t} = \varphi(w_1 f_{1,t}, w_2 f_{2,t}, \dots, w_n f_{n,t}) \quad t = 1, 2, \dots, T \quad (2.21)$$

เมื่อ  $\varphi$  คือ ฟังก์ชันเชิงเส้นหรือฟังก์ชันไม่เชิงเส้นสำหรับการผสมตัวแบบ  
 $w_i f_i$  คือ ค่าการพยากรณ์แบบถ่วงน้ำหนักของตัวแบบที่  $i$  โดยที่  $i = 1, 2, \dots, n$   
 $T$  คือ จำนวนข้อมูลพื้นฐาน  
 $n$  คือ จำนวนตัวแบบพื้นฐานแต่ละตัวที่ใช้เป็นส่วนประกอบการผสม



รูปที่ 10 โครงสร้างการผสมแบบขนาน (Hajirahimi และ Khashei, 2019)

2) โครงสร้างการผสมแบบอนุกรม (Series Hybrid Structure) เป็นโครงสร้างการผสมที่ใช้แนวคิดของขั้นตอนการสร้างแบบจำลองตามลำดับ ดังรูปที่ 11 โดยการแยกอนุกรมเวลาออกเป็นส่วนที่เป็นเชิงเส้นและส่วนที่ไม่เป็นเชิงเส้น จากนั้นจึงนำสร้างตัวแบบที่ประกอบด้วยตัวแบบที่สามารถจับรูปแบบเชิงเส้น เช่น ตัวแบบ SARIMA และตัวแบบที่สามารถจับรูปแบบไม่เชิงเส้น เช่น ตัวแบบ ANN ตัวแบบ SVR เป็นต้น ซึ่งจะมีลำดับขั้นตอนการสร้างตัวแบบ 2 แบบคือ การสร้างตัวแบบตามลำดับแบบเชิงเส้นต่อด้วยแบบไม่เชิงเส้น (Linear-Nonlinear Sequential Modeling) เช่น การสร้างตัวแบบ SARIMA เพื่อจับรูปแบบเชิงเส้นของอนุกรมเวลา จากนั้นจึงนำค่าส่วนเหลือ (Residual) ไปเป็นข้อมูลป้อนเข้าตัวแบบ ANN เพื่อจับรูปแบบไม่เชิงเส้นของอนุกรมเวลา จากนั้นจึงนำค่าการพยากรณ์ทั้ง 2 ตัวแบบรวมกัน ซึ่งจะเรียกว่าตัวแบบ SARIMA-ANN และแบบที่ 2 คือ การสร้างตัวแบบตามลำดับแบบไม่เชิงเส้นต่อด้วยแบบเชิงเส้น (Nonlinear-Linear Sequential Modeling) ซึ่งจะมีลำดับการสร้างตัวแบบที่ตรงกันข้ามกับแบบแรก เช่น การสร้างตัวแบบ ANN เพื่อจับรูปแบบไม่เชิงเส้นของอนุกรมเวลา จากนั้นจึงนำค่าส่วนเหลือ (Residual) ไปเป็นข้อมูลป้อนเข้าตัวแบบ SARIMA เพื่อจับรูปแบบเชิงเส้นของอนุกรมเวลา จากนั้นจึงนำค่าการพยากรณ์ทั้ง 2 ตัวแบบรวมกัน ซึ่งจะเรียกว่าตัวแบบ ANN-SARIMA สำหรับโครงสร้างการผสมแบบขนานรูปแบบสมการของคือ

$$f_{combined,t} = \hat{f}_{1,t} + \hat{f}'_{2,t} + \hat{f}'_{3,t} + \dots + \hat{f}'_{n,t} \quad t = 1, 2, \dots, T \quad (2.22)$$

เมื่อ  $\hat{f}_i$  คือ ตัวแบบเดี่ยวแต่ละตัวแบบลำดับที่  $i$  โดยที่  $i = 1, 2, \dots, n$

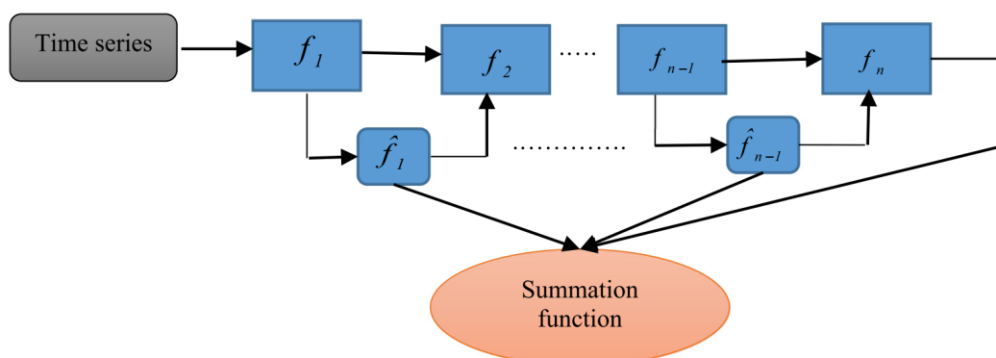
$\hat{f}_1$  คือ ค่าพยากรณ์ของตัวแบบแรกที่ได้จากการป้อนเข้าจากข้อมูลดิบ

$\hat{f}'_2$  คือ ค่าพยากรณ์ของตัวแบบที่สองที่ได้จากการป้อนเข้าจากค่าพยากรณ์ของตัวแบบแรก

$f'_n$  คือ ค่าพยากรณ์ของตัวแบบสุดท้ายที่ได้จากการป้อนเข้าจากค่าพยากรณ์ของตัวแบบก่อนหน้า

$T$  คือ จำนวนข้อมูลพื้นฐาน

$n$  คือ จำนวนตัวแบบพื้นฐานแต่ละตัวแบบ



รูปที่ 11 โครงสร้างการผสมแบบอนุกรม (Hajirahimi และ Khashei, 2019)

3) โครงสร้างการผสมแบบขนาน-อนุกรม (Parallel-Series Hybrid Structure) เป็นโครงสร้างการผสมโดยการผสมรูปแบบโครงสร้างการผสมแบบขนานและแบบอนุกรมเข้าด้วยกันเพื่อสกัดเอาข้อดีของทั้งสองโครงสร้าง ซึ่งจะมีโครงสร้างที่ซับซ้อนมากขึ้น

ผู้วิจัยเลือกศึกษาการผสมตัวแบบที่มีการพื้นฐานจากการผสมผสานส่วนประกอบของตัวแบบตามโครงสร้างการผสมแบบอนุกรมและโครงสร้างการผสมแบบขนานตามวิธีการที่ได้ศึกษาในงานวิจัยของ Hajirahimi และ Khashei (2019) โดยการใช้โครงสร้างการผสมแบบขนานสามารถช่วยลดความเสี่ยงจากการใช้แบบจำลองที่ไม่เหมาะสม และการใช้โครงสร้างการผสมแบบอนุกรมสามารถช่วยปรับปรุงความแม่นยำของการพยากรณ์ให้ดีขึ้นเนื่องจากการตรวจจบบรูปแบบข้อมูลและการสร้างตัวแบบที่ครอบคลุม ซึ่งงานของผู้วิจัยที่ได้ศึกษาเพิ่มเติมมีความแตกต่างจากงานวิจัยอื่น ๆ คือผู้วิจัยจะนำผลการพยากรณ์ของการสร้างตัวแบบผสมตามโครงสร้างการผสมแบบอนุกรมจำนวน 2 ตัวแบบคือตัวแบบ SARIMAX-ANN และตัวแบบ SARIMAX-SVR มาผสมกันตามโครงสร้างการผสมแบบขนานด้วยการรวมผลการพยากรณ์ด้วยวิธีการถดถอยเชิงเส้นพหุคูณ ซึ่งผู้วิจัยคิดว่าจะช่วยตรวจจบบรูปแบบข้อมูลและสร้างตัวแบบที่ครอบคลุมรูปแบบข้อมูลได้ดีขึ้น โดยคาดว่าจะให้ผลการพยากรณ์ที่แม่นยำมากขึ้นและมีความความแกร่ง (Robust) มากขึ้น

สำหรับตัวแบบผสมที่ใช้โครงสร้างการผสมแบบขนาน ผู้วิจัยเลือกใช้ฟังก์ชันการผสมแบบเส้นตรงโดยจะรวมตัวแบบเดี่ยวด้วยวิธี Random Sample Consensus (RANSAC) ซึ่งเป็นวิธีการวน

ซ้ำ (Iterative Method) สำหรับการประมาณค่าพารามิเตอร์ของแบบจำลองทางคณิตศาสตร์จากชุดข้อมูลป้อนเข้าที่มีค่าผิดปกติ (Outlier) และในงานวิจัยนี้ผู้วิจัยเลือกใช้แบบจำลองทางคณิตศาสตร์เป็นตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับการประมาณค่าพารามิเตอร์ของตัวแบบ หรือเรียกว่า RANSAC Regression นอกจากนี้ผู้วิจัยมีความสนใจที่จะศึกษาผลลัพธ์ของการพยากรณ์ตัวแบบเดี่ยวที่ใช้ตัวแบบ RANSAC Regression ซึ่งผู้วิจัยคาดว่าตัวแบบ RANSAC Regression น่าจะเป็นตัวแบบที่แกร่งมากขึ้นจากตัวแบบการถดถอยเชิงเส้นพหุคูณเนื่องจากตัวแบบ RANSAC Regression จะหลีกเลี่ยงการนำค่าผิดปกติของข้อมูลป้อนเข้ามาใช้ในการประมาณค่าพารามิเตอร์ของตัวแบบ โดยผู้วิจัยได้ศึกษาตัวอย่างงานวิจัยที่มีการนำ RANSAC ไปประยุกต์ใช้ ได้แก่ 1) งานวิจัยของ Li และ Gans (2017) ที่ได้นำอัลกอริทึม RANSAC ที่มีความแกร่งและมีประสิทธิภาพไปใช้สร้างและติดตามตัวแบบที่มีค่าผิดปกติและค่าผิดพลาดมาก (Heavy Noise) ซึ่งผลลัพธ์พบว่าการพยากรณ์ด้วย RANSAC ให้ผลลัพธ์ที่ดีในเรื่องของความแม่นยำในการประมาณค่าและใช้เวลาในการคำนวณน้อย 2) งานวิจัยของ Ghahremani et al. (2021) ได้นำ RANSAC ไปใช้การวิเคราะห์ข้อมูลของกลุ่มจุดสามมิติที่เก็บค่าตำแหน่งขอบของวัตถุใน 3 มิติ หรือเรียกว่า 3D Point Cloud Data แล้วใช้จำลองอวัยวะต่างๆ ของพืช ซึ่งการใช้อัลกอริทึม RANSAC ทำให้เทคนิคในการวิเคราะห์ข้อมูล 3D Point Cloud Data นั้นแข็งแกร่งขึ้นต่อค่าผิดปกติที่ไม่เป็นที่ต้องการ

จากการทบทวนวรรณกรรมที่เกี่ยวข้อง ผู้วิจัยได้สรุปเนื้อหาของงานวิจัยที่เกี่ยวข้องกับการพยากรณ์น้ำมันสำเร็จรูปและการพยากรณ์ความต้องการด้วยเทคนิคการเรียนรู้ของเครื่อง ดังแสดงในตารางที่ 1 และ 2 ตามลำดับ จะเห็นได้ว่าในการพยากรณ์ที่เกี่ยวข้องกับปริมาณการบริโภค น้ำมันสำเร็จรูปมีการใช้เทคนิคการพยากรณ์ที่หลากหลายทั้งวิธีการพยากรณ์ด้วยวิธีอนุกรมเวลาเชิงสถิติ วิธีการพยากรณ์เชิงสาเหตุ วิธีการพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง รวมถึงมีการผสมวิธีการพยากรณ์เพื่อให้ประสิทธิภาพของตัวแบบการพยากรณ์ดีขึ้นกว่าการใช้ตัวแบบการพยากรณ์ด้วยวิธีเดียว ดังนั้นในงานวิจัยนี้จึงเลือกใช้ตัวแบบอนุกรมเวลาเชิงสถิติเพื่อศึกษาอนุกรมเวลาของข้อมูลปริมาณการบริโภคน้ำมันสำเร็จรูป ตัวแบบการถดถอยเชิงเส้นพหุคูณเพื่อศึกษาปัจจัยที่ส่งผลต่อปริมาณการบริโภคน้ำมันสำเร็จรูปของประเทศไทย ตัวแบบการเรียนรู้ของเครื่องเพื่อศึกษาวิธีการพยากรณ์ที่ไม่ใช้สมมติฐานทางสถิติมาเกี่ยวข้อง และตัวแบบผสมเพื่อศึกษาวิธีการผสมตัวแบบต่าง ๆ และพัฒนาตัวแบบผสมให้มีประสิทธิภาพในการพยากรณ์ให้ดีขึ้น โดยศึกษาวิธีการพยากรณ์และเลือกตัวแบบที่เหมาะสมสำหรับการพยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทย โดยเปรียบเทียบผลการพยากรณ์ของแต่ละตัวแบบดังกล่าว เพื่อให้ได้แบบจำลองที่เหมาะสมและแม่นยำสำหรับการพยากรณ์นี้

ตารางที่ 1 สรุปการทบทวนวรรณกรรมที่เกี่ยวข้องการพยากรณ์น้ำมันสำเร็จรูป

ผู้เขียน	ข้อมูลที่พยากรณ์	ความถี่ข้อมูล	ตัวแปรอิสระที่ใช้	วิธีการพยากรณ์	วิธีวัดความแม่นยำ
Melikoglu (2014)	ความต้องการใช้น้ำมันเบนซิน ดีเซลและก๊าซปิโตรเลียมเหลวของประเทศตุรกี	รายปี	ไม่มี	- เบนซินใช้ตัวแบบ Gaussian, Modified Gaussian และ Lorentzian - ดีเซลและก๊าซปิโตรเลียมเหลวใช้ตัวแบบ Linear, Quadratic และ Exponential	RMSE และ MAPE
Liddle และ Huntington (2020)	ความต้องการน้ำมันเบนซินและน้ำมันดีเซลของกลุ่ม OECD 35 ประเทศ และไม่ใช่ OECD 83 ประเทศ	รายปี	GDP per capita ราคา น้ำมันเชื้อเพลิง และค่าข้อมูลย้อนหลังที่เวลาย้อนไป t-1 คาบของแต่ละตัวแปรอิสระ	Dynamic Adjustment Model ที่ไม่ใช่เชิงเส้น	RMSE
Karakurt (2021)	การบริโภคน้ำมันของกลุ่มประเทศ BRICS-T	รายปี	จำนวนประชากรทั้งหมด จำนวนประชากรในเมือง GDP per capita	ใช้ตัวแบบ Multiple Linear Regression	MAD, RMSE, PER และ MAPE
Azadeh et al. (2009)	การบริโภคน้ำมันของสหรัฐอเมริกา ญี่ปุ่น ออสเตรเลีย และแคนาดา (ค.ศ.1990 - 2005)	รายปี	จำนวนประชากรรายปี, ค่าใช้จ่ายในการนำเข้าน้ำมันดิบ ปริมาณการผลิตน้ำมัน และ GDP per capita	ใช้ตัวแบบ Fuzzy Regression และ Conventional Regression	MAPE
Al-Fattah (2020)	ความต้องการน้ำมันเบนซินของประเทศซาอุดีอาระเบีย ใช้ข้อมูล (ค.ศ.1970 - 2016)	รายปี	ราคาน้ำมันเบนซินและน้ำมันดีเซล จำนวนประชากร จำนวนรถยนต์ขนาดเล็กและขนาดใหญ่ และ GDP	ใช้ตัวแบบปัญญาประดิษฐ์ GANNATS ที่มีพื้นฐานจากขั้นตอนวิธีทางพันธุกรรม โครงข่ายประสาทเทียม และการทำเหมืองข้อมูล	MRPE, MAPE, MSE, RMSE และ NSE
Sagheer และ Kotb (2019)	ปริมาณการผลิตผลิตภัณฑ์ปิโตรเลียม	รายเดือน	ไม่มี	ใช้ตัวแบบ ARIMA, RNN, DGRU, NEA และ DLSTM	MSE, RMSE, และ MAPE
พิศาล (2564) *งานของผู้วิจัย	การบริโภคน้ำมันสำเร็จรูป (น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน) ของประเทศไทย	รายไตรมาส	ปริมาณการผลิตน้ำมันสำเร็จรูป ราคาน้ำมันสำเร็จรูป GDP จำนวนรถจดทะเบียนใหม่	ใช้ตัวแบบ Holt-Winters, SARIMA, SARIMAX, Multiple Linear Regression, KNN, SVR, ANN, Adaboost และ Hybrid	MAPE

ตารางที่ 2 สรุปการทบทวนวรรณกรรมที่เกี่ยวข้องการพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง

ผู้เขียน	ข้อมูลที่พยากรณ์	ความถี่ข้อมูล	ตัวแปรอิสระที่ใช้	วิธีการพยากรณ์	วิธีวัดความแม่นยำ
Xiao et al. (2018)	การบริโภคพลังงานในประเทศจีน ค.ศ.1978 - 2014	รายปี	ไม่มี	ตัวแบบการพยากรณ์แบบผสม: AdaBoost Ensemble, Back Propagation Neural Network, Support Vector Regression, Genetic Programming, และ Radical Basis Function Neural Network	RMSE และ MAPE
Lu et al. (2019)	พยากรณ์โหลดการใช้พลังงานระยะสั้นของก๊าซในเมือง	รายวัน	อุณหภูมิเฉลี่ยของอากาศ รายวัน สภาพอากาศ วันที่วันหยุด ปริมาณการใช้ก๊าซในเมืองย้อนหลัง 7 วัน	ตัวแบบการพยากรณ์แบบผสมที่รวม Fruit Fly Optimization Algorithm, Simulated Annealing Algorithm, Cross Factor และ Support Vector Machine เข้าด้วยกัน เทียบกับตัวแบบ PSO-SVM, BPNN, GM (1,1) และ ARIMA	RMSE, MAPE และ MSE
Maehashi และ Shintani (2020)	พยากรณ์อนุกรมเวลาเศรษฐกิจมหภาคของประเทศไทย ปีงบประมาณ 1973 ถึง มิถุนายน 2018	รายเดือน	ตัวแปรทางเศรษฐกิจมหภาค 219 ตัวแปร	- Regularized Least Squares Methods: Lasso, Ridge และ Elastic Net - Ensemble Machine Learning based on Regression Trees: Bagging, Random Forests, Boosting - Neural Networks: Feedforward Neural Networks, Convolutional Neural Networks, Recurrent Neural Networks	MSFEs
Al-Dosary et al. (2019)	พยากรณ์การใช้เชื้อเพลิงของระบบแทรกเตอร์ไถลั่ว	-	กำลังของแทรกเตอร์ ความกว้างการไถ ความลึกและความเร็วการไถ ร้อยละของทราย ตะกอน และโคลนในดิน ความชื้นเริ่มต้นของดิน และความหนาแน่นเริ่มต้นของดิน	K Nearest Neighbor และ Multiple Linear Regression	MAE และ RMSE
Kristjanpoller et al. (2020)	พยากรณ์ผลรวมความต้องการพลังงานในระยะยาว	รายเดือน	ไม่มี	Ensemble Model: ARIMA, ANN, Fuzzy Inference System, Adaptive, SVR, Extreme Machine Learning และ GA	MSE และ MAPE

## บทที่ 3

### วิธีการดำเนินงานวิจัย

#### 3.1 การเตรียมข้อมูล

งานวิจัยนี้จะใช้ข้อมูลการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักรายไตรมาสของประเทศไทยสำหรับสร้างและทดสอบตัวแบบการพยากรณ์เพื่อคาดการณ์ปริมาณการใช้ในอนาคต โดยมีน้ำมันสำเร็จรูป 3 ชนิด คือ น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน และผู้วิจัยได้ทบทวนวรรณกรรมเพื่อศึกษาตัวแปรอิสระที่มีอิทธิพลต่อการบริโภคน้ำมันสำเร็จรูปจากวรรณกรรมดังแสดงในตารางที่ 3

ตารางที่ 3 ตัวแปรอิสระที่มีอิทธิพลต่อการบริโภคน้ำมันสำเร็จรูปจากการทบทวนวรรณกรรม

ผู้วิจัย	ข้อมูลที่พยากรณ์	ตัวแปรอิสระที่ใช้
Azadeh et al. (2009)	การบริโภคน้ำมันของประเทศสหรัฐอเมริกา แคนาดา ญี่ปุ่น และออสเตรเลีย	จำนวนประชากรรายปี ค่าใช้จ่ายในการนำเข้า น้ำมันดิบ ผลิตภัณฑ์มวลรวมในประเทศต่อหัว ประชากร และปริมาณการผลิตน้ำมัน
Al-Fattah (2020)	ความต้องการน้ำมันเบนซินของประเทศ ซาอุดีอาระเบีย	ราคาน้ำมันเบนซินและน้ำมันดีเซล จำนวนประชากร ผลิตภัณฑ์มวลรวมในประเทศ จำนวนรถยนต์ขนาดเล็กและขนาดใหญ่
Karakurt (2021)	การบริโภคน้ำมันของกลุ่มประเทศ BRICS-T ซึ่งเป็นกลุ่มประเทศเศรษฐกิจเกิดใหม่และกำลังเติบโต	จำนวนประชากรทั้งหมด จำนวนประชากรในเมือง ผลิตภัณฑ์มวลรวมในประเทศต่อหัว ประชากร
Liddle และ Huntington (2020)	ความต้องการน้ำมันเบนซินและน้ำมันดีเซลของกลุ่มประเทศ OECD 35 ประเทศ และไม่ใช่ OECD 83 ประเทศ	ผลิตภัณฑ์มวลรวมในประเทศต่อหัวประชากร ราคาน้ำมันเบนซินและดีเซล และค่าข้อมูลย้อนหลังที่เวลาย้อนไปหนึ่งคาบเวลาของแต่ละตัวแปรอิสระ

ผู้วิจัยตัดสินใจเลือกใช้ตัวแปรอิสระที่อาจส่งผลกระทบต่อปริมาณการใช้น้ำมันสำเร็จรูปของประเทศไทยจากตัวแปรอิสระที่ถูกเลือกใช้จากวรรณกรรมที่ได้ศึกษาดังแสดงในตารางที่ 3 และจากข้อมูลที่ผู้วิจัยสามารถรวบรวมได้จากหน่วยงานภาครัฐ โดยมีตัวแปรอิสระดังต่อไปนี้

- 1) ข้อมูลราคาเฉลี่ยรายไตรมาสในประเทศไทยของน้ำมันดีเซลและน้ำมันเบนซินที่เกิดจากการนำข้อมูลรายเดือนที่ได้จากเว็บไซต์ของสำนักงานนโยบายและแผนพลังงานมาทำการเฉลี่ยให้เป็นรายไตรมาส
- 2) ข้อมูลปริมาณการผลิตน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินในประเทศไทยรายเดือน
- 3) ข้อมูลผลิตภัณฑ์มวลรวมในประเทศไทยรายไตรมาส



- 4) ข้อมูลจำนวนรถใหม่ที่จดทะเบียนตามกฎหมายว่าด้วยรถยนต์และกฎหมายว่าด้วยการขนส่งทางบกของประเทศไทย โดยรวมข้อมูลเป็นข้อมูลจำนวนรถใหม่ที่จดทะเบียนรายเดือน
- 5) ข้อมูลจำนวนประชากรรายปีที่ได้จากเว็บไซต์ของสำนักงานนโยบายและแผนพลังงาน งานวิจัยนี้เลือกการพยากรณ์เป็นแบบรายไตรมาสเพราะมุ่งศึกษาการพยากรณ์อนุกรมเวลาที่อาจมีผลของฤดูกาลด้วย จึงพยากรณ์ในระดับย่อยที่สุดได้ในระดับไตรมาส เพราะข้อมูลผลิตภัณฑ์มวลรวมในประเทศไทยซึ่งเป็นตัวแปรอิสระที่หลาย ๆ งานวิจัยศึกษาและพบว่ามิผลกระทบต่อปริมาณการใช้ น้ำมันสำเร็จรูปนั้นมีข้อมูลย่อยที่สุดที่รวบรวมได้เป็นแบบรายไตรมาส ข้อมูลปริมาณการบริโภค ปริมาณการผลิตของน้ำมันสำเร็จรูปและข้อมูลจำนวนรถใหม่ที่จดทะเบียนตามกฎหมายนั้นเป็นรายเดือนจึงนำแต่ละเดือนมารวมกันให้เป็นแบบรายไตรมาส ส่วนข้อมูลจำนวนประชากรนั้นข้อมูลที่รวบรวมมาได้มีข้อมูลย่อยที่สุดรายปีแต่เป็นตัวแปรที่ถูกศึกษาในหลาย ๆ งานวิจัยแล้วพบว่ามิผลต่อการบริโภคน้ำมันสำเร็จรูป จึงทำการประมาณค่าในช่วง (Interpolation) ให้เป็นรายไตรมาส

### 3.2 การพยากรณ์ด้วยตัวแบบอนุกรมเวลาเชิงสถิติ

การพยากรณ์ด้วยตัวแบบอนุกรมเวลาเชิงสถิติเป็นเทคนิคทั่วไปที่ใช้ได้ดีและไม่ซับซ้อนสำหรับพยากรณ์ที่ข้อมูลที่มีลักษณะเป็นอนุกรมเวลา ซึ่งงานวิจัยนี้จะใช้ตัวแบบอนุกรมเวลาเชิงสถิติสำหรับเปรียบเทียบกับวิธีการพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง

#### 3.2.1 ข้อมูลที่ใช้สำหรับการพยากรณ์ด้วยตัวแบบอนุกรมเวลาเชิงสถิติ

สำหรับการพยากรณ์ระยะสั้น ข้อมูลอนุกรมเวลารายไตรมาสของข้อมูลปริมาณการบริโภค น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน จะถูกใช้พยากรณ์แบบเลื่อนตำแหน่งช่วงเวลา ดังนี้

ช่วงที่ 1 ข้อมูลในปี พ.ศ. 2536 ถึงปี พ.ศ. 2552 จำนวน 68 ข้อมูลจะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน และ พ.ศ. 2553 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ

ช่วงที่ 2 ข้อมูลในปี พ.ศ. 2537 ถึงปี พ.ศ. 2553 จำนวน 68 ข้อมูลจะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน และ พ.ศ. 2554 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ

ช่วงที่ 3 ข้อมูลในปี พ.ศ. 2538 ถึงปี พ.ศ. 2554 จำนวน 68 ข้อมูลจะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน และ พ.ศ. 2555 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ

ช่วงที่ 4 ข้อมูลในปี พ.ศ. 2539 ถึงปี พ.ศ. 2555 จำนวน 68 ข้อมูลจะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน และ พ.ศ. 2556 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ

ช่วงที่ 5 ข้อมูลในปี พ.ศ. 2540 ถึงปี พ.ศ. 2556 จำนวน 68 ข้อมูลจะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน และ พ.ศ. 2557 จะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ

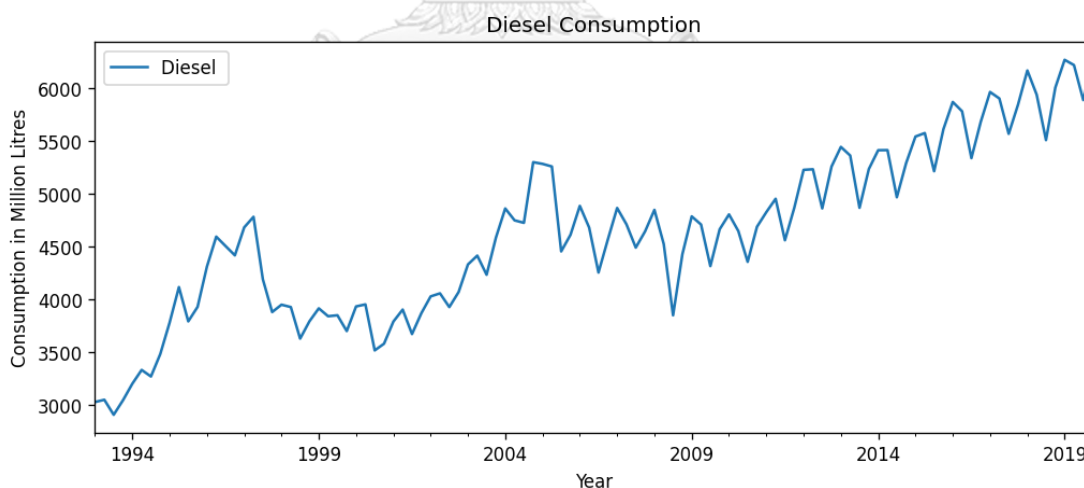
สำหรับการพยากรณ์ระยะกลาง ข้อมูลอนุกรมเวลารายไตรมาสของข้อมูลปริมาณการบริโภค น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินในช่วงเวลา พ.ศ. 2536 ถึง พ.ศ. 2552 จำนวนชุดข้อมูลละ 68 ข้อมูลจะถูกใช้เป็นชุดข้อมูลสำหรับฝึกสอน และข้อมูลในช่วงเวลา พ.ศ. 2553 ถึง พ.ศ. 2557 จำนวนชุดข้อมูลละ 20 ข้อมูลจะถูกใช้เป็นชุดข้อมูลสำหรับตรวจสอบ

### 3.2.2 การตรวจสอบและสำรวจข้อมูลเบื้องต้น

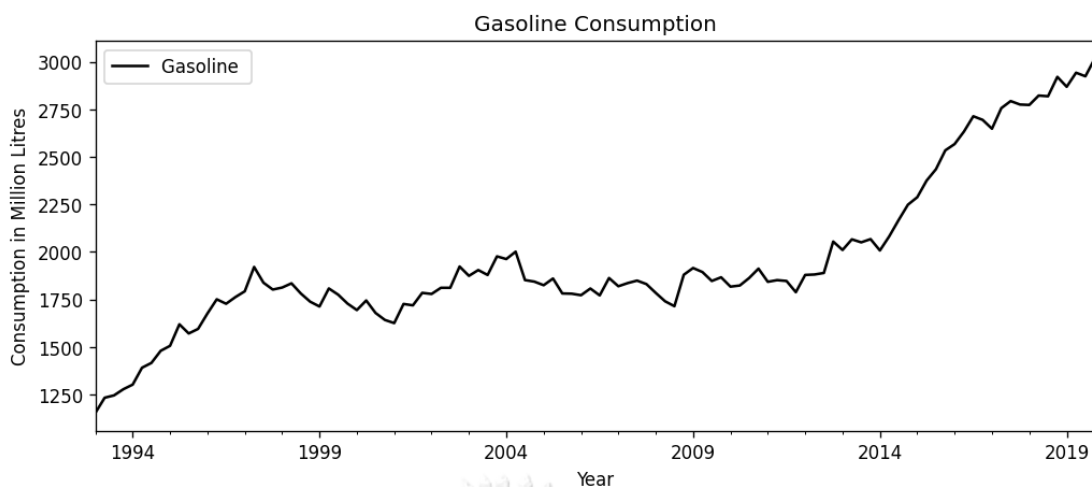
สำหรับขั้นตอนการตรวจสอบและสำรวจข้อมูลเบื้องต้นจะใช้โปรแกรม Python บนแพลตฟอร์ม Google Colab โดยมี Python Library ที่เกี่ยวข้องคือ dateutil.parser matplotlib seaborn numpy pandas และ statsmodel

#### 3.2.2.1 การสร้างกราฟอนุกรมเวลา

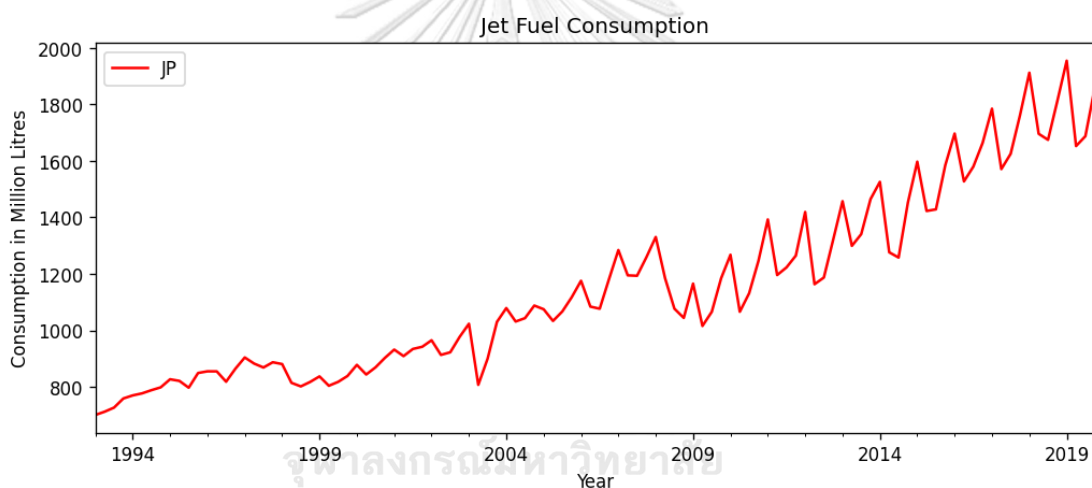
เมื่อนำข้อมูลปริมาณการบริโภค น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินรายไตรมาส ในช่วงเวลา พ.ศ. 2536 ถึง พ.ศ. 2562 (ค.ศ. 1993 ถึง ค.ศ. 2019) มาสร้างกราฟอนุกรมเวลาแสดงดังรูปที่ 12 รูปที่ 13 และรูปที่ 14 ตามลำดับ



รูปที่ 12 กราฟอนุกรมเวลาของปริมาณการบริโภค น้ำมันดีเซล



รูปที่ 13 กราฟอนุกรมเวลาของปริมาณการบริโภคน้ำมันเบนซิน



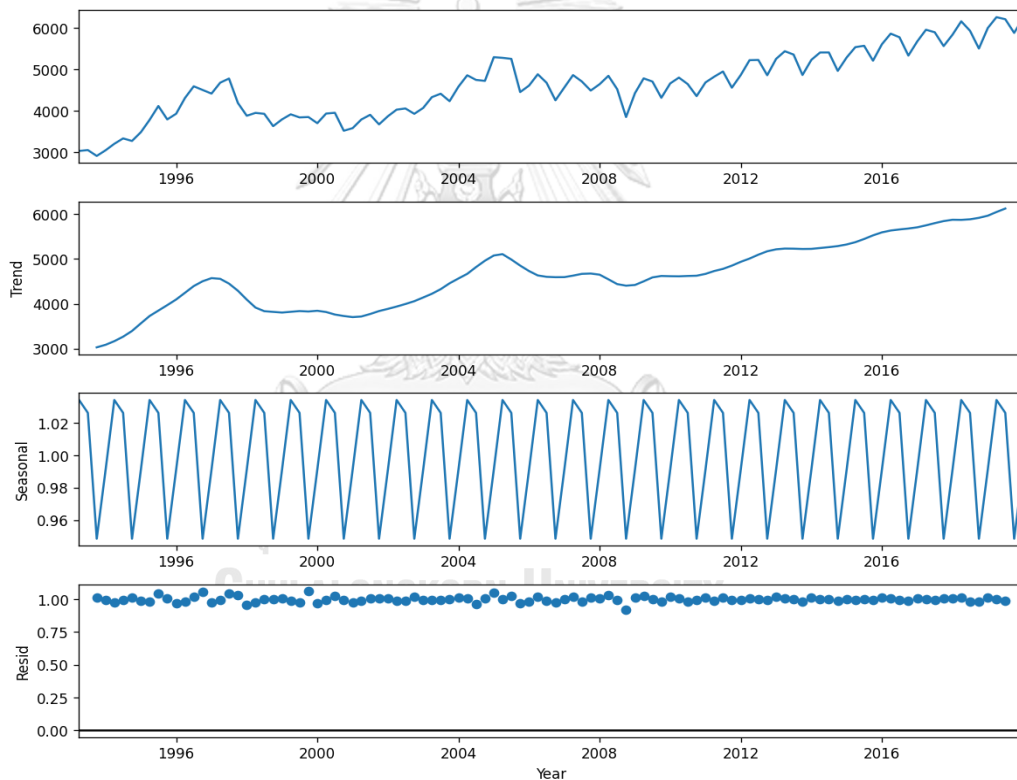
รูปที่ 14 กราฟอนุกรมเวลาของปริมาณการบริโภคน้ำมันเครื่องบิน

จากการพิจารณากราฟอนุกรมเวลาปริมาณของการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน พบว่ากราฟของน้ำมันทั้ง 3 ชนิดมีรูปแบบแนวโน้มสูงขึ้นเมื่อเวลาเปลี่ยนแปลงไปและอาจมีรูปแบบของฤดูกาล เพื่อให้มั่นใจว่าอนุกรมเวลาปริมาณของการบริโภคน้ำมันดีเซล น้ำมันเบนซินและน้ำมันเครื่องบินมีรูปแบบแนวโน้มและฤดูกาลจึงจะใช้วิธีการแยกส่วนประกอบของอนุกรมเวลาในการตรวจสอบรูปแบบของข้อมูลอนุกรมเวลา

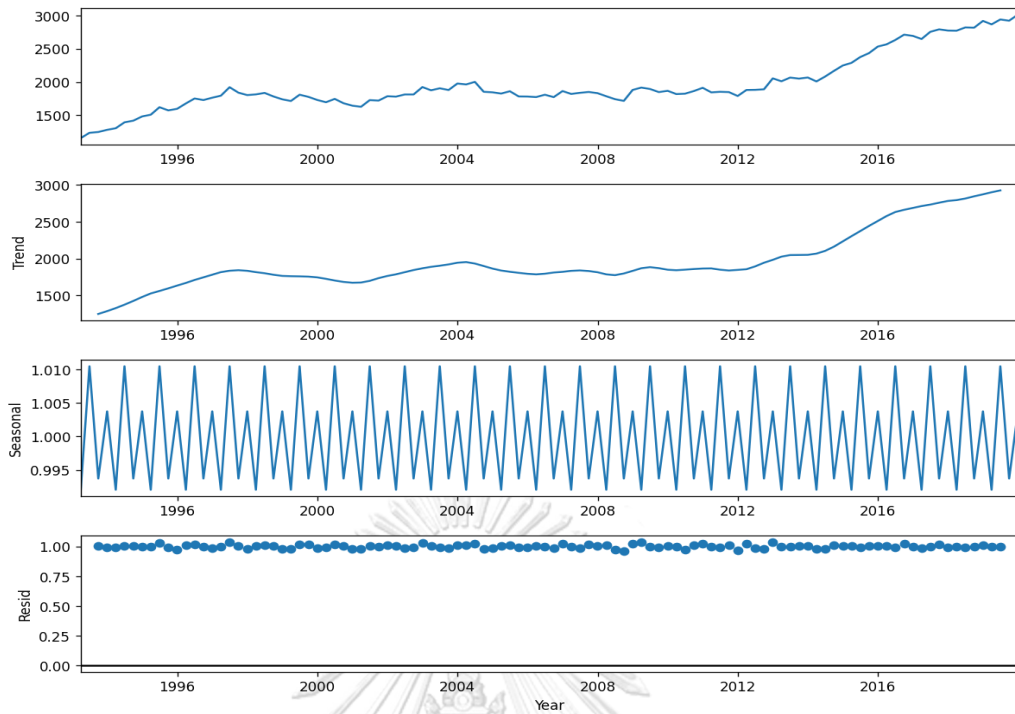
### 3.2.2.2 การตรวจสอบรูปแบบอนุกรมเวลาด้วยการแยกส่วนประกอบของอนุกรมเวลา

ส่วนประกอบของอนุกรมเวลาประกอบด้วย แนวโน้ม (Trend) ฤดูกาล (Seasonal) วัฏจักร (Cyclical) และความผิดปกติ (Irregular) ซึ่งงานวิจัยนี้จะใช้การแยกส่วนประกอบของ

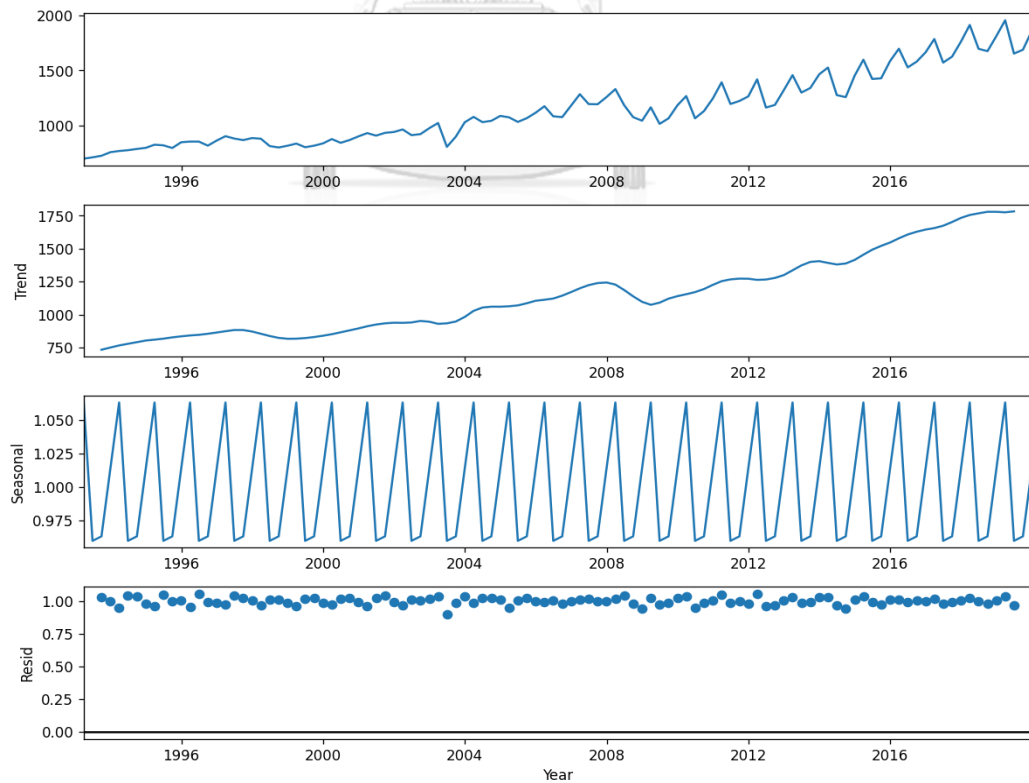
อนุกรมเวลาแบบการคูณ (Multiplicative) สำหรับข้อมูลอนุกรมเวลาของน้ำมันดีเซลและน้ำมันเครื่องบิน เนื่องจากเมื่อพิจารณาจากรูปที่ 12 และ 14 พบว่าข้อมูลมีลักษณะการเคลื่อนไหวของฤดูกาลแบบที่การผันแปร (Variation) เพิ่มขึ้นในขณะที่เวลาเปลี่ยนแปลงไป แต่ข้อมูลอนุกรมเวลาของน้ำมันเบนซินจะใช้การแยกส่วนประกอบของอนุกรมเวลาแบบการบวก (Additive) เนื่องจากเมื่อพิจารณาจากรูปที่ 13 พบว่าข้อมูลมีลักษณะการเคลื่อนไหวของฤดูกาลแบบที่การผันแปรไม่เปลี่ยนแปลงในขณะที่เวลาเปลี่ยนแปลงไป (Hyndman และ Athanasopoulos, 2018) งานวิจัยใช้ Python Library: statsmodels.tsa.seasonal ในการแยกส่วนประกอบของอนุกรมเวลา ซึ่งสามารถเห็นรูปแบบอนุกรมเวลาได้ชัดเจนและได้ผลลัพธ์ดังแสดงในรูปที่ 15 16 และ 17



รูปที่ 15 กราฟส่วนประกอบของอนุกรมเวลาของน้ำมันดีเซล



รูปที่ 16 กราฟส่วนประกอบของอนุกรมเวลาของน้ำมันเบนซิน



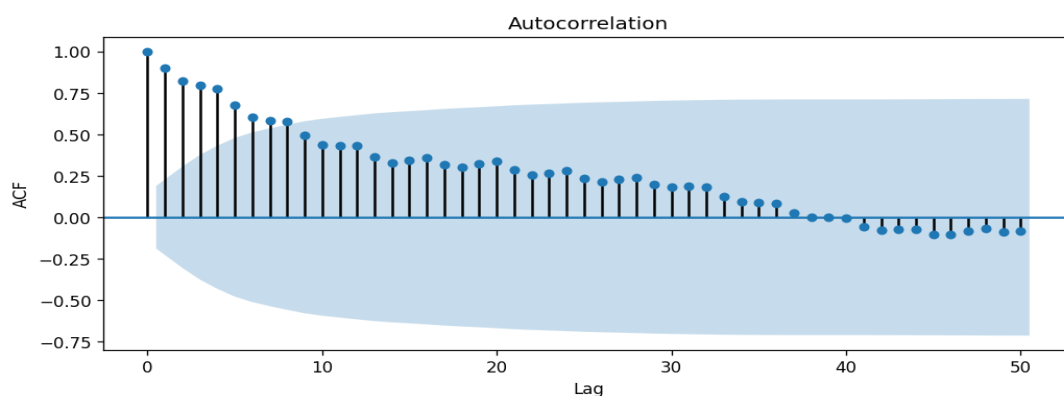
รูปที่ 17 กราฟส่วนประกอบของอนุกรมเวลาของน้ำมันเครื่องบิน

รูปที่ 15 16 และ 17 แสดงให้เห็นส่วนประกอบของอนุกรมเวลาดังต่อไปนี้ น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน มีส่วนประกอบที่เป็นแนวโน้ม ดูได้จากกราฟที่แสดง Trend โดยน้ำมันทั้ง 3 ประเภทมีแนวโน้มสูงขึ้นเมื่อเวลาเปลี่ยนแปลงไป นอกจากนี้ น้ำมันทั้ง 3 ประเภทยังมีส่วนประกอบที่เป็นฤดูกาลอีกด้วย ดูได้จากกราฟ Seasonal

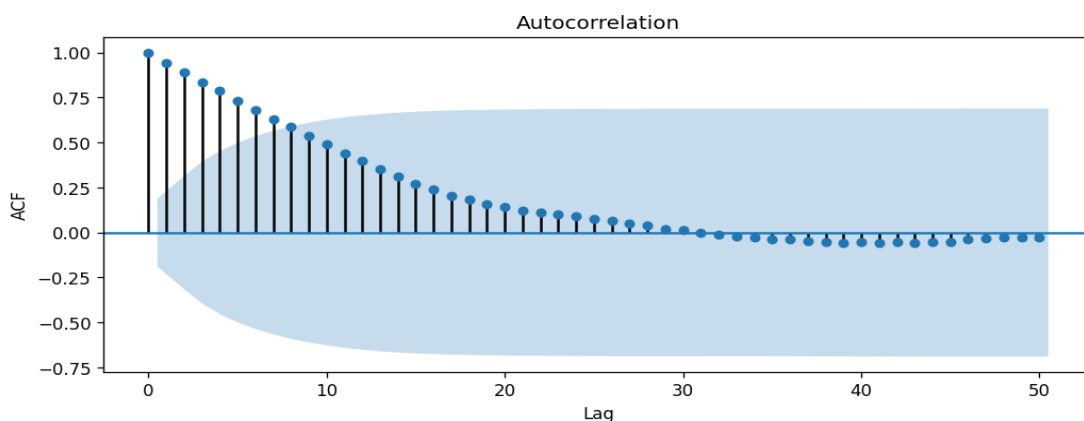
### 3.2.2.3 การตรวจสอบรูปแบบอนุกรมเวลาด้วยฟังก์ชันอัตสหสัมพันธ์ (ACF)

ฟังก์ชันอัตสหสัมพันธ์ (Autocorrelation Function, ACF) เป็นหนึ่งในวิธีที่ใช้ในการตรวจสอบรูปแบบของอนุกรมเวลา โดยในงานวิจัยนี้จะใช้ Python Library: statsmodels ในการสร้างกราฟ ACF เพื่อใช้ในการหารูปแบบของอนุกรมเวลาที่ปรากฏในข้อมูลปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซินและน้ำมันเครื่องบิน โดยใช้ Python Library: statsmodels.graphics.tsaplots ในการหา ACF ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน

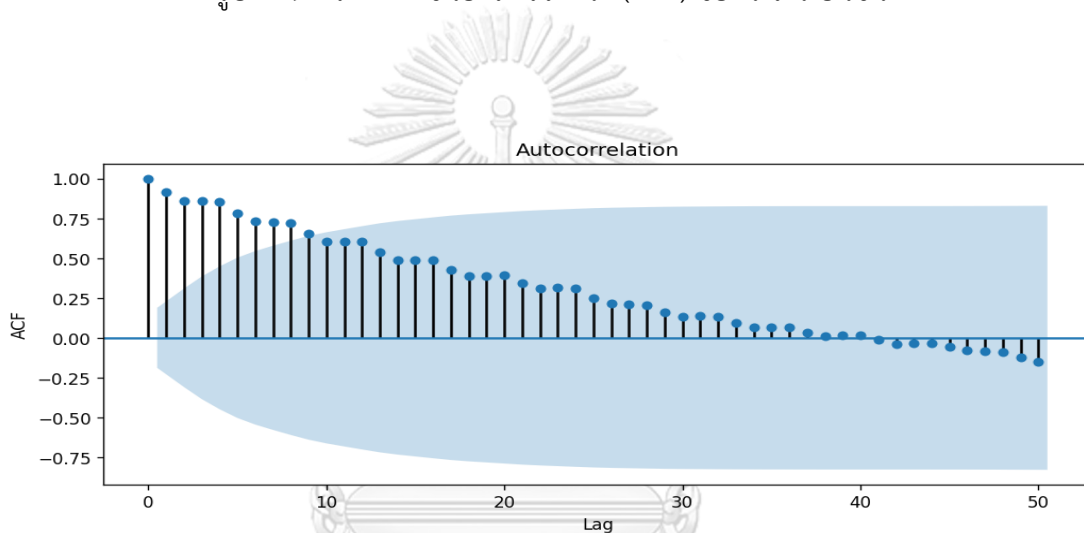
ในการตรวจสอบรูปแบบอนุกรมเวลาของข้อมูลปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซินและน้ำมันเครื่องบิน โดยใช้ ACF ดังแสดงในรูปที่ 18 19 และ 20 ตามลำดับ พบว่าการลดลงของค่า ACF อย่างช้า ๆ ในขณะที่ Lag เพิ่มขึ้นนั้นแสดงให้ว่าข้อมูลมีรูปแบบแนวโน้มปรากฏอยู่ และการที่ ACF มีลักษณะการเคลื่อนไหวเป็นรูปเส้นโค้งเปลือกหอย (Scalloped Shape) ปรากฏอยู่ แสดงว่าข้อมูลมีรูปแบบฤดูกาลปรากฏอยู่ด้วย (Hyndman และ Athanasopoulos, 2018) แต่ในรูปที่ 19 กราฟ ACF มีลักษณะการเคลื่อนไหวเป็นรูปเส้นโค้งเปลือกหอยไม่ชัดเจนนักเมื่อเทียบกับรูป 18 และ 20 แต่เมื่อพิจารณาร่วมกับวิธีการแยกส่วนประกอบอนุกรมเวลาแล้วจึงสรุปได้ว่าข้อมูลปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซินและน้ำมันเครื่องบิน มีรูปแบบแนวโน้มและฤดูกาลปรากฏอยู่



รูปที่ 18 กราฟฟังก์ชันอัตสหสัมพันธ์ (ACF) ของน้ำมันดีเซล



รูปที่ 19 กราฟฟังก์ชันอัตสหสัมพันธ์ (ACF) ของน้ำมันเบนซิน



รูปที่ 20 กราฟฟังก์ชันอัตสหสัมพันธ์ (APF) ของน้ำมันเครื่องบิน

### 3.2.3 การสร้างตัวแบบอนุกรมเวลาเชิงสถิติ

ในขั้นตอนนี้จะเลือกตัวแบบที่เหมาะสมสำหรับใช้พยากรณ์อนุกรมเวลาที่มีองค์ประกอบของแนวโน้มและฤดูกาล โดยในงานวิจัยนี้จะใช้เลือกใช้ตัวแบบการปรับเรียบเอกซ์โพเนนเชียลแบบโฮลด์และวินเทอร์ (Holt-Winters Exponential Smoothing) จากนั้นไปจะเรียกว่าตัวแบบ Holt-Winters ตัวแบบบ็อกซ์-เจนกินส์ (Box-Jenkins: SARIMA) จากนั้นไปจะเรียกว่าตัวแบบ SARIMA ตัวแบบ Seasonal Autoregressive Integrated Moving Average with Exogenous Variables จากนั้นไปจะเรียกว่าตัวแบบ SARIMAX ในการสร้างตัวแบบการพยากรณ์อนุกรมเวลาของข้อมูลปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน โดยผู้วิจัยได้เลือกใช้โปรแกรม Python บนแพลตฟอร์ม Google Colab เป็นเครื่องมือในการสร้างตัวแบบทั้งหมดในงานวิจัยนี้

### 3.2.3.1 การสร้างตัวแบบ Holt-Winters

จากผลการตรวจสอบและสำรวจข้อมูลเบื้องต้น ผู้วิจัยจึงเลือกใช้ตัวแบบ Holt-Winters แบบการบวก ดังสมการที่ 1 ถึง 4 สำหรับข้อมูลปริมาณการบริโภคของน้ำมันเบนซิน และเลือกใช้ตัวแบบ Holt-Winters แบบการคูณ ดังสมการที่ 5 ถึง 8 สำหรับข้อมูลปริมาณการบริโภคของน้ำมันดีเซลและน้ำมันเครื่อง ยานยนต์ โดยผู้วิจัยใช้ Python Library: statsmodels.tsa.holtwinters ในการสร้างตัวแบบ Holt-Winters เพื่อให้ผลการพยากรณ์จากตัวแบบมีความแม่นยำมากที่สุดจึงต้องมีการปรับค่าพารามิเตอร์  $\alpha$   $\beta$  และ  $\gamma$  ให้เหมาะสมด้วย Python Library ดังกล่าว ซึ่งจะใช้ชุดข้อมูลฝึกสอนตามที่กำหนดสำหรับการพยากรณ์ระยะสั้นและระยะยาว เมื่อได้ตัวแบบ Holt-Winters ที่มีค่าพารามิเตอร์ที่เหมาะสมแล้วต่อไปให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 สำหรับการพยากรณ์ระยะกลาง และทำการพยากรณ์ไปล่วงหน้าครั้งละ 4 ไตรมาสหรือครั้งละ 1 ปี จนครบ 20 ไตรมาส สำหรับการพยากรณ์ระยะสั้น เพื่อนำผลการพยากรณ์ไปเปรียบเทียบกัน

### 3.2.3.2 การสร้างตัวแบบ SARIMA

ผู้วิจัยเลือกใช้ Python Library: pmdarima.arima และใช้ชุดคำสั่งสำหรับสร้างตัวแบบ SARIAMA เป็น pmdarima.arima.auto\_arima ซึ่งสามารถหาตัวแบบ SARIMA ได้ดังในสมการที่ 2.9 และเขียนตัวแบบอย่างง่ายเป็น SARIMA(p,d,q)(P,D,Q)s โดยใช้เทคนิคในการหาค่าเหมาะสมของพารามิเตอร์ p d q P D และ Q กับการจัดกลุ่ม (Combination) ที่แตกต่างกัน และค่าเหมาะสมของพารามิเตอร์สุดท้ายของแต่ละตัวแบบที่จะใช้พยากรณ์ปริมาณการบริโภคน้ำมันดีเซล เบนซิน และน้ำมันเครื่อง ยานยนต์จะถูกเลือกจากค่าพารามิเตอร์ที่ทำให้ Akaike Information Criterion (AIC) มีค่าต่ำที่สุดและทำการทดสอบ z- test กับสัมประสิทธิ์ (Coefficients) เพื่อทดสอบสัมประสิทธิ์ทุกตัวของตัวแบบว่ามีนัยสำคัญภายใต้ระดับความเชื่อมั่น 95% จากนั้นนำข้อมูลความผิดพลาด (Residual) ที่ได้จากการหาค่าความแตกต่างระหว่างช่วงค่าชุดข้อมูลฝึกสอนกับช่วงค่าที่คำนวณได้จากตัวแบบ SARIMA ไปตรวจสอบรูปแบบ (Diagnostic Checking) ว่ารูปแบบที่สร้างขึ้นเหมาะสมกับอนุกรมเวลาหรือไม่ โดยต้องสอดคล้องกับเงื่อนไขดังต่อไปนี้ 1) กราฟ Standardized Residual มีลักษณะการกระจายตัวอยู่รอบ ๆ ค่า 0 2) ตรวจสอบกราฟ ACF ของค่าความผิดพลาดจากตัวแบบต้องไม่เหลือรูปแบบอนุกรมเวลาอื่นใด เช่น แนวโน้ม ฤดูกาล นอกเหนือจากรูปแบบสุ่ม (Random) 3) ตรวจสอบลักษณะการกระจายตัวของข้อมูลค่าความผิดพลาดต้องมีการแจกแจงแบบปกติ (Normal Distribution) เมื่อได้ตัวแบบ SARIMA ที่มีค่าพารามิเตอร์ที่



เหมาะสมกับตัวแบบการพยากรณ์ระยะสั้นและระยะกลางแล้ว ต่อไปให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึงพ.ศ. 2557 สำหรับการพยากรณ์ระยะกลาง และทำการพยากรณ์ไปล่วงหน้าครั้งละ 4 ไตรมาสหรือครั้งละ 1 ปี จนครบ 20 ไตรมาส สำหรับการพยากรณ์ระยะสั้น เพื่อนำผลการพยากรณ์ไปเปรียบเทียบกัน

### 3.2.3.3 การสร้างตัวแบบ SARIMAX

ผู้วิจัยเลือกใช้ Python Library: pmdarima.arima ซึ่งสามารถหาตัวแบบ SARIMAX ได้ตั้งในสมการที่ 2.16 และเขียนตัวแบบอย่างง่ายเป็น SARIMAX(p,d,q)(P,D,Q)s โดยจะใช้วิธีการสร้างตัวแบบ SARIMAX เหมือนกันกับตัวแบบ SARIMA แต่มีการใช้ตัวแปรอิสระที่มีผลต่อปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินที่ถูกคัดเลือกจากเทคนิคการถดถอยแบบขั้นตอนมาใช้เป็นตัวแปรอิสระภายนอกเพิ่มเข้าไปในตัวแบบ โดยจะเพิ่มอาร์กิวเมนต์ exogeneous ในชุดคำสั่ง pmdarima.arima.auto\_arima เมื่อได้ตัวแบบ SARIMAX ที่มีค่าพารามิเตอร์ที่เหมาะสมกับตัวแบบการพยากรณ์ระยะสั้นและระยะกลางแล้ว ต่อไปให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 สำหรับการพยากรณ์ระยะกลาง และทำการพยากรณ์ไปล่วงหน้าครั้งละ 4 ไตรมาสหรือครั้งละ 1 ปี จนครบ 20 ไตรมาส สำหรับการพยากรณ์ระยะสั้น เพื่อนำผลการพยากรณ์ไปเปรียบเทียบกัน

## 3.3 การพยากรณ์เชิงสาเหตุ

จากการทบทวนงานวิจัยที่เกี่ยวข้องในบทที่ 2 สำหรับการพยากรณ์เชิงสาเหตุในงานวิจัยนี้จะใช้ตัวแปรอิสระที่สามารถอธิบายความสัมพันธ์ที่อาจเกี่ยวข้องกับตัวแปรตามโดยจะกำหนดตัวแปรดังนี้

- 1) ตัวแปรตาม
  - ปริมาณการบริโภคน้ำมันดีเซลในหน่วยล้านลิตร; Diesel Consumption
  - ปริมาณการบริโภคน้ำมันเบนซินในหน่วยล้านลิตร; Gasoline Consumption
  - ปริมาณการบริโภคน้ำมันเครื่องบินในหน่วยล้านลิตร; JP Consumption
- 2) ตัวแปรอิสระที่อาจมีความสัมพันธ์กับตัวแปรตามมีดังนี้
  - ปริมาณการผลิตน้ำมันดีเซลในหน่วยล้านลิตร; Diesel Production
  - ปริมาณการผลิตน้ำมันเบนซินในหน่วยล้านลิตร; Gasoline Production
  - ปริมาณการผลิตน้ำมันเครื่องบินในหน่วยล้านลิตร; JP Production

- ปริมาณการผลิตก๊าซปิโตรเลียมเหลวในหน่วยล้านลิตร; LPG Production
- ราคาน้ำมันดีเซลในหน่วยบาทต่อลิตร; Diesel Price
- ราคาน้ำมันเบนซินในหน่วยบาทต่อลิตร; Gasoline Price
- ราคาก๊าซปิโตรเลียมเหลวในหน่วยบาทต่อกิโลกรัม; LPG Price
- ผลิตภัณฑ์มวลรวม; ExpeGDP
- จำนวนรถใหม่ที่จดทะเบียน; Total Vehicle
- จำนวนประชากร; Population

ตัวแปรอิสระที่ผู้วิจัยได้คิดว่าอาจส่งผลกระทบต่อประมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักประเทศไทยนั้นผู้วิจัยได้ทบทวนวรรณกรรมที่เกี่ยวข้องกับการพยากรณ์น้ำมันสำเร็จรูป พบว่าผลิตภัณฑ์มวลรวมและจำนวนประชากรของแต่ละประเทศนั้นมีความสัมพันธ์ที่ส่งผลกระทบต่อปริมาณการบริโภคน้ำมันสำเร็จรูปชนิดต่าง ๆ ในแต่ละประเทศที่ถูกศึกษา ซึ่งเป็นตัวแปรอิสระที่สำคัญ ผู้วิจัยจึงคาดว่าจะส่งผลกระทบต่อประมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักประเทศไทยด้วย ส่วนตัวแปรอิสระกลุ่มราคาของน้ำมันสำเร็จรูปแต่ละชนิดที่ต่างกัน ผู้วิจัยคิดว่าอาจมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันสำเร็จรูปที่ชนิดต่างกันด้วย เช่น ราคาน้ำมันเบนซินมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันดีเซล เป็นต้น และตัวแปรกลุ่มปริมาณการผลิตน้ำมันสำเร็จรูปแต่ละชนิดที่มีปริมาณการผลิตค่อนข้างสูง ผู้วิจัยคิดว่าอาจมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันสำเร็จรูปที่ชนิดต่างกันด้วยเช่นกัน ดังนั้นในการคัดเลือกตัวแปรอิสระที่ส่งผลกระทบต่อปริมาณการบริโภคน้ำมันสำเร็จรูปแต่ละชนิด ตัวแปรอิสระทั้งหมดดังในรายการที่ถูกลำถึงข้างต้นนี้จะถูกนำมาพิจารณาในการคัดเลือกตัวแปรอิสระที่ส่งผลกระทบต่อปริมาณการบริโภคน้ำมันสำเร็จรูปทั้ง 3 ชนิด

### 3.3.1 การพยากรณ์ด้วยตัวแบบการถดถอยเชิงเส้นพหุคูณ (MLR)

สมมติฐาน (Hypothesis) ที่ใช้ในงานวิจัยนี้คือ

$$H_0 : \beta_i = 0 \forall i$$

$H_a$  : มี  $\beta_i$  อย่างน้อย 1 ตัวที่มีค่าไม่เท่ากับ 0

ซึ่งกล่าวได้ว่า

$H_0$  : ไม่มีตัวแปรอิสระใดที่มีความสัมพันธ์กับตัวแปรตามอย่างมีนัยสำคัญ

$H_a$  : มีตัวแปรอิสระอย่างน้อย 1 ตัวที่มีความสัมพันธ์กับตัวแปรตามอย่างมีนัยสำคัญ

งานวิจัยนี้จะใช้ตัวแบบการถดถอยพหุคูณดังสมการที่ 2.18 และจะใช้เทคนิคการถดถอยแบบขั้นตอน (Stepwise Regression) ในการคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามที่ Stepwise Selection Term ( $\alpha$ ) = 0.05 โดยใช้โปรแกรม Minitab เวอร์ชัน 18 เมื่อได้ตัวแบบสำหรับการพยากรณ์ระยะสั้นและระยะยาวของน้ำมันสำเร็จรูปทั้ง 3 ชนิดแล้ว จึงทำการทดสอบ

สมมติฐานของการถดถอยเชิงเส้นกับตัวแบบที่สร้างขึ้นโดยมีข้อสมมติฐานพื้นฐาน 4 ข้อคือ ที่ค่าใด ๆ ของตัวแปรอิสระ 1) ประชากรของค่าความผิดพลาดมีค่าเฉลี่ยเท่ากับศูนย์ 2) ประชากรของค่าความผิดพลาดมีความแปรปรวนคงที่ 3) ประชากรของค่าความผิดพลาดมีการแจกแจงแบบปกติ และ 4) ค่าความผิดพลาดมีความเป็นอิสระต่อกัน (นันทชัย, 2555) จากนั้นให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 สำหรับการพยากรณ์ระยะกลาง และทำการพยากรณ์ไปล่วงหน้าครั้งละ 4 ไตรมาสหรือครั้งละ 1 ปี จนครบ 20 ไตรมาส สำหรับการพยากรณ์ระยะสั้น เพื่อนำผลการพยากรณ์ไปเปรียบเทียบกัน

### 3.3.2 การพยากรณ์ด้วยตัวแบบ Random Sample Consensus Regression (RANSAC-REG)

ตัวแบบ Random Sample Consensus Regression ต่อไปจะเรียกว่าตัวแบบ RANSAC-REG ผู้วิจัยเลือกสร้างตัวแบบโดยใช้ Python Library คือ `sklearn.linear_model` และใช้คำสั่ง `RANSACRegressor` โดยมีการกำหนดให้ทดลองใช้ค่า `random_state` คือตัวเลขที่ใช้กำหนดการทำงานของตัวสุ่มตัวเลขของจุดศูนย์กลางเริ่มต้น โดยใช้ `random_state` เป็น 1, 123, 6789, 12345 และ 7891011 และเลือกพารามิเตอร์ `base_estimator` เป็นค่าเริ่มต้น (default=None) นั่นคือไม่ใส่ค่าใด ๆ ในพารามิเตอร์ `base_estimator` ซึ่งค่าเริ่มต้นของ `base_estimator` คือการเรียกใช้ตัวแบบ Linear Regression สำหรับค่าเป้าหมาย โดย `random_state` และ `base_estimator` เป็นการปรับค่าพารามิเตอร์ก่อนสร้างตัวแบบ เมื่อได้ตัวแบบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองน้อยสุดจากการทดลองปรับค่าพารามิเตอร์ที่กำหนดให้ทดลองใช้แล้ว จากนั้นให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 สำหรับการพยากรณ์ระยะกลาง และทำการพยากรณ์ไปล่วงหน้าครั้งละ 4 ไตรมาสหรือครั้งละ 1 ปี จนครบ 20 ไตรมาส สำหรับการพยากรณ์ระยะสั้น เพื่อนำผลการพยากรณ์ไปเปรียบเทียบกัน

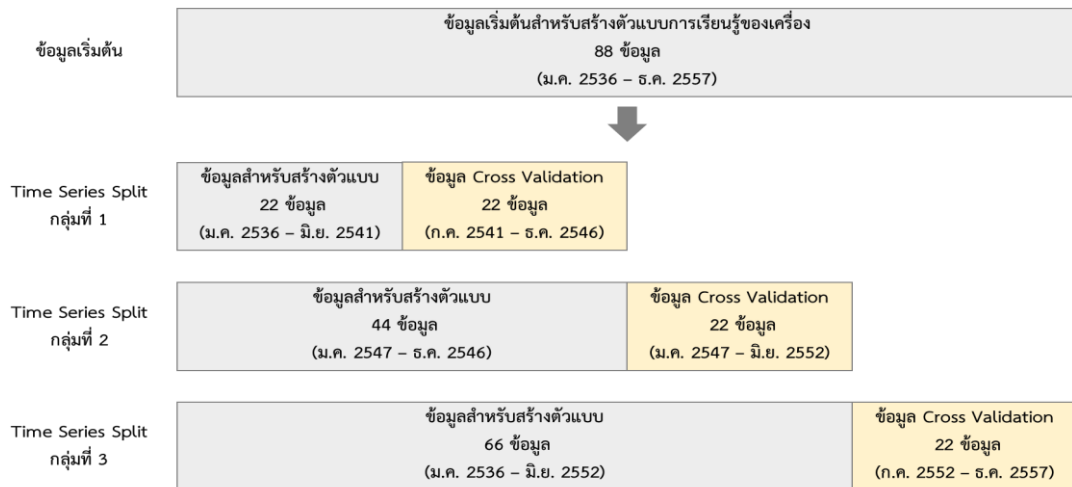
### 3.4 การพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง

ในขั้นตอนนี้จะมีการสร้างตัวแบบการพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่องโดยใช้ตัวแบบ ดังนี้ ตัวแบบ K-nearest Neighbors ตัวแบบ Support Vector Regression ตัวแบบ Adaboost ตัวแบบ Artificial Neural Network และตัวแบบ Hybrid

#### 3.4.1 การปรับไฮเปอร์พารามิเตอร์ของตัวแบบการเรียนรู้ของเครื่อง

ไฮเปอร์พารามิเตอร์ (Hyperparameter) เป็นพารามิเตอร์ของอัลกอริทึมการเรียนรู้ (ไม่ใช่ของตัวแบบ) ดังนั้นจึงไม่ได้รับผลกระทบจากอัลกอริทึมการเรียนรู้ ซึ่งไม่ใช่พารามิเตอร์ที่เกิดจากการ

เรียนรู้ของอัลกอริทึมการเรียนรู้ของเครื่อง โดยจะถูกระบุค่าพารามิเตอร์ไว้ก่อนที่จะสร้างตัวแบบและยังคงมีค่าคงที่ในระหว่างการสร้างตัวแบบการเรียนรู้ของเครื่อง ซึ่งการปรับไฮเปอร์พารามิเตอร์เป็นสิ่งที่สำคัญส่วนหนึ่งของการสร้างระบบการเรียนรู้ของเครื่อง ผู้วิจัยเลือกใช้เทคนิคการปรับการปรับไฮเปอร์พารามิเตอร์ของตัวแบบการเรียนรู้ของเครื่องเป็น GridSearchCV จาก Python Library คือ sklearn.model\_selection เนื่องจากเป็นเทคนิคที่ใช้งานได้ไม่ยุ่งยากและเป็นที่ยอมรับ ในการทำ GridSearchCV นั้นจะต้องทำการกำหนดค่าพารามิเตอร์ที่ต้องการจะใช้ทดลองในการสร้างตัวแบบการเรียนรู้ของเครื่อง โดยจะมีการประเมินผลลัพธ์ของตัวแบบทุกการรวมกัน (Combination) ของแต่ละค่าพารามิเตอร์ที่เป็นไปได้ โดยใช้วิธีการแบ่งชุดข้อมูลตรวจสอบออกเป็นหลายชุด (Cross Validation) แต่ละตัวแบบจะถูกประเมินผลลัพธ์หนึ่งครั้งต่อหนึ่งชุดข้อมูลสำหรับตรวจสอบโดยทำซ้ำจนครบชุดข้อมูลที่ถูกแบ่งด้วยวิธีวัดประสิทธิภาพของตัวแบบที่เลือกใช้ จากนั้นจึงหาค่าเฉลี่ยของค่าที่ได้จากการประเมินผลลัพธ์ของตัวแบบที่ได้จากแต่ละชุดข้อมูลตรวจสอบที่ถูกแบ่ง ซึ่งจะหาค่าการวัดประสิทธิภาพของตัวแบบที่แม่นยำมากขึ้น แต่อย่างไรก็ตามยิ่งแบ่งชุดข้อมูลตรวจสอบเป็นจำนวนมากขึ้นจะยิ่งใช้เวลาในการสร้างตัวแบบมากขึ้นด้วยเช่นกัน (Géron, 2019) ผู้วิจัยเลือกใช้วิธีการแบ่งชุดข้อมูลสำหรับตรวจสอบออกเป็นหลายชุดโดยใช้คำสั่ง TimeSeriesSplit จาก Python Library คือ sklearn.model\_selection ซึ่งเป็นวิธีที่เหมาะสมกับข้อมูลอนุกรมเวลาโดยชุดข้อมูลตรวจสอบจะอยู่ในช่วงเวลาตามหลังชุดข้อมูลที่ใช้สร้างตัวแบบเสมอ ผู้วิจัยเลือกกำหนดกลุ่มการแบ่งชุดข้อมูลตรวจสอบเป็น 3 กลุ่มดังแสดงในรูปที่ 21 เนื่องจากใช้เวลาสร้างตัวแบบไม่มากนักและยังให้ผลลัพธ์การประเมินตัวแบบที่แม่นยำ สำหรับค่าเริ่มต้นของจำนวนข้อมูลตรวจสอบจะคำนวณได้จากการหารค่าจำนวนข้อมูลเริ่มต้นสำหรับสร้างตัวแบบด้วยผลรวมของค่าจำนวนกลุ่มการแบ่งชุดข้อมูลตรวจสอบกับหนึ่ง แล้วปัดเศษขึ้นให้เป็นจำนวนเต็ม ดังนั้นเมื่อค่าจำนวนข้อมูลเริ่มต้นสำหรับสร้างตัวแบบและสำหรับการทำ Cross Validation คือ 88 ข้อมูล ค่าจำนวนกลุ่มการแบ่งชุดข้อมูลตรวจสอบคือ 3 กลุ่ม จะได้จำนวนข้อมูลตรวจสอบเป็น 22 ข้อมูล สำหรับการสร้างตัวแบบการเรียนรู้ของเครื่องของการพยากรณ์น้ำมันสำเร็จรูปกลุ่มหลักแบบระยะกลางจะใช้ข้อมูลอนุกรมเวลารายไตรมาสของข้อมูลปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินในช่วงเวลา พ.ศ. 2536 ถึง พ.ศ. 2552 จำนวนชุดข้อมูลละ 68 ข้อมูลจะถูกใช้เป็นส่วนข้อมูลสำหรับฝึกสอน (Training Data Set) และข้อมูลในช่วงเวลา พ.ศ. 2553 ถึง พ.ศ. 2557 จำนวนชุดข้อมูลละ 20 ข้อมูลจะถูกใช้เป็นส่วนข้อมูลสำหรับตรวจสอบ (Validation Data Set) ผู้วิจัยจะนำชุดข้อมูลที่ใช้สำหรับฝึกสอนและชุดข้อมูลสำหรับตรวจสอบรวมเข้าด้วยกันเพื่อนำไปใช้ในการปรับไฮเปอร์พารามิเตอร์ของตัวแบบการเรียนรู้ของเครื่องด้วยวิธี GridSearchCV โดยในขั้นตอนการทำ Cross Validation ด้วย TimeSeriesSplit นั้นชุดข้อมูลสำหรับตรวจสอบจะไม่ถูกนำมาใช้ฝึกสอนตัวแบบ แต่จะเป็นส่วนหนึ่งที่จะใช้เป็นส่วนข้อมูล Cross Validation ดังแสดงในรูปที่ 21



รูปที่ 21 การแบ่งชุดข้อมูลแบบ Time Series Split สำหรับชุดข้อมูลปริมาณน้ำมันสำเร็จรูป

ผู้วิจัยเลือกใช้ตัวแปรอิสระที่ถูกคัดเลือกจากเทคนิคการถดถอยแบบขั้นตอน และเลือกใช้ตัวแปรอิสระทั้งหมดที่ใช้ในงานวิจัยนี้ในการสร้างตัวแบบ K-nearest Neighbors ตัวแบบ Support Vector Regression ตัวแบบ Artificial Neural Network และตัวแบบ Adaboost เพื่อเปรียบเทียบผลลัพธ์ของตัวแบบที่ได้จากการใช้ตัวแปรอิสระทั้งสองแบบ และก่อนทำการสร้างตัวแบบ จะต้องทำการปรับสเกลของคุณลักษณะหรือตัวแปรที่จะใช้ในการสร้างตัวแบบแบบ K-nearest Neighbors ตัวแบบ Support Vector Regression ตัวแบบ Artificial Neural Network ซึ่งเรียกว่า Feature Scaling โดยผู้วิจัยเลือกใช้ Python Library คือ sklearn.preprocessing และใช้คำสั่ง MinMaxScaler โดยกำหนดค่าสูงสุดในช่วงเป็น 0.99 และค่าต่ำสุดเป็น 0.01 ส่วนตัวแบบในกลุ่ม Ensemble เช่น ตัวแบบ Random Forest ตัวแบบ Adaboost ไม่จำเป็นต้องทำ Feature Scaling และสำหรับการประเมินผลลัพธ์ของตัวแบบที่ได้จากการปรับไฮเปอร์พารามิเตอร์จะใช้ฟังก์ชันวัตถุประสงค์ (Objective Function) เป็นแบบฟังก์ชันที่เราต้องการหาค่าต่ำสุดหรือเรียกว่า Loss Function โดยผู้วิจัยเลือกใช้เป็นฟังก์ชันค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Squared Error, MSE) ซึ่งเป็น Loss Function ที่เหมาะสมกับการพยากรณ์ที่มีลักษณะปัญหาเป็นการถดถอย มีสมการดังนี้

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3.1)$$

เมื่อ  $Y_i$  คือ ค่าของข้อมูลจริง

$\hat{Y}_i$  คือ ค่าพยากรณ์ของข้อมูล

$n$  คือ จำนวนข้อมูล

สำหรับตัวแบบการเรียนรู้ของเครื่องทั้งหมด ผู้วิจัยจะทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาส หรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึงพ.ศ. 2557 สำหรับการพยากรณ์ในระยะกลาง ส่วนการพยากรณ์ระยะสั้นจะพิจารณาความแม่นยำของตัวแบบที่ได้จากการพยากรณ์ระยะกลางก่อน ซึ่งหากตัวแบบระยะกลางมีความแม่นยำในระดับที่ผู้วิจัยยอมรับได้แล้วคือ MAPE ของชุดข้อมูลตรวจสอบมีค่าไม่เกิน 5% ผู้วิจัยจะไม่ทำการพยากรณ์ระยะสั้นสำหรับตัวแบบการเรียนรู้ของเครื่องที่จะทำการศึกษาทั้งหมด แต่จะเลือกใช้ตัวแบบการพยากรณ์ระยะกลาง ซึ่งมีช่วงการพยากรณ์ที่ครอบคลุมการพยากรณ์ระยะสั้นอยู่แล้ว

#### 3.4.2 การสร้างตัวแบบ K-nearest Neighbors

ตัวแบบ K-nearest Neighbors ซึ่งต่อไปจะเรียกว่าตัวแบบ KNN ผู้วิจัยเลือกสร้างตัวแบบโดยใช้ Python Library คือ `sklearn.neighbors` และใช้คำสั่ง `KNeighborsRegressor` และ จากนั้นจะทำการกำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ KNN ที่จะใช้ทดลองสร้างตัวแบบดังแสดงในตารางที่ 4 ซึ่งประกอบด้วยไฮเปอร์พารามิเตอร์ดังนี้

- 1) `n_neighbors` คือ จำนวนเพื่อนบ้านใกล้สุด  $k$  ตัว ในงานวิจัยนี้จะทดลองใช้ค่าตั้ง 1 ถึง 22 เนื่องจากมีเงื่อนไขในการเลือกใช้  $k$  ในคำสั่ง `KNeighborsRegressor` นั่นคือ  $k \leq$  จำนวนข้อมูลที่ใช้สร้างตัวแบบ ซึ่งจากการทำ Cross Validation แบบ Time Series Split นั้น จำนวนข้อมูลสำหรับใช้สร้างตัวแบบที่น้อยที่สุดคือ 22 ข้อมูล
- 2) `metric` คือ มาตรฐานระยะทาง งานวิจัยนี้จะทดลองใช้เป็น `'euclidean'` `'manhattan'` และ `'minkowski'`

ตารางที่ 4 ค่าไฮเปอร์พารามิเตอร์ที่ทดลองใช้ในการสร้างตัวแบบ KNN

ไฮเปอร์พารามิเตอร์	ค่ากำหนดที่ใช้ใน GridSearchCV
<code>n_neighbors</code> หรือค่า $k$	1, 2, 3, 4, ..., 22
<code>metric</code>	<code>'euclidean'</code> , <code>'manhattan'</code> , <code>'minkowski'</code>

ผู้วิจัยใช้ GridSearchCV ในการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ KNN เมื่อได้ตัวแบบที่มีค่าความคลาดเคลื่อนกำลังสองต่ำสุดจากการทดลองปรับค่าไฮเปอร์พารามิเตอร์ที่กำหนดให้ทดลองใช้แล้ว จากนั้นให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึงพ.ศ. 2557

### 3.4.3 การสร้างตัวแบบ Support Vector Regression

ตัวแบบ Support Vector Regression ซึ่งต่อไปจะเรียกว่าตัวแบบ SVR ผู้วิจัยเลือกสร้างตัวแบบโดยใช้ Python Library คือ `sklearn.svm` และใช้คำสั่ง SVR จากนั้นจะทำการกำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ SVR ที่จะใช้ทดลองสร้างตัวแบบดังแสดงในตารางที่ 5 ซึ่งประกอบด้วยไฮเปอร์พารามิเตอร์ดังนี้

ตารางที่ 5 ค่าไฮเปอร์พารามิเตอร์ที่ทดลองใช้ในตัวแบบ Support Vector Regression

Hyperparameter	ค่ากำหนดที่ใช้ใน GridSearchCV
C	0.1, 1, 100, 1000, 3000, 10000
kernel	'rbf', 'poly', 'sigmoid', 'linear'
degree	1, 2, 3, 4, 5
gamma	1, 0.1, 0.01, 0.001, 0.0001

- 1) C คือ เร็กกูลารีไรเซชันพารามิเตอร์ (Regularization Parameter) ระดับของเร็กกูลารีไรเซชันจะแปรผกผันกับค่า C โดยค่า C ต้องมีค่าทางบวกเท่านั้น และมี Penalty Function คือ ฟังก์ชันกำลังสองของ Ridge Regression (l2)
- 2) kernel คือ Kernel Trick ที่ใช้ในอัลกอริทึม SVR โดยที่ 'rbf' คือ Radial Basis Function Kernel, 'poly' คือ Polynomial Kernel, 'sigmoid' คือ Sigmoid Kernel และ 'linear' คือ Linear Kernel ซึ่ง 'rbf', 'poly', 'sigmoid' เป็น Kernel Function สำหรับตัวแบบที่ไม่เชิงเส้น ส่วน 'linear' เป็น Kernel Function สำหรับตัวแบบเชิงเส้น
- 3) degree คือ ดีกรีพหุนามของ Polynomial Kernel
- 4) gamma คือ ค่าสัมประสิทธิ์ของ Radial Basis Function Kernel, Polynomial Kernel และ Sigmoid Kernel

ผู้วิจัยใช้ GridSearchCV ในการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ SVR เมื่อได้ตัวแบบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดจากการทดลองปรับค่าไฮเปอร์พารามิเตอร์ที่กำหนดให้ทดลองใช้แล้ว จากนั้นให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557

### 3.4.4 การสร้างตัวแบบ Artificial Neural Network

ตัวแบบ Artificial Neural Network ซึ่งต่อไปจะเรียกว่าตัวแบบ ANN ผู้วิจัยเลือกสร้างตัวแบบโดยใช้ Python Library คือ `keras.models` `keras.layers` และ `keras.wrappers.scikit_learn` โดย

ใช้คำสั่ง KerasRegressor จากนั้นจะทำการกำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ ANN ที่จะใช้ทดลองสร้างตัวแบบดังแสดงในตารางที่ 6 ซึ่งประกอบด้วยไฮเปอร์พารามิเตอร์ดังนี้

ตารางที่ 6 ค่าไฮเปอร์พารามิเตอร์ที่ทดลองใช้ในตัวแบบ Artificial Neural Network

Hyperparameter	ค่ากำหนดที่ใช้ใน GridSearchCV
batch_size	24, 48, 88
epochs	25, 50, 100, 500, 1000
optimizer	'adam', 'sgd'
units	3, 5, 15, 30, 50
numLayers	1, 2
activation	'relu', 'elu', 'selu'
dropout	0.0, 0.2, 0.4
seed	1, 23, 456

- 1) batch\_size คือ การแบ่งจำนวนข้อมูลที่ใช้สร้างตัวแบบที่ทำการเรียนรู้ตัวแบบด้วย optimizer หนึ่งครั้ง ซึ่ง batch size ขนาดเล็กจะใช้เวลาในการเรียนรู้ตัวแบบได้เร็วกว่าขนาดใหญ่ แต่ batch size ขนาดใหญ่จะให้ค่าประมาณของการเคลื่อนลงตามความชัน (Gradient Descent) ที่แม่นยำกว่า
- 2) epochs คือ จำนวนรอบที่ตัวแบบทำการเรียนรู้ด้วย optimizer จนครบจำนวนข้อมูลที่ใช้สร้างตัวแบบทั้งหมด
- 3) optimizer คือ อัลกอริทึมการปรับให้เหมาะสมสำหรับการเคลื่อนลงตามความชันสำหรับการฝึกตัวแบบการเรียนรู้ของตัวแบบ ANN
- 4) units คือ จำนวนโหนดของชั้นซ่อน (Hidden Layer)
- 5) numLayers คือ จำนวนชั้นซ่อน
- 6) activation คือ ฟังก์ชันกระตุ้น (Activation Function) ที่ใช้งานในชั้นซ่อนและชั้นข้อมูลออก (Output Layer) โดยที่ 'relu' คือฟังก์ชัน Rectified Linear Unit, 'elu' คือฟังก์ชัน Exponential Linear Unit และ 'selu' คือฟังก์ชัน Scaled Exponential Linear Unit
- 7) dropout คือ ความน่าจะเป็นในการลดจำนวนโหนดในแต่ละชั้น (Layer) ของตัวแบบ ANN ออกแบบสุ่ม หรือเรียกว่า Dropout Rate ซึ่งเป็นหนึ่งในเทคนิคการเร็กกิวลาร์ไรเซชันที่นิยมของตัวแบบ ANN



8) seed คือ ตัวเลขที่ใช้กำหนดการทำงานของตัวสุ่มตัวเลข เมื่อกำหนด seed เป็นตัวเลขเดียวกันทุกครั้งที่ในการสร้างตัวแบบจะได้ผลลัพธ์เหมือนเดิม

ผู้วิจัยใช้ GridSearchCV ในการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ ANN เมื่อได้ตัวแบบที่มีค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดจากการทดลองปรับค่าไฮเปอร์พารามิเตอร์ที่กำหนดให้ทดลองใช้แล้ว จากนั้นให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557

#### 3.4.5 การสร้างตัวแบบ Adaboost

ตัวแบบ AdaBoost เป็นตัวแบบในกลุ่ม Ensemble Learning ต่อไปจะเรียกว่าตัวแบบ ADA ซึ่งตัวแบบนี้เกิดจากการรวมผลการพยากรณ์ของตัวแบบ Regression Tree การสร้างตัวแบบ ADA จะใช้ Python Library คือ sklearn.ensemble จะใช้คำสั่ง AdaBoostRegressor จากนั้นจะทำการกำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ RF และตัวแบบ ADA ที่จะใช้ทดลองสร้างตัวแบบดังแสดงในตารางที่ 7 ซึ่งประกอบด้วยไฮเปอร์พารามิเตอร์ดังนี้

ตารางที่ 7 ค่า Hyperparameter ที่กำหนดใช้ในตัวแบบ Adaboost

Hyperparameter	ค่ากำหนดที่ใช้ใน GridSearchCV
n_estimators	10, 50, 100, 500, 1000
learning_rate	0.0001, 0.001, 0.01, 0.1, 1.0
random_state	1, 23, 456, 7890, 111213

- 1) n\_estimators คือ จำนวนต้นไม้การถดถอย (Regression Tree) ในตัวแบบ RF
- 2) learning\_rate คือ อัตราการเรียนรู้ของอัลกอริทึมที่ใช้ในการเรียนรู้
- 3) random\_state คือ ตัวเลขที่ใช้กำหนดการทำงานของตัวสุ่มตัวเลขของการควบคุมทั้งการสุ่มของการบวตสเตรปของตัวอย่างที่ใช้ในการสร้างต้นไม้และการสุ่มตัวอย่างคุณสมบัติที่ต้องพิจารณาเมื่อมองหาการแยกที่ดีที่สุดในแต่ละโหนด เมื่อกำหนดตัวเลขเดียวกันทุกครั้งที่ในการสร้างตัวแบบจะได้ผลลัพธ์เหมือนเดิม

ผู้วิจัยใช้ GridSearchCV ในการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ ADA เมื่อได้ตัวแบบที่มีค่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำสุดจากการทดลองปรับค่าไฮเปอร์พารามิเตอร์ที่กำหนดให้ทดลองใช้แล้ว จากนั้นให้ทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557

### 3.4.6 การสร้างตัวแบบผสม (Hybrid Model)

วิธีผสมตัวแบบเป็นเทคนิคที่เป็นที่รู้จักที่มีแนวโน้มให้ผลลัพธ์ที่แม่นยำยิ่งขึ้นสำหรับการสร้างแบบจำลองและการพยากรณ์อนุกรมเวลา ซึ่งความแม่นยำเป็นปัจจัยสำคัญในการเลือกรูปแบบการพยากรณ์ที่เหมาะสม ผู้วิจัยเลือกศึกษาการผสมตัวแบบที่มีการพื้นฐานจากการผสมผสานส่วนประกอบของตัวแบบตามโครงสร้างการผสมแบบขนานที่สามารถช่วยลดความเสี่ยงจากการใช้แบบจำลองที่ไม่เหมาะสม และตามโครงสร้างการผสมแบบอนุกรมที่สามารถช่วยปรับปรุงความแม่นยำของการพยากรณ์ให้ดีขึ้นเนื่องจากการตรวจจบบรูปแบบข้อมูลและการสร้างตัวแบบที่ครอบคลุม

#### 3.4.6.1 การสร้างตัวแบบผสมตามโครงสร้างการผสมแบบขนาน

การสร้างตัวแบบผสมตามโครงสร้างการผสมแบบขนาน มีขั้นตอนการผสมตัวแบบดังนี้

ขั้นตอนที่ 1 สร้างตัวแบบเดี่ยวแต่ละตัวแบบโดยใช้ตัวแบบในกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติกลุ่ม ได้แก่ ตัวแบบ SARIMA และตัวแบบ SARIMAX

ขั้นตอนที่ 2 สร้างตัวแบบเดี่ยวแต่ละตัวแบบโดยใช้ตัวแบบการถดถอยเชิงเส้นพหุคูณและตัวแบบในกลุ่มตัวแบบการเรียนรู้ของเครื่อง ได้แก่ ตัวแบบ KNN ตัวแบบ SVR ตัวแบบ ANN และตัวแบบ ADA

ขั้นตอนที่ 3 เลือกตัวแบบเดี่ยวที่มีความแม่นยำสูงและเหมาะสมในกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติมา 1 ตัวแบบซึ่งจะได้ค่าพยากรณ์เป็น  $\hat{Y}_t^1$  และเลือกตัวแบบเดี่ยวจากกลุ่มตัวแบบการถดถอยเชิงเส้นพหุคูณและตัวแบบในกลุ่มตัวแบบการเรียนรู้ของเครื่องมา 1 ตัวแบบ เพื่อใช้เป็นตัวแบบเดี่ยวในการผสมตัวแบบ ซึ่งจะได้ค่าพยากรณ์เป็น  $\hat{Y}_t^2$

ขั้นตอนที่ 4 นำค่าพยากรณ์  $\hat{Y}_t^1$  และ  $\hat{Y}_t^2$  มาเป็นตัวแปรต้นในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณโดยใช้อัลกอริทึม RANSAC (ตัวแบบ RANSAC-REG) และใช้ค่าของปริมาณการบริโภคน้ำมันสำเร็จรูปแต่ละชนิดที่ใช้สำหรับสร้างตัวแบบมาเป็นตัวแปรตาม  $Y_t$

ขั้นตอนที่ 5 นำค่าสัมประสิทธิ์การถดถอยที่ได้จากขั้นตอนที่ 4 มาเป็นค่าถ่วงน้ำหนักในการรวมผลการพยากรณ์ เขียนเป็นสมการได้ดังนี้

$$\hat{Y}_t = \beta_0 + \beta_1 \hat{Y}_t^1 + \beta_2 \hat{Y}_t^2 \quad (3.2)$$

### 3.4.6.2 การสร้างตัวแบบผสมตามโครงสร้างการผสมแบบอนุกรม

การสร้างตัวแบบผสมตามโครงสร้างการผสมแบบอนุกรมมีขั้นตอนการผสมตัวแบบ ดังนี้

ขั้นตอนที่ 1 สร้างตัวแบบเดี่ยวแต่ละตัวแบบโดยใช้ตัวแบบในกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติที่เป็นฟังก์ชันเชิงเส้น ได้แก่ ตัวแบบ SARIMA และตัวแบบ SARIMAX

ขั้นตอนที่ 2 สร้างตัวแบบเดี่ยวแต่ละตัวแบบโดยใช้ตัวแบบการถดถอยเชิงเส้นพหุคูณ และตัวแบบในกลุ่มตัวแบบการเรียนรู้ของเครื่อง ได้แก่ ตัวแบบ SVR ตัวแบบ ANN ตัวแบบ ADA ซึ่งเป็นฟังก์ชันไม่เชิงเส้น

ขั้นตอนที่ 3 เลือกตัวแบบเดี่ยวที่มีความแม่นยำสูงและเหมาะสมในกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติมา 1 ตัวแบบซึ่งจะได้ค่าตัวแปรตามที่ได้จากการสร้างตัวแบบเป็น  $\hat{F}_t^{stat}$  และได้ค่าพยากรณ์เป็น  $\hat{Y}_t^{stat}$  และเลือกตัวแบบเดี่ยวจากกลุ่มตัวแบบการเรียนรู้ของเครื่องมา 1 ตัวแบบ เพื่อใช้เป็นตัวแบบเดี่ยวในการผสมตัวแบบ ซึ่งจะได้ค่าของตัวแปรตามตามที่ได้จากการสร้างตัวแบบเป็น  $\hat{F}_t^{ML}$  และได้ค่าพยากรณ์เป็น  $\hat{Y}_t^{ML}$

ขั้นตอนที่ 4 แบ่งโครงสร้างการผสมแบบอนุกรมเป็น 2 กลุ่มดังนี้

1) โครงสร้างตามลำดับแบบฟังก์ชันเชิงเส้นต่อด้วยฟังก์ชันไม่เชิงเส้น

1.1) เลือกตัวแบบเดี่ยวที่มีความแม่นยำสูงและเหมาะสมในกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติมา 1 ตัวแบบซึ่งเป็นฟังก์ชันเชิงเส้นและจะได้ค่าพยากรณ์เป็น  $\hat{Y}_t^{stat}$

1.2) คำนวณค่าส่วนเหลือ (Residual,  $\hat{\epsilon}_t$ ) จากตัวแบบอนุกรมเวลาเชิงสถิติที่เลือก โดยที่

$$\hat{\epsilon}_t = Y_t - \hat{F}_t^{stat} \quad (3.3)$$

1.3) นำค่าส่วนเหลือที่ได้ไปพยากรณ์ข้อมูลในส่วนที่เป็นฟังก์ชันไม่เชิงเส้นด้วยตัวแบบ ANN หรือตัวแบบ SVR และจะได้ค่าพยากรณ์เป็น  $\hat{Y}_t'^{ML}$

1.4) คำนวณหาค่าการพยากรณ์ของตัวแบบผสมที่ได้จาก

$$\hat{Y}_t = \hat{Y}_t^{stat} + \hat{Y}_t'^{ML} \quad (3.4)$$

2) โครงสร้างตามลำดับแบบฟังก์ชันไม่เชิงเส้นต่อด้วยฟังก์ชันเชิงเส้น

2.1) เลือกตัวแบบเดี่ยว ANN หรือ SVR สำหรับพยากรณ์ข้อมูลในส่วนที่เป็นฟังก์ชันไม่เชิงเส้น และจะได้ค่าพยากรณ์เป็น  $\hat{Y}_t^{ML}$

2.2) คำนวณค่าส่วนเหลือ (Residual,  $\hat{\epsilon}_t$ ) จากตัวแบบสำหรับพยากรณ์ข้อมูลในส่วนที่เป็นฟังก์ชันไม่เชิงเส้น โดยที่

$$\hat{\epsilon}_t = Y_t - \hat{F}_t^{ML} \quad (3.5)$$

2.3) นำค่าส่วนเหลือที่ได้ไปพยากรณ์ข้อมูลในส่วนที่เป็นฟังก์ชันเชิงเส้นด้วยตัวแบบที่เลือกจากกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติ และจะได้ค่าพยากรณ์เป็น  $\hat{Y}_t^{stat}$

2.4) คำนวณหาค่าการพยากรณ์ของตัวแบบผสมที่ได้จาก

$$\hat{Y}_t = \hat{Y}_t^{ML} + \hat{Y}_t^{stat} \quad (3.6)$$

### 3.5 การวัดประสิทธิภาพของตัวแบบพยากรณ์

การวัดประสิทธิภาพของตัวแบบพยากรณ์เพื่อหาตัวแบบที่เหมาะสมจะใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error, MAPE) มีสมการดังนี้

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - \hat{Y}_i|}{Y_i} \times 100\% \quad (3.7)$$

เมื่อ  $Y_i$  คือ ค่าของข้อมูลจริง

$\hat{Y}_i$  คือ ค่าพยากรณ์ของข้อมูล

$n$  คือ จำนวนข้อมูล

ซึ่งจากการทบทวนวรรณกรรมที่เกี่ยวข้องพบว่า MAPE เป็นตัวชี้วัดประสิทธิภาพของตัวแบบพยากรณ์ที่มีการใช้งานในหลาย ๆ งานวิจัยที่เกี่ยวกับการพยากรณ์ความต้องการ ในประสิทธิภาพของตัวแบบพยากรณ์เพื่อคัดเลือกตัวแบบที่มีความแม่นยำมากที่สุด จะใช้ชุดข้อมูลตรวจสอบ (Cross Validation Data) กับค่าพยากรณ์ของแต่ละตัวแบบที่ได้การพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 สำหรับการพยากรณ์ระยะกลาง และกับค่าพยากรณ์ล่วงหน้าครั้งละ 4 ไตรมาสหรือครั้งละ 1 ปี จนครบ 20 ไตรมาส สำหรับการพยากรณ์ระยะสั้น จากนั้น

จึงคำนวณหาค่า MAPE และทำการวิเคราะห์ผลและเปรียบเทียบประสิทธิภาพของแต่ละตัวแบบการพยากรณ์ที่ได้

### 3.6 การคัดเลือกตัวแบบการพยากรณ์

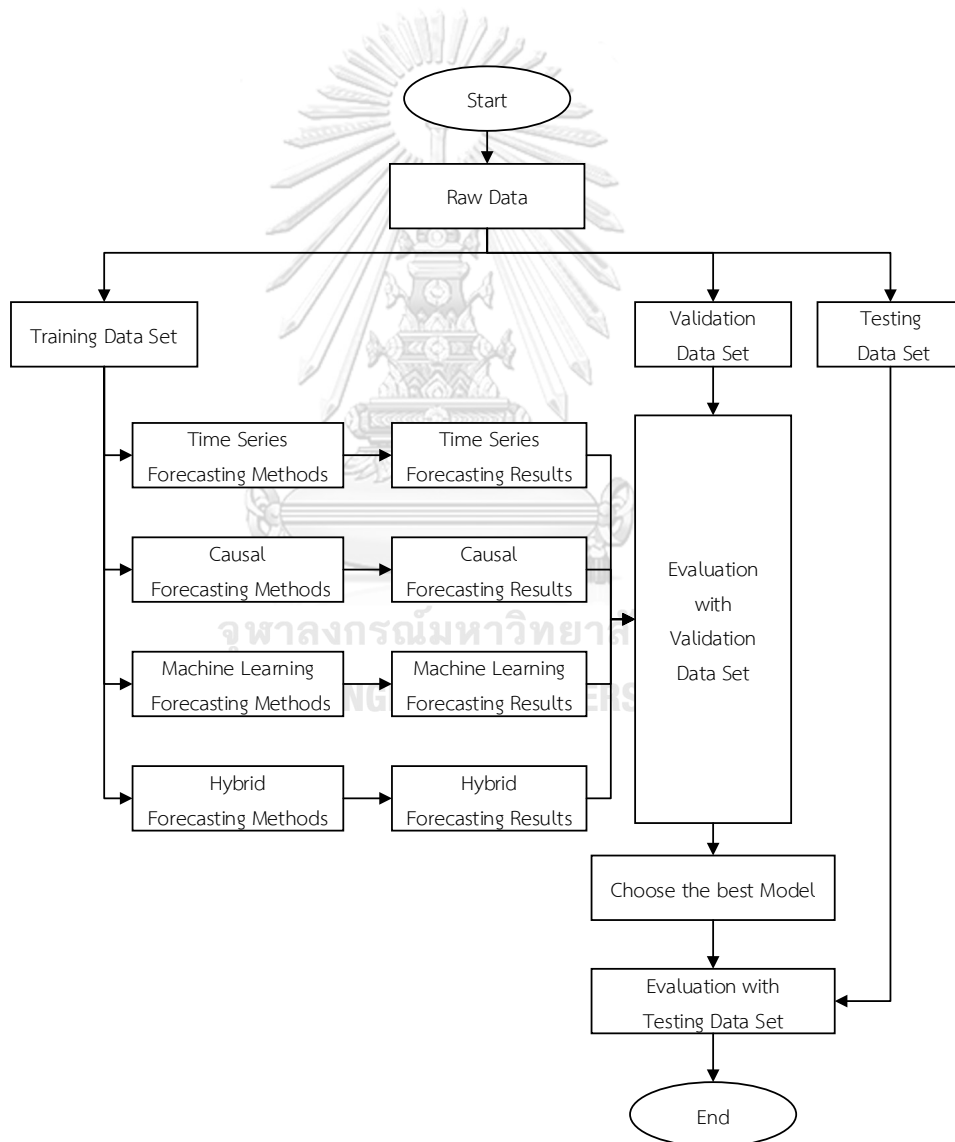
ในงานวิจัยนี้จะใช้วิธีการคัดเลือกตัวแบบที่แม่นยำในการพยากรณ์มากที่สุดจากค่า MAPE ของตัวแบบที่ได้ เมื่อได้ตัวแบบของแต่ละชนิดน้ำมันสำเร็จรูปที่มีความแม่นยำมากที่สุดแล้ว จากนั้นจะทำการพยากรณ์ปริมาณการใช้น้ำมันสำเร็จรูปแต่ละชนิดกับชุดข้อมูลสำหรับทดสอบ (Testing Data) และทำการวัดประสิทธิภาพตัวแบบด้วย MAPE



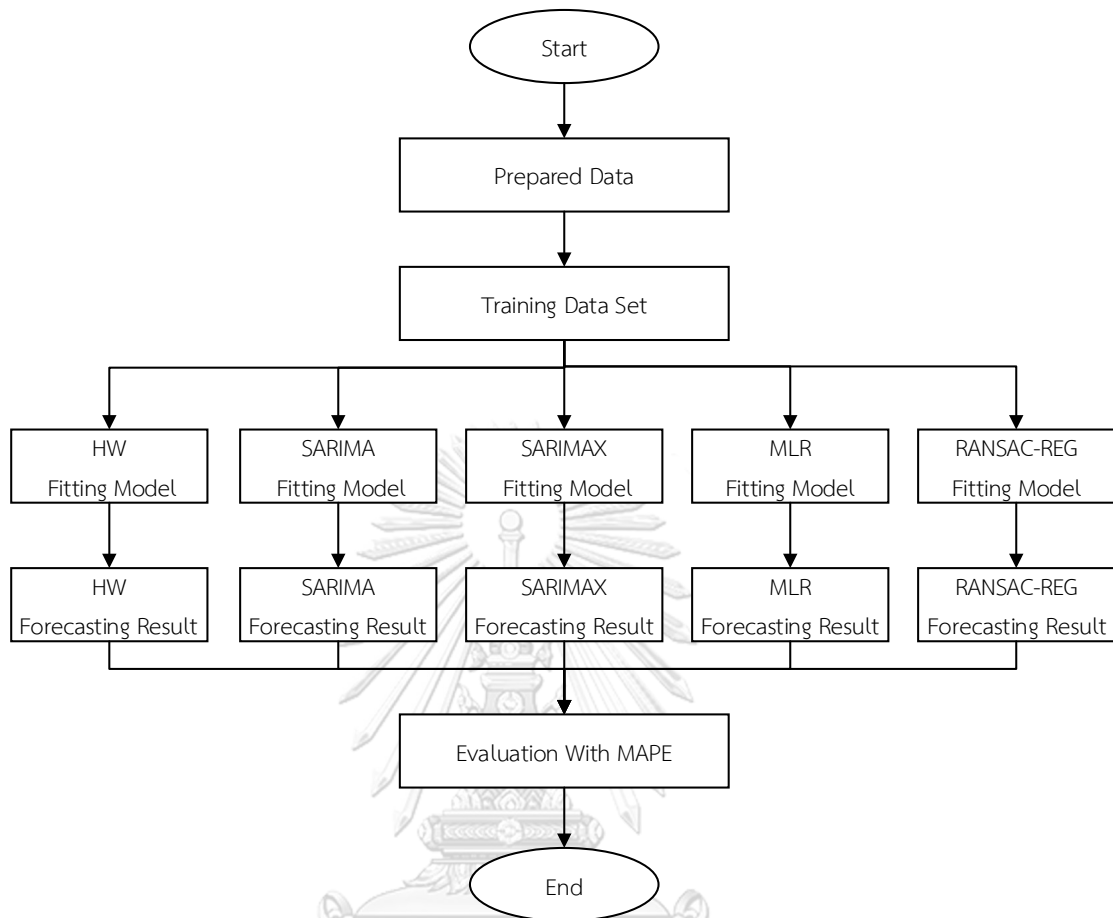
## บทที่ 4

### ผลการดำเนินการ

จากการดำเนินงานวิจัย ได้ทำการพยากรณ์ข้อมูลน้ำมันสำเร็จรูปกลุ่มหลักทั้ง 3 ชนิด ได้แก่ น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน โดยใช้ตัวแบบอนุกรมเวลาเชิงสถิติ ตัวแบบการพยากรณ์เชิงสาเหตุ ตัวแบบการเรียนรู้ของเครื่อง และตัวแบบผสม มีภาพรวมกระบวนการพยากรณ์ ดังแสดงในรูปที่ 22 โดยในส่วนแรกผู้วิจัยจะนำเสนอผลการพยากรณ์โดยใช้ตัวแบบ Holt-Winters ตัวแบบ SARIMA ตัวแบบ SARIMAX ตัวแบบการถดถอยเชิงเส้นพหุคูณ และตัวแบบ RANSAC-REG ซึ่งมีกระบวนการสร้างตัวแบบดังแสดงในรูปที่ 23 โดยมีผลการดำเนินการดังนี้



รูปที่ 22 กระบวนการพยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลัก



รูปที่ 23 กระบวนการสร้างตัวแบบอนุกรมเวลาเชิงสถิติและตัวแบบการพยากรณ์เชิงสาเหตุ  
จุฬาลงกรณ์มหาวิทยาลัย

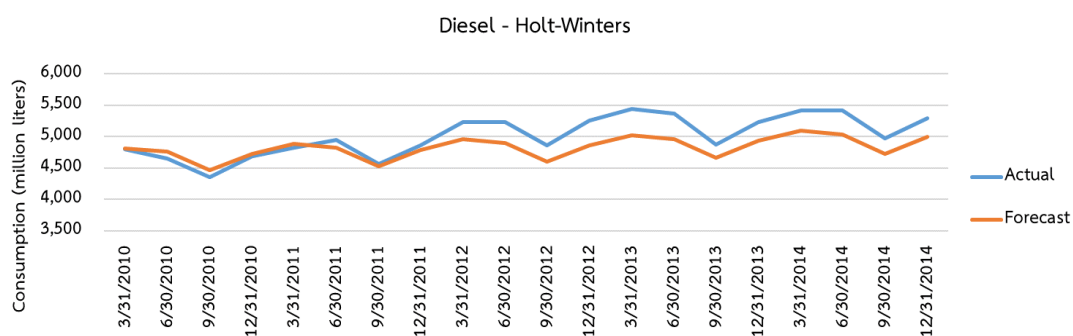
#### 4.1 ตัวแบบ Holt-Winters

##### 1) ตัวแบบการพยากรณ์ระยะกลาง

ในการสร้างตัวแบบ Holt-Winters และการปรับค่าพารามิเตอร์ของตัวแบบให้เหมาะสมที่จะทำให้ตัวแบบมีค่า MAPE น้อยที่สุดสำหรับการพยากรณ์ระยะกลาง ผลลัพธ์ของตัวแบบของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินได้ค่า  $\alpha$   $\beta$  และ  $\gamma$  ดังแสดงในตารางที่ 8 เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 24 25 และ 26

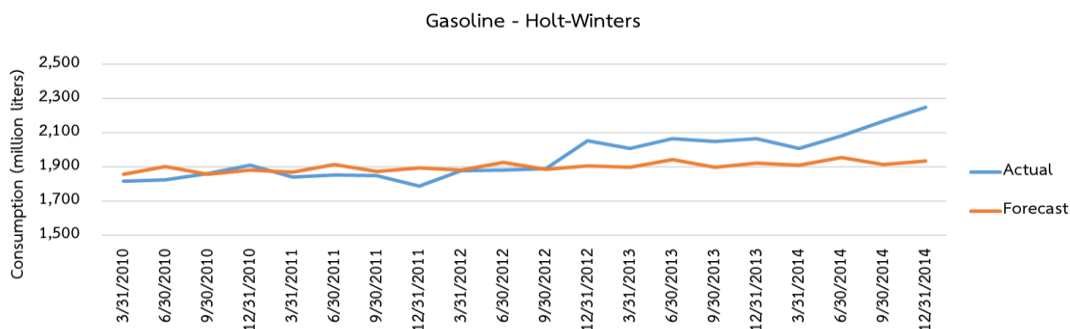
ตารางที่ 8 ค่าพารามิเตอร์ของตัวแบบ Holt-Winters ที่ได้จากการสร้างตัวแบบด้วยชุดข้อมูลฝึกสอนสำหรับการพยากรณ์ระยะกลาง

ชนิดน้ำมันสำเร็จรูป	ค่าพารามิเตอร์ของของตัวแบบ Holt-Winters ที่ได้		
	$\alpha$	$\beta$	$\gamma$
ดีเซล	0.9242857	0.0001000	0.0757143
เบนซิน	0.7121428	0.0593454	0.2110952
เครื่องบิน	0.8040465	0.0838543	0.0051869

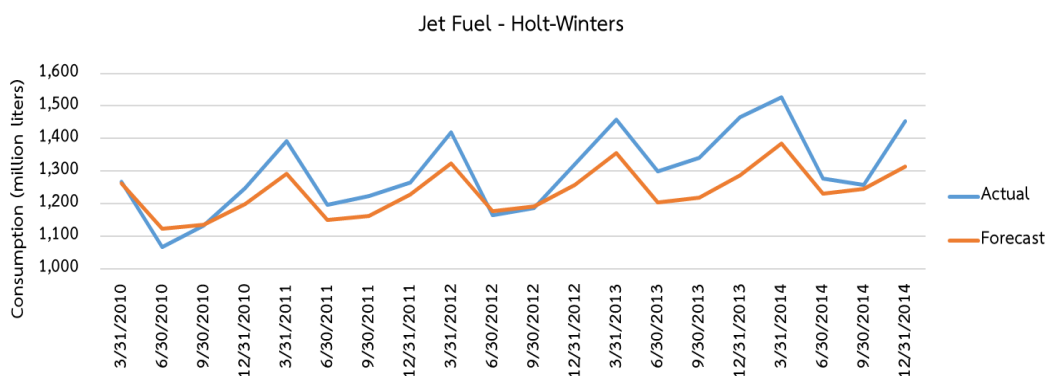


รูปที่ 24 การพยากรณ์ระยะกลางด้วยตัวแบบ Holt-Winters ของน้ำมันดีเซล





รูปที่ 25 การพยากรณ์ระยะกลางด้วยตัวแบบ Holt-Winters ของน้ำมันเบนซิน



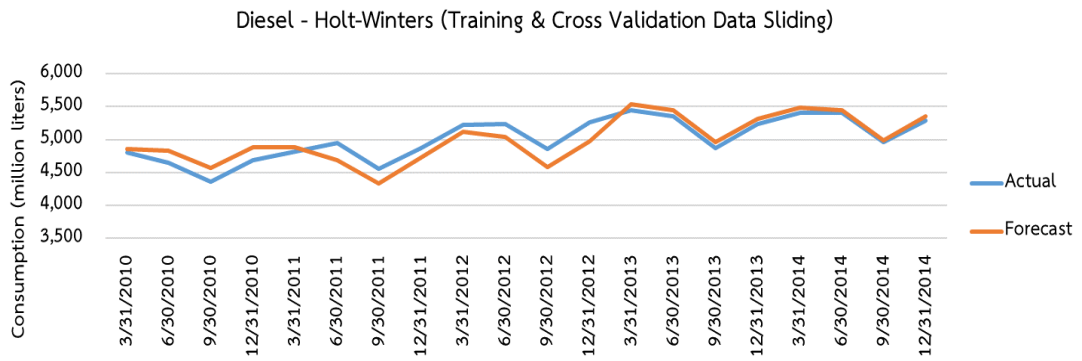
รูปที่ 26 การพยากรณ์ระยะกลางด้วยตัวแบบ Holt-Winters ของน้ำมันเครื่องบิน

## 2) ตัวแบบการพยากรณ์ระยะสั้น

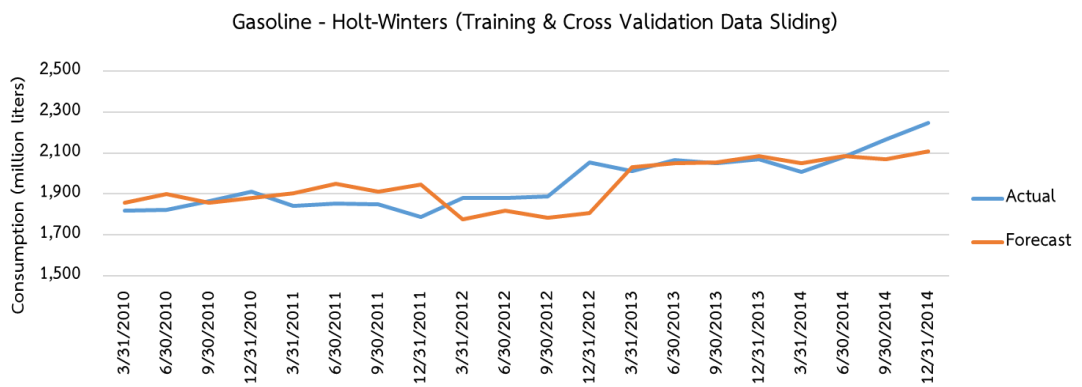
ในส่วนของการพยากรณ์ระยะสั้นได้ผลลัพธ์ของตัวแบบของน้ำมันดีเซล น้ำมันเบนซิน และ น้ำมันเครื่องบินดังแสดงในตารางที่ 9 ซึ่งค่าพารามิเตอร์ของแต่ละช่วงข้อมูลที่ใช้สร้างตัวแบบ Holt-Winters จะไม่เท่ากัน เมื่อทำการพยากรณ์ไปล่วงหน้า 4 ไตรมาสหรือจำนวน 1 ปี โดยมีการเลื่อน (Sliding) การพยากรณ์ไปครั้งละ 4 ไตรมาส จำนวน 5 ครั้ง จนครบ 20 ไตรมาส ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 27 28 และ 29

ตารางที่ 9 ค่าพารามิเตอร์ของตัวแบบ Holt-Winters ที่ได้จากการสร้างตัวแบบด้วยชุดข้อมูลฝึกสอน สำหรับการพยากรณ์ระยะสั้น จำนวน 5 ช่วงเวลา

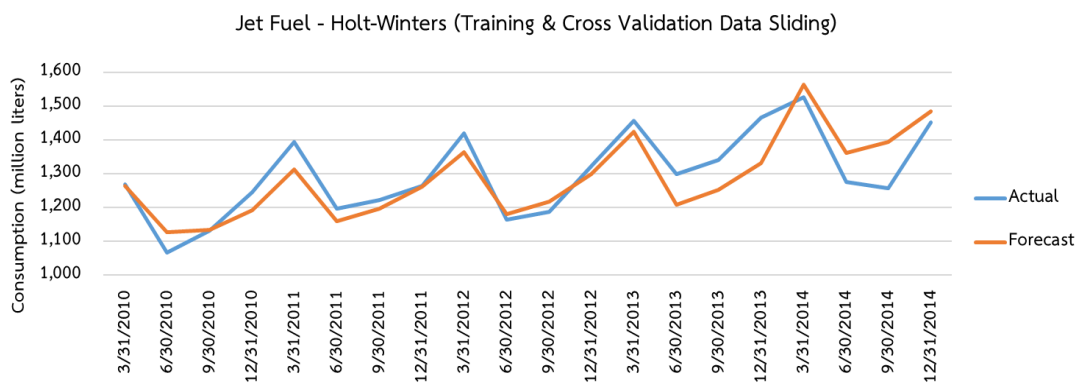
ชนิดน้ำมันสำเร็จรูป	ช่วงข้อมูลที่ใช้สร้างตัวแบบ	ค่าพารามิเตอร์ของของตัวแบบ Holt-Winters ที่ได้		
		$\alpha$	$\beta$	$\gamma$
ดีเซล	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	0.9242857	0.0001000	0.0757143
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	0.7729772	0.1328191	0.2043432
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	0.7475000	0.0897000	0.2525000
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	0.7475000	0.0299000	0.2525000
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	0.6448683	0.1318932	0.3288858
เบนซิน	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	0.9685940	0.0739012	0.0114681
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	0.9364432	0.0245409	0.0133901
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	0.9947734	0.0018272	0.0038087
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	0.8943975	0.0629896	0.1056025
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	0.9977361	0.0290222	0.0005202
เครื่องบิน	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	0.7341436	0.0532994	0.2594083
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	0.7543179	0.0185077	0.2407863
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	0.7259644	0.0030969	0.2702704
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	0.6989494	0.0033514	0.2946056
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	0.7468480	0.0000000	0.2531520



รูปที่ 27 การพยากรณ์ระยะสั้นด้วยตัวแบบ Holt-Winters ของน้ำมันดีเซล



รูปที่ 28 การพยากรณ์ระยะสั้นด้วยตัวแบบ Holt-Winters ของน้ำมันเบนซิน



รูปที่ 29 การพยากรณ์ระยะสั้นด้วยตัวแบบ Holt-Winters ของน้ำมันเครื่องบิน

## 4.2 ตัวแบบ SARIMA

### 1) ตัวแบบการพยากรณ์ระยะกลาง

ผลลัพธ์ของการสร้างตัวแบบ SARIMA สำหรับการพยากรณ์ระยะกลางจากชุดข้อมูลฝึกสอนของปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน จะถูกเลือกจากค่าที่ทำให้ Akaike Information Criterion (AIC) มีค่าต่ำที่สุดดังแสดงในรูปที่ 30 31 และ 32 ซึ่งเขียนเป็นตัวแบบอย่างง่ายได้ดังนี้ SARIMA(0,1,0)(0,1,1)<sub>4</sub> SARIMA(0,1,1)(0,1,1)<sub>4</sub> และ SARIMA(0,1,0)(0,1,1)<sub>4</sub> ตามลำดับ โดยมีค่า AIC ที่ต่ำที่สุดเป็น 846.182 670.326 และ 667.046 ตามลำดับ เมื่อทำการทดสอบ z- test กับสัมประสิทธิ์ (Coefficients) ของแต่ละตัวแบบพบว่าสัมประสิทธิ์ทุกตัวของทุกตัวแบบมีนัยสำคัญภายใต้ระดับความเชื่อมั่น 95% ดังแสดงในรูปที่ 33 34 และ 35 เมื่อนำข้อมูลความผิดพลาด (Residual) ที่ได้จากการหาค่าความแตกต่างระหว่างช่วงค่าชุดข้อมูลฝึกสอนกับช่วงค่าที่คำนวณได้จากตัวแบบ SARIMA ไปตรวจสอบรูปแบบ (Diagnostic Checking) ว่ารูปแบบที่สร้างขึ้นเหมาะสมกับอนุกรมเวลาหรือไม่ โดยต้องสอดคล้องกับเงื่อนไขดังต่อไปนี้ 1) กราฟ Standardized Residual มีลักษณะการกระจายตัวอยู่รอบ ๆ ค่า 0 2) ตรวจสอบกราฟ ACF ของค่าความผิดพลาดจากตัวแบบโดยต้องไม่เหลือรูปแบบอนุกรมเวลาอื่นใดเช่น แนวโน้ม ฤดูกาล นอกเหนือจากรูปแบบสุ่ม (Random) 3) ตรวจสอบลักษณะการกระจายตัวของข้อมูลค่าความผิดพลาดต้องมีการแจกแจงแบบปกติ (Normal Distribution) ผลลัพธ์ของการตรวจสอบรูปแบบของตัวแบบปริมาณการบริโภคน้ำมันสำเร็จรูปทั้ง 3 กลุ่มพบว่าต้องสอดคล้องกับเงื่อนไขข้างต้นทุกประการ ดังนั้นตัวแบบที่ได้มีความเหมาะสมกับข้อมูลอนุกรมเวลา

```

Performing stepwise search to minimize aic
ARIMA(0,1,0)(0,1,0)[4] : AIC=870.562, Time=0.02 sec
ARIMA(1,1,0)(1,1,0)[4] : AIC=856.733, Time=0.11 sec
ARIMA(0,1,1)(0,1,1)[4] : AIC=847.422, Time=0.11 sec
ARIMA(0,1,1)(0,1,0)[4] : AIC=872.513, Time=0.04 sec
ARIMA(0,1,1)(1,1,1)[4] : AIC=849.067, Time=0.19 sec
ARIMA(0,1,1)(0,1,2)[4] : AIC=848.965, Time=0.16 sec
ARIMA(0,1,1)(1,1,0)[4] : AIC=856.799, Time=0.08 sec
ARIMA(0,1,1)(1,1,2)[4] : AIC=inf, Time=0.25 sec
ARIMA(0,1,0)(0,1,1)[4] : AIC=846.182, Time=0.05 sec
ARIMA(0,1,0)(1,1,1)[4] : AIC=847.992, Time=0.09 sec
ARIMA(0,1,0)(0,1,2)[4] : AIC=847.957, Time=0.10 sec
ARIMA(0,1,0)(1,1,0)[4] : AIC=855.167, Time=0.04 sec
ARIMA(0,1,0)(1,1,2)[4] : AIC=850.006, Time=0.24 sec
ARIMA(1,1,0)(0,1,1)[4] : AIC=847.425, Time=0.10 sec
ARIMA(1,1,1)(0,1,1)[4] : AIC=inf, Time=0.28 sec
ARIMA(0,1,0)(0,1,1)[4] intercept : AIC=847.857, Time=0.12 sec

Best model: ARIMA(0,1,0)(0,1,1)[4]
Total fit time: 1.985 seconds

```

รูปที่ 30 การหาค่าเหมาะสมของพารามิเตอร์ในตัวแบบ SARIMA(p,d,q)(P,D,Q)<sub>s</sub>  
ของน้ำมันดีเซล

Performing stepwise search to minimize aic

ARIMA(0,1,0)(0,1,0)[4]	: AIC=690.079, Time=0.01 sec
ARIMA(1,1,0)(1,1,0)[4]	: AIC=674.470, Time=0.11 sec
ARIMA(0,1,1)(0,1,1)[4]	: AIC=670.326, Time=0.09 sec
ARIMA(0,1,1)(0,1,0)[4]	: AIC=690.534, Time=0.03 sec
ARIMA(0,1,1)(1,1,1)[4]	: AIC=672.795, Time=0.17 sec
ARIMA(0,1,1)(0,1,2)[4]	: AIC=672.316, Time=0.16 sec
ARIMA(0,1,1)(1,1,0)[4]	: AIC=671.092, Time=0.07 sec
ARIMA(0,1,1)(1,1,2)[4]	: AIC=672.736, Time=0.47 sec
ARIMA(0,1,0)(0,1,1)[4]	: AIC=673.581, Time=0.05 sec
ARIMA(1,1,1)(0,1,1)[4]	: AIC=671.499, Time=0.12 sec
ARIMA(0,1,2)(0,1,1)[4]	: AIC=671.297, Time=0.14 sec
ARIMA(1,1,0)(0,1,1)[4]	: AIC=672.037, Time=0.10 sec
ARIMA(1,1,2)(0,1,1)[4]	: AIC=673.288, Time=0.40 sec
ARIMA(0,1,1)(0,1,1)[4] intercept	: AIC=671.388, Time=0.12 sec

Best model: ARIMA(0,1,1)(0,1,1)[4]  
Total fit time: 2.044 seconds

รูปที่ 31 การหาค่าเหมาะสมของพารามิเตอร์ในตัวแทน SARIMA(p,d,q)(P,D,Q)<sub>s</sub>  
ของน้ำมันเบนซิน

Performing stepwise search to minimize aic

ARIMA(0,1,0)(0,1,0)[4]	: AIC=687.911, Time=0.02 sec
ARIMA(1,1,0)(1,1,0)[4]	: AIC=678.044, Time=0.06 sec
ARIMA(0,1,1)(0,1,1)[4]	: AIC=668.966, Time=0.09 sec
ARIMA(0,1,1)(0,1,0)[4]	: AIC=688.594, Time=0.03 sec
ARIMA(0,1,1)(1,1,1)[4]	: AIC=670.810, Time=0.13 sec
ARIMA(0,1,1)(0,1,2)[4]	: AIC=670.720, Time=0.14 sec
ARIMA(0,1,1)(1,1,0)[4]	: AIC=677.962, Time=0.07 sec
ARIMA(0,1,1)(1,1,2)[4]	: AIC=672.955, Time=0.23 sec
ARIMA(0,1,0)(0,1,1)[4]	: AIC=667.046, Time=0.05 sec
ARIMA(0,1,0)(1,1,1)[4]	: AIC=668.864, Time=0.09 sec
ARIMA(0,1,0)(0,1,2)[4]	: AIC=668.762, Time=0.10 sec
ARIMA(0,1,0)(1,1,0)[4]	: AIC=676.328, Time=0.04 sec
ARIMA(0,1,0)(1,1,2)[4]	: AIC=671.040, Time=0.17 sec
ARIMA(1,1,0)(0,1,1)[4]	: AIC=668.989, Time=0.08 sec
ARIMA(1,1,1)(0,1,1)[4]	: AIC=inf, Time=0.29 sec
ARIMA(0,1,0)(0,1,1)[4] intercept	: AIC=669.006, Time=0.06 sec

Best model: ARIMA(0,1,0)(0,1,1)[4]  
Total fit time: 1.639 seconds

รูปที่ 32 การหาค่าเหมาะสมของพารามิเตอร์ในตัวแทน SARIMA(p,d,q)(P,D,Q)<sub>s</sub>  
ของน้ำมันเครื่องบิน

SARIMAX Results

Dep. Variable:	y	No. Observations:	68
Model:	SARIMAX(0, 1, 0)x(0, 1, [1], 4)	Log Likelihood	-421.091
Date:	Sun, 24 Jan 2021	AIC	846.182
Time:	10:39:55	BIC	850.468
Sample:	0	HQIC	847.868
	- 68		

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ma.S.L4	-0.6946	0.113	-6.151	0.000	-0.916	-0.473
sigma2	3.605e+04	5247.169	6.871	0.000	2.58e+04	4.63e+04
Ljung-Box (L1) (Q):	0.83	Jarque-Bera (JB):	6.72			
Prob(Q):	0.36	Prob(JB):	0.03			
Heteroskedasticity (H):	1.51	Skew:	-0.61			
Prob(H) (two-sided):	0.35	Kurtosis:	4.03			

รูปที่ 33 ผลลัพธ์การสร้างตัวแบบ SARIMA ของน้ำมันดีเซล

SARIMAX Results

Dep. Variable:	y	No. Observations:	68
Model:	SARIMAX(0, 1, 1)x(0, 1, 1, 4)	Log Likelihood	-332.163
Date:	Sun, 24 Jan 2021	AIC	670.326
Time:	10:39:58	BIC	676.755
Sample:	0	HQIC	672.854
	- 68		

Covariance Type: opg

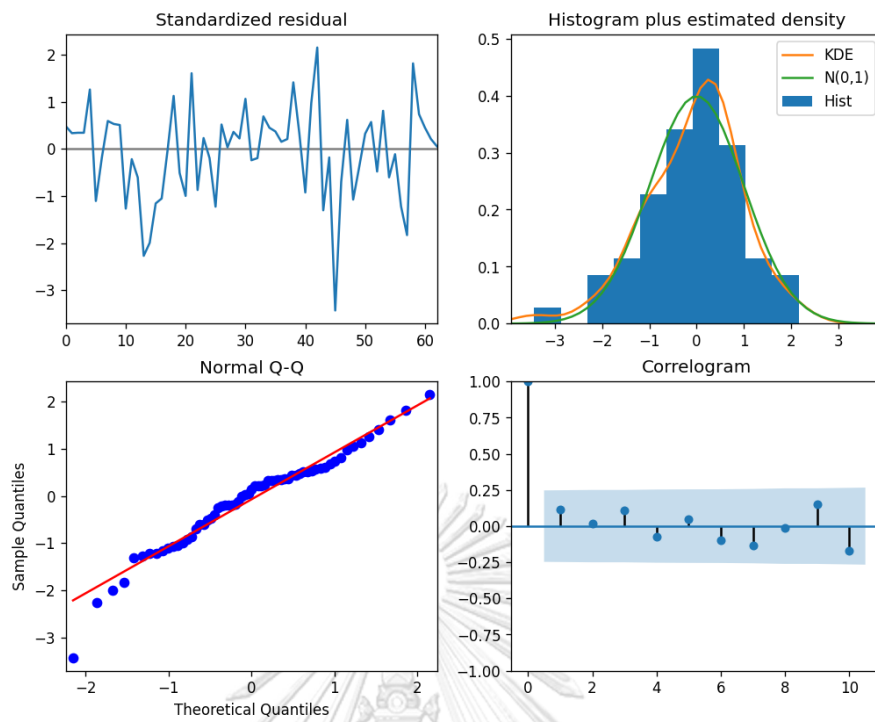
	coef	std err	z	P> z	[0.025	0.975]
ma.L1	0.3428	0.126	2.720	0.007	0.096	0.590
ma.S.L4	-0.6898	0.131	-5.269	0.000	-0.946	-0.433
sigma2	2129.3768	378.733	5.622	0.000	1387.074	2871.680
Ljung-Box (L1) (Q):	0.18	Jarque-Bera (JB):	0.87			
Prob(Q):	0.67	Prob(JB):	0.65			
Heteroskedasticity (H):	1.54	Skew:	0.21			

รูปที่ 34 ผลลัพธ์การสร้างตัวแบบ SARIMA ของน้ำมันเบนซิน

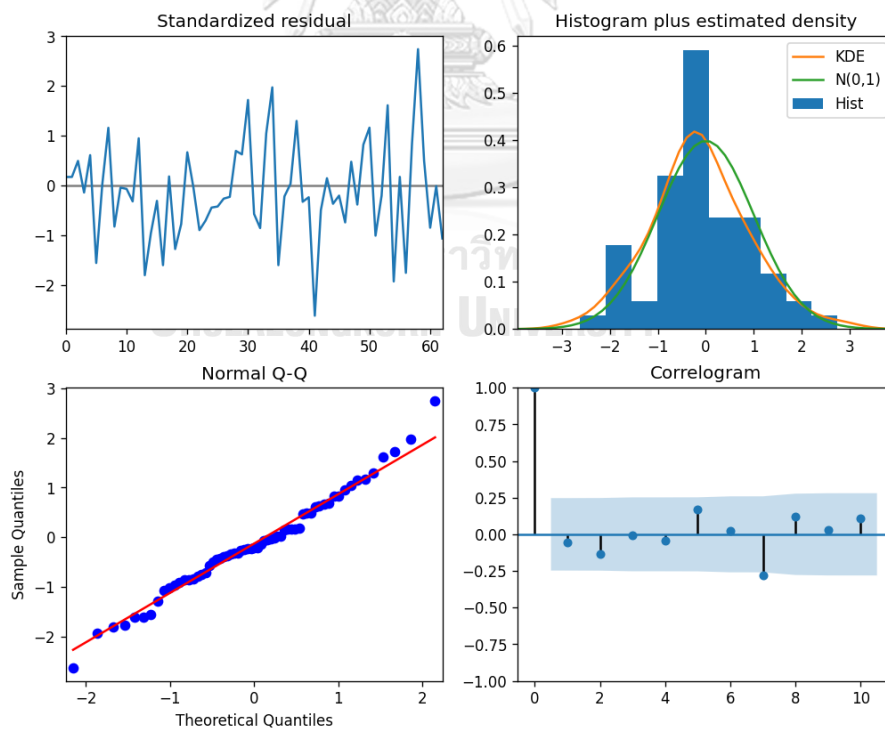
SARIMAX Results						
<b>Dep. Variable:</b>	y			<b>No. Observations:</b>	68	
<b>Model:</b>	SARIMAX(0, 1, 0)x(0, 1, [1], 4)			<b>Log Likelihood</b>	-331.523	
<b>Date:</b>	Sun, 24 Jan 2021			<b>AIC</b>	667.046	
<b>Time:</b>	10:40:03			<b>BIC</b>	671.333	
<b>Sample:</b>	0			<b>HQIC</b>	668.732	
	- 68					
<b>Covariance Type:</b> opg						
	<b>coef</b>	<b>std err</b>	<b>z</b>	<b>P&gt; z </b>	<b>[0.025</b>	<b>0.975]</b>
<b>ma.S.L4</b>	-0.6751	0.104	-6.473	0.000	-0.880	-0.471
<b>sigma2</b>	2097.1760	239.309	8.763	0.000	1628.139	2566.213
<b>Ljung-Box (L1) (Q):</b>	0.04		<b>Jarque-Bera (JB):</b> 32.24			
<b>Prob(Q):</b>	0.83		<b>Prob(JB):</b> 0.00			
<b>Heteroskedasticity (H):</b>	6.03		<b>Skew:</b> -0.89			
<b>Prob(H) (two-sided):</b>	0.00		<b>Kurtosis:</b> 6.02			

รูปที่ 35 ผลลัพธ์การสร้างตัวแบบ SARIMA ของน้ำมันเครื่องบิน

เมื่อนำข้อมูลความผิดพลาด (Residual) ที่ได้จากการหาค่าความแตกต่างระหว่างช่วงค่าชุดข้อมูลฝึกสอนกับช่วงค่าที่คำนวณได้จากตัวแบบ SARIMA ไปตรวจสอบรูปแบบ (Diagnostic Checking) ผลลัพธ์ที่ได้พบว่าต้องคล้องกับเงื่อนไขข้างต้นทุกประการดังแสดงในรูปที่ 36 37 และ 38 ดังนั้นตัวแบบที่ได้มีความเหมาะสมกับข้อมูลอนุกรมเวลา

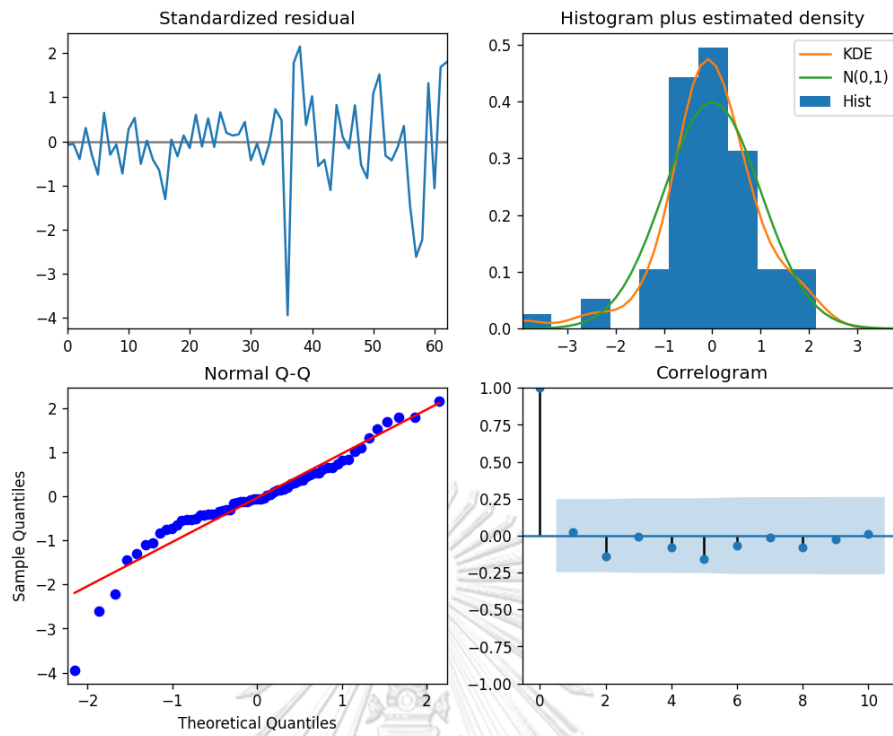


รูปที่ 36 การตรวจสอบรูปแบบ (Diagnostic Checking) ของตัวแบบ SARIMA ของน้ำมันดีเซล



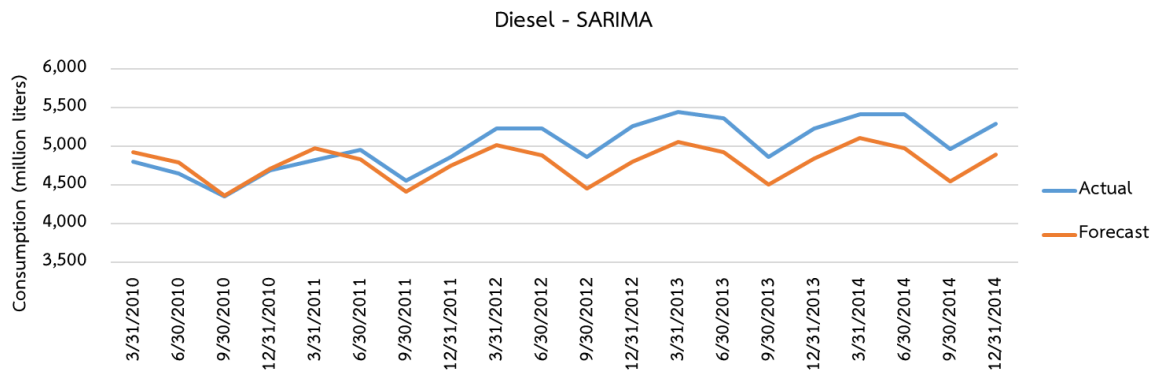
รูปที่ 37 การตรวจสอบรูปแบบ (Diagnostic Checking) ของตัวแบบ SARIMA ของน้ำมันเบนซิน



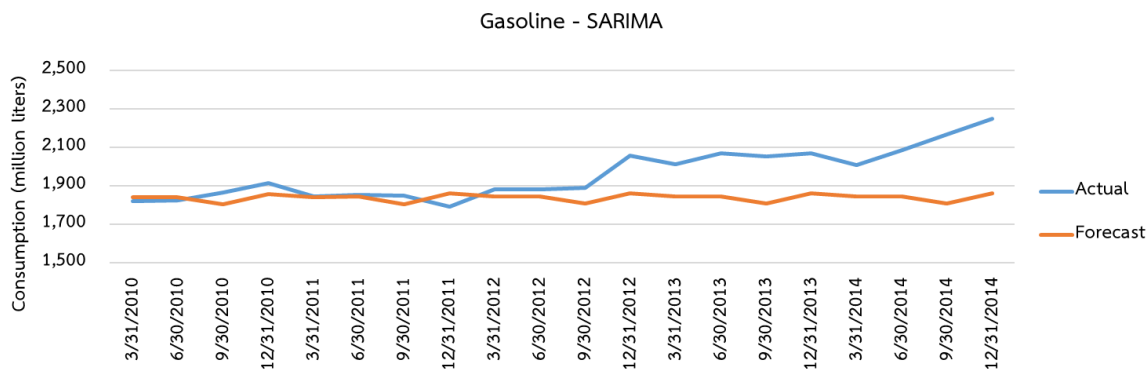


รูปที่ 38 การตรวจสอบรูปแบบ (Diagnostic Checking) ของตัวแบบ SARIMA ของน้ำมันเครื่องบิน

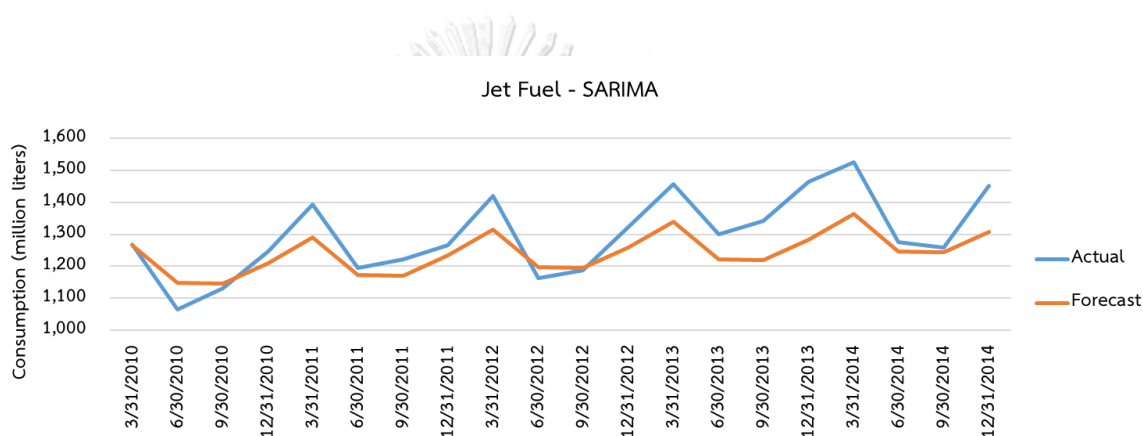
เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 39 40 และ 41



รูปที่ 39 การพยากรณ์ระยะกลางด้วยตัวแบบ SARIMA ของน้ำมันดีเซล



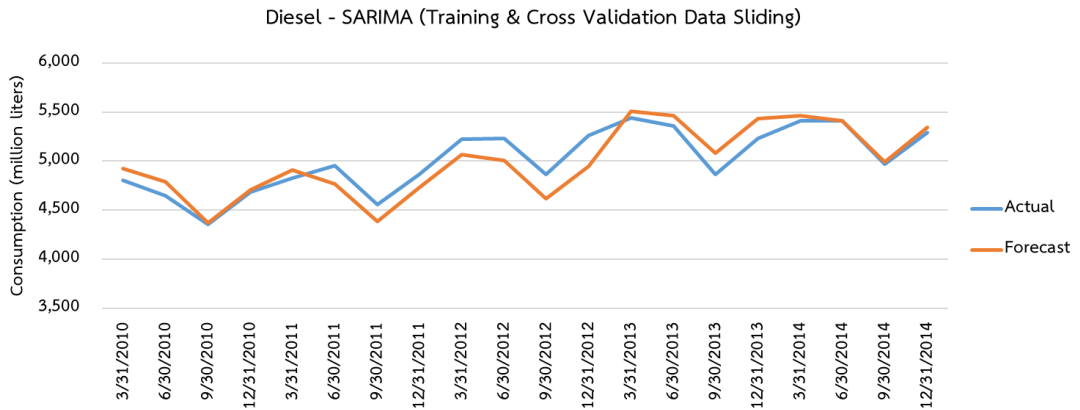
รูปที่ 40 การพยากรณ์ระยะกลางด้วยตัวแบบ SARIMA ของน้ำมันเบนซิน



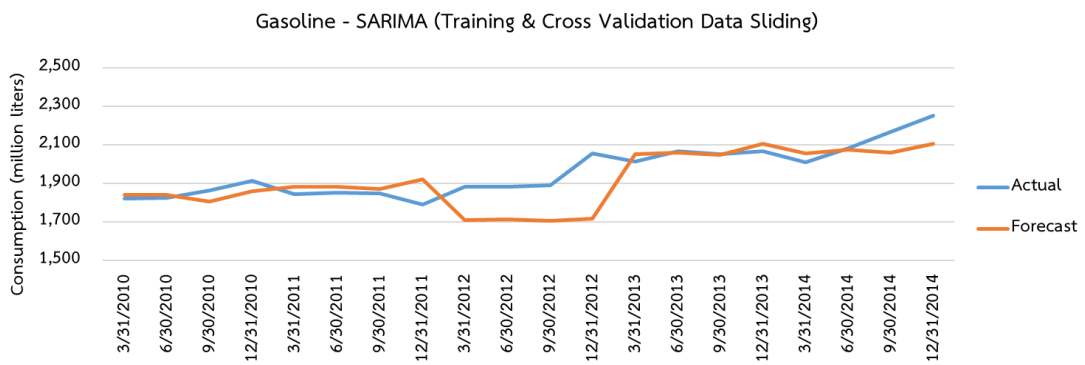
รูปที่ 41 การพยากรณ์ระยะกลางด้วยตัวแบบ SARIMA ของน้ำมันเครื่องบิน

## 2) ตัวแบบการพยากรณ์ระยะสั้น

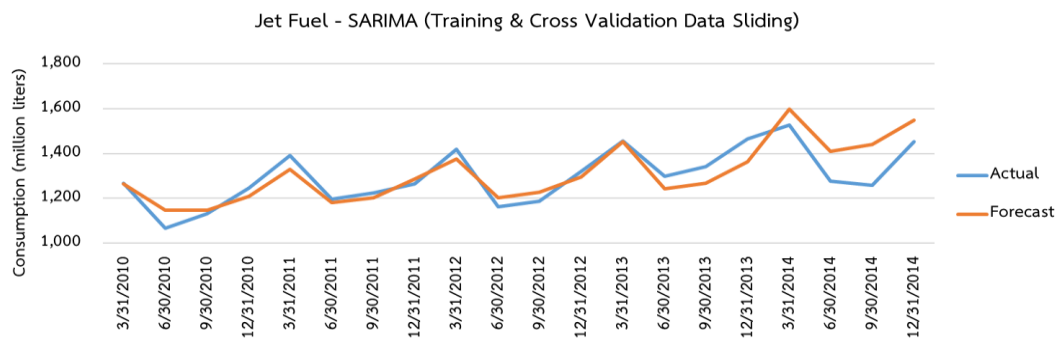
ผลลัพธ์ของตัวแบบ SARIMA ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน สำหรับการพยากรณ์ระยะสั้นซึ่งใช้การเลื่อนชุดข้อมูลฝึกสอนไปข้างหน้าในการสร้างตัวแบบครึ่งละ 4 ไตรมาส ได้ตัวแบบ SARIMA ดังแสดงในตารางที่ 10 จะเห็นได้ว่าตัวแบบ SARIMA ของน้ำมันดีเซลและน้ำมันเครื่องบินในแต่ละช่วงข้อมูลฝึกสอนที่ใช้นั้นได้ตัวแบบที่เหมือนกันทั้ง 5 ช่วง แต่ตัวแบบ SARIMA ของน้ำมันดีเซลได้ตัวแบบในช่วงที่ 4 และ 5 แตกต่างไปจากตัวแบบในช่วงที่ 1 2 และ 3 เมื่อทำการตรวจสอบรูปแบบของตัวแบบ SARIMA เช่นเดียวกันกับตัวแบบ SARIMA สำหรับการพยากรณ์ระยะกลาง ซึ่งผลลัพธ์ที่ได้พบว่ามี ความคล้อยกับเงื่อนไขดังกล่าว จากนั้นทำการพยากรณ์ไปล่วงหน้า 4 ไตรมาสหรือจำนวน 1 ปี โดยมีการเลื่อน (Sliding) การพยากรณ์ไปครึ่งละ 4 ไตรมาส จำนวน 5 ครั้ง จนครบ 20 ไตรมาส ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 42 43 และ 44



รูปที่ 42 การพยากรณ์ระยะสั้นด้วยตัวแบบ SARIMA ของน้ำมันดีเซล



รูปที่ 43 การพยากรณ์ระยะสั้นด้วยตัวแบบ SARIMA ของน้ำมันเบนซิน



รูปที่ 44 การพยากรณ์ระยะสั้นด้วยตัวแบบ SARIMA ของน้ำมันเครื่องบิน

ตารางที่ 10 ตัวแบบ SARIMA ของน้ำมันสำเร็จรูปที่ได้จากการสร้างตัวแบบด้วยชุดข้อมูลฝีกสอน สำหรับการพยากรณ์ระยะสั้น จำนวน 5 ช่วงเวลา

ชนิดน้ำมันสำเร็จรูป	ช่วงข้อมูลที่ใช้สร้างตัวแบบ	ตัวแบบ SARIMA ที่ได้
ดีเซล	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
เบนซิน	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	SARIMA(0,1,1)(0,1,1) <sub>4</sub>
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	SARIMA(0,1,1)(0,1,1) <sub>4</sub>
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	SARIMA(0,1,1)(0,1,1) <sub>4</sub>
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
เครื่องบิน	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	SARIMA(0,1,0)(0,1,1) <sub>4</sub>

จากตารางที่ 10 จะเห็นได้ว่าตัวแบบ SARIMA ของน้ำมันดีเซลและน้ำมันเครื่องบินในแต่ละช่วงเวลาได้ผลลัพธ์เป็นตัวแบบเดียวกันทั้งหมด ส่วนของตัวแบบ SARIMA ของน้ำมันเบนซินในช่วงที่ 4 และ 5 ได้ผลลัพธ์แตกต่างจากช่วงที่ 1 2 และ 3 ส่วนที่แตกต่างของตัวแบบคือเทอมของ  $q$  หรือเรียกว่าอันดับของตัวแบบ Moving Average ที่ไม่มีฤดูกาล ซึ่งผู้วิจัยมีความเห็นว่าน่าจะเกิดจากข้อมูลปริมาณการบริโภคน้ำมันเบนซินในช่วงปี พ.ศ. 2536 ถึง พ.ศ. 2538 (ค.ศ. 1993 ถึง ค.ศ. 1995) นั้น มีรูปแบบแนวโน้มในลักษณะชันขึ้นมาก แต่หลังจากปี พ.ศ. 2539 จนถึง พ.ศ. 2557 (ค.ศ. 1996 ถึง ค.ศ. 2014) พบว่าข้อมูลมีรูปแบบค่อนข้างคงที่ ดังแสดงในรูปที่ 13 เมื่อทำสร้างตัวแบบการพยากรณ์ระยะสั้นแบบการเลื่อนตำแหน่งช่วงเวลา ในช่วงที่ 4 และ 5 ตัวแบบจึงไม่ครอบคลุมในช่วงที่มีแนวโน้ม

ชั้นขึ้นมาก ซึ่งจะครอบคลุมช่วงที่ค่อนข้างคงที่เป็นส่วนใหญ่ ผลลัพธ์ของการสร้างตัวแบบน้ำมันเบนซินช่วงที่ 4 และ 5 จึงได้ผลลัพธ์ที่เหมือนกันและแตกต่างไปจากช่วงอื่น

### 4.3 ตัวแบบการถดถอยเชิงเส้นพหุคูณ (MLR)

#### 1) การพยากรณ์ระยะกลาง

การศึกษาปัจจัยที่ส่งผลต่อการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยเป็นหนึ่งในผลที่คาดว่าจะได้รับจากงานวิจัยนี้ ซึ่งตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับการพยากรณ์ระยะกลางของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินสามารถอธิบายได้ด้วยตาราง ANOVA จากโปรแกรม Minitab ดังแสดงในรูปที่ 45 46 และ 47 ตามลำดับ

#### Stepwise Selection of Terms

$\alpha$  to enter = 0.05,  $\alpha$  to remove = 0.05

#### Regression Equation

Diesel Consumption = 5273 + 0.9569 Gasoline Production - 53.64 Diesel Price  
+ 0.002085 ExpeGDP + 0.001169 Total Vehicle - 0.000095 Population

#### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5273	1825	2.89	0.005	
Gasoline Production	0.9569	0.0988	9.69	0.000	2.72
Diesel Price	-53.64	6.75	-7.95	0.000	5.72
ExpeGDP	0.002085	0.000275	7.58	0.000	12.52
Total Vehicle	0.001169	0.000225	5.20	0.000	3.27
Population	-0.000095	0.000035	-2.76	0.008	6.61

#### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
177.377	90.90%	90.16%	88.92%

#### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	19478935	3895787	123.82	0.000
Gasoline Production	1	2954155	2954155	93.89	0.000
Diesel Price	1	1987432	1987432	63.17	0.000
ExpeGDP	1	1806252	1806252	57.41	0.000
Total Vehicle	1	850984	850984	27.05	0.000
Population	1	239122	239122	7.60	0.008
Error	62	1950687	31463		
Total	67	21429621			

รูปที่ 45 ผลลัพธ์การวิเคราะห์การถดถอยของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง

จากรูปที่ 45 พบว่าผลลัพธ์จากการทำการถดถอยแบบขั้นตอนของข้อมูลน้ำมันดีเซลได้เลือกตัวแปรอิสระ ได้แก่ ปริมาณการผลิตน้ำมันเบนซิน ราคา น้ำมันดีเซล ผลิตรถยนต์รวม จำนวนรถใหม่ที่จะจดทะเบียน และจำนวนประชากร โดยตัวแปรอิสระดังกล่าวมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันสำเร็จรูปที่ระดับนัยสำคัญ 0.05 โดยผู้วิจัยเห็นด้วยกับความสัมพันธ์ของปริมาณการบริโภคน้ำมันดีเซลที่มีความสัมพันธ์เชิงเส้นกับราคา น้ำมันดีเซลในทิศทางตรงข้าม อาจเป็นเพราะเมื่อราคาน้ำมันดีเซลสูงขึ้น ผู้บริโภคอาจเปลี่ยนไปบริโภคน้ำมันสำเร็จรูปชนิดอื่นทดแทน เช่น มีการบริโภคน้ำมันดีเซลชนิดที่มีการผสมกับไบโอดีเซล (เชื้อเพลิงดีเซลทางเลือกที่ผลิตจาก น้ำมันพืช หรือ ไขมันสัตว์) เพิ่มขึ้น ทำให้ปริมาณการบริโภคน้ำมันดีเซลลดลง เป็นต้น ในส่วนของความสัมพันธ์ระหว่างปริมาณการบริโภคน้ำมันดีเซลกับผลิตรถยนต์รวมและจำนวนรถจดทะเบียนใหม่ซึ่งมีทิศทางเดียวกัน ผู้วิจัยมีความเห็นเช่นเดียวกัน ซึ่งคิดว่าเมื่อผลิตรถยนต์รวมและจำนวนรถใหม่ที่จะจดทะเบียนสูงขึ้น นั้นหมายถึงเศรษฐกิจดีขึ้น ผู้บริโภคมีกำลังซื้อเพิ่มขึ้น เกิดการใช้จ่ายใช้สอยและการเดินทางทั้งรถส่วนตัวและระบบขนส่งมวลชนเพิ่มมากขึ้น ปริมาณการบริโภคน้ำมันดีเซลจึงเพิ่มขึ้นด้วย สำหรับจำนวนประชากรที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันดีเซลในทิศทางตรงข้ามซึ่งผู้วิจัยไม่เห็นด้วย โดยผู้วิจัยเชื่อว่าเมื่อจำนวนประชากรเพิ่มมากขึ้นปริมาณการบริโภคน้ำมันดีเซลหรือน้ำมันสำเร็จรูปชนิดอื่นจะเพิ่มขึ้นด้วย เพราะมีจำนวนผู้บริโภคเพิ่มขึ้น และตัวแปรอิสระสุดท้ายที่ถูกเลือกคือปริมาณการผลิตน้ำมันเบนซิน ผู้วิจัยคิดว่าควรจะเป็นปริมาณการผลิตน้ำมันดีเซลที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันดีเซลมากกว่าปริมาณการผลิตน้ำมันเบนซิน เพราะหากมีการผลิตสูงขึ้นก็จะนำเข้าน้ำมันสำเร็จรูปลดลง ราคาของน้ำมันดีเซลอาจจะถูกลง ทำให้ผู้บริโภคมีกำลังซื้อน้ำมันดีเซลมากขึ้น เมื่อนำตัวแปรอิสระที่ถูกเลือกโดยเทคนิคการถดถอยแบบขั้นตอนมาเขียนสมการของตัวแบบการถดถอยพหุคูณของปริมาณการบริโภคน้ำมันดีเซลกับตัวแปรปัจจัยที่ส่งผลได้ดังนี้

$$\begin{aligned} \text{Diesel} &= 5273 + 0.9569 \text{ Gasoline Production} - 53.64 \text{ Diesel Price} \\ \text{Consumption} &+ 0.002085 \text{ ExpeGDP} + 0.001169 \text{ Total Vehicle} \\ &- 0.000095 \text{ Population} \end{aligned} \quad (4.1)$$

ตัวแบบการถดถอยพหุคูณที่ได้มีค่าสัมประสิทธิ์การตัดสินใจปรับแต่ง (Adjusted Coefficient of Determination,  $R^2(\text{adj})$ ) เท่ากับ 90.16% แสดงให้เห็นว่าชุดข้อมูลฝึกสอนมีความเข้ากันกับตัวแบบการถดถอยเชิงเส้นพหุคูณที่ได้เป็นอย่างดี นอกจากนี้ยังพบปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) ซึ่งเห็นได้จากค่า Variance Inflation Factor (VIF) ถ้าหากค่า VIF มากกว่า 10 แสดงว่าเกิดปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ (Sheather, 2009)

### Stepwise Selection of Terms

$\alpha$  to enter = 0.05,  $\alpha$  to remove = 0.05

### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	2123904	303415	132.76	0.000
Diesel Production	1	13944	13944	6.10	0.016
Gasoline Production	1	39254	39254	17.18	0.000
LPG Production	1	61667	61667	26.98	0.000
Diesel Price	1	59511	59511	26.04	0.000
ExpeGDP	1	20902	20902	9.15	0.004
Total Vehicle	1	23200	23200	10.15	0.002
Population	1	128416	128416	56.19	0.000
Error	60	137123	2285		
Total	67	2261027			

### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
47.8057	93.94%	93.23%	92.23%

### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-5172	785	-6.59	0.000	
Diesel Production	0.0665	0.0269	2.47	0.016	29.14
Gasoline Production	0.2010	0.0485	4.14	0.000	9.04
LPG Production	-0.3292	0.0634	-5.19	0.000	28.30
Diesel Price	-10.45	2.05	-5.10	0.000	7.25
ExpeGDP	0.000234	0.000077	3.02	0.004	13.65
Total Vehicle	0.000201	0.000063	3.19	0.002	3.55
Population	0.000104	0.000014	7.50	0.000	14.78

### Regression Equation

$$\begin{aligned} \text{Gasoline Consumption} = & -5172 + 0.0665 \text{ Diesel Production} + 0.2010 \text{ Gasoline Production} \\ & - 0.3292 \text{ LPG Production} - 10.45 \text{ Diesel Price} + 0.000234 \text{ ExpeGDP} \\ & + 0.000201 \text{ Total Vehicle} + 0.000104 \text{ Population} \end{aligned}$$

รูปที่ 46 ผลลัพธ์การวิเคราะห์การถดถอยของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง

จากรูปที่ 46 พบว่าผลลัพธ์จากการทำการถดถอยแบบขั้นตอนของข้อมูลข้อมูลน้ำมันเบนซิน ได้เลือกตัวแปรอิสระ ได้แก่ ปริมาณการผลิตน้ำมันดีเซล ปริมาณการผลิตน้ำมันเบนซิน ปริมาณการผลิตก๊าซปิโตรเลียมเหลว ราคาน้ำมันดีเซล ผลผลิตทั้งหมดรวม จำนวนรถใหม่ที่จดทะเบียน และจำนวนประชากร ซึ่งตัวแปรอิสระดังกล่าวมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันสำเร็จรูปที่ระดับนัยสำคัญ 0.05 ซึ่งผู้วิจัยเห็นด้วยว่าผลผลิตทั้งหมดรวม จำนวนรถยนต์จดทะเบียนใหม่ และจำนวนประชากรมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันเบนซินในทิศทางเดียวกัน เหตุผลเป็นเช่นเดียวกันกับที่กล่าวไว้ในส่วนของน้ำมันดีเซล สำหรับตัวแปรราคาน้ำมันดีเซลที่ถูกเลือกและมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันเบนซินในทิศทางตรงข้ามนั้น ผู้วิจัยไม่เห็นด้วยและคิดว่าราคาน้ำมันเบนซินควรถูกเลือกมากกว่า เพราะราคาน้ำมันเบนซินเมื่อสูงขึ้น กำลังซื้อของผู้บริโภคอาจ

ลดลง หรือผู้บริโภคลือกใช้แก๊สโซฮอล์ทดแทน ซึ่งเป็นน้ำมันเบนซินที่ผสมกับเอทานอล ทำให้ปริมาณการบริโภคน้ำมันเบนซินโดยรวมลดลง แต่ผู้วิจัยคิดว่ากรณีที่ตัวแบบได้เลือกราคาน้ำมันดีเซล อาจเป็นเพราะราคาน้ำมันดีเซลและราคาน้ำมันเบนซินอาจมีความสัมพันธ์กันในทิศทางเดียวกัน โดยจากข้อมูลที่ใช้ในการสร้างตัวแบบอาจทำให้ราคาน้ำมันดีเซลมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันเบนซินมากกว่าราคาน้ำมันเบนซิน และในส่วนของกลุ่มตัวแปรอิสระปริมาณการผลิตน้ำมันสำเร็จรูปที่ตัวแบบได้เลือกมา ผู้วิจัยเห็นด้วยกับปริมาณการผลิตน้ำมันเบนซิน ซึ่งผู้วิจัยคิดว่าหากมีการผลิตน้ำมันเบนซินสูงขึ้นก็จะนำเข้าน้ำมันเบนซินลดลง ราคาของน้ำมันเบนซินอาจจะถูกลง ทำให้ผู้บริโภคมักกำลังซื้อน้ำมันเบนซินมากขึ้นสำหรับปริมาณการผลิตน้ำมันดีเซลและก๊าซปิโตรเลียมเหลว นั้น ผู้วิจัยคิดว่าอาจส่งผลต่อปริมาณการบริโภคน้ำมันเบนซิน เนื่องจากในการกลั่นน้ำมันสำเร็จรูปจากน้ำมันดิบนั้น ได้ผลิตภัณฑ์หลายชนิด เมื่อผลิตน้ำมันเบนซินเพิ่มขึ้น จะได้น้ำมันชนิดอื่นเพิ่มขึ้นด้วย ดังนั้นเมื่อนำชุดข้อมูลฝึกสอนมาสร้างตัวแบบด้วยเทคนิคการถดถอยแบบขั้นตอน ข้อมูลปริมาณการผลิตน้ำมันสำเร็จรูปชนิดอื่น ๆ จึงอาจมีอิทธิพลต่อปริมาณการบริโภคน้ำมันเบนซิน เมื่อเขียนสมการของตัวแบบการถดถอยพหุคูณของปริมาณการบริโภคน้ำมันเบนซินกับตัวแปรปัจจัยที่ส่งผลได้สมการดังนี้

$$\begin{aligned}
 \text{Gasoline} &= -5172 + 0.0665 \text{ Diesel Production} \\
 \text{Consumption} &+ 0.2010 \text{ Gasoline Production} \\
 &- 0.3292 \text{ LPG Production} - 10.45 \text{ Diesel Price} \\
 &+ 0.000234 \text{ ExpeGDP} \\
 &+ 0.000201 \text{ Total Vehicle} + 0.000104 \text{ Population} \quad (4.2)
 \end{aligned}$$

ตัวแบบการถดถอยพหุคูณที่ได้มีค่าสัมประสิทธิ์การตัดสินใจปรับแต่ง  $R^2(\text{adj})$  เท่ากับ 93.23% แสดงให้เห็นว่าชุดข้อมูลฝึกสอนมีความเข้ากันกับตัวแบบการถดถอยเชิงเส้นพหุคูณที่ได้เป็นอย่างดี และยังพบปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ



### Stepwise Selection of Terms

$\alpha$  to enter = 0.05,  $\alpha$  to remove = 0.05

### Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1407255	1407255	615.33	0.000
ExpeGDP	1	1407255	1407255	615.33	0.000
Error	66	150943	2287		
Total	67	1558197			

### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
47.8227	90.31%	90.17%	89.52%

### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	181.0	31.6	5.73	0.000	
ExpeGDP	0.000520	0.000021	24.81	0.000	1.00

### Regression Equation

$$JP \text{ Consumption} = 181.0 + 0.000520 \text{ ExpeGDP}$$

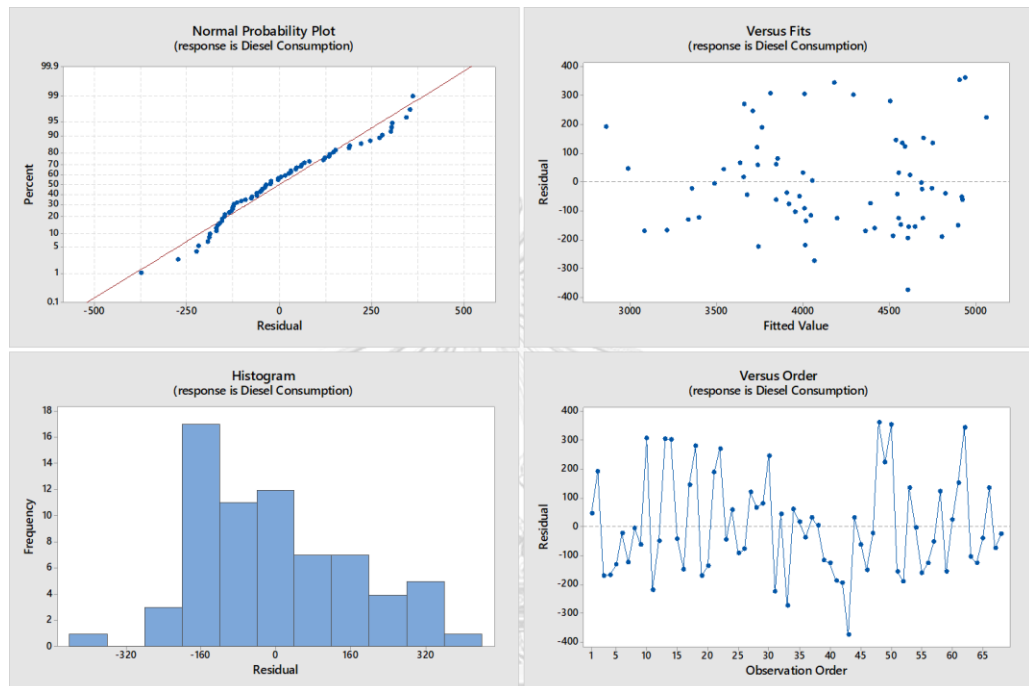
รูปที่ 47 ผลลัพธ์การวิเคราะห์การถดถอยของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะกลาง

จากรูปที่ 47 พบว่าผลลัพธ์จากการทำการถดถอยแบบขั้นตอนของข้อมูลน้ำมันเครื่องบินได้เลือกตัวแปรอิสระที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันสำเร็จรูปที่ระดับนัยสำคัญ 0.05 เพียงตัวแปรเดียวคือ ผลิตภัณฑ์มวลรวม ซึ่งผู้วิจัยเห็นด้วยเพราะการที่ผลิตภัณฑ์มวลรวมสูงขึ้น แสดงถึงผู้บริโภคมีการใช้จ่ายใช้สอยมากขึ้น กำลังซื้อเพิ่มขึ้น การเดินทางด้วยเครื่องบินอาจเพิ่มขึ้น ส่งผลทำให้การบริโภคน้ำมันเครื่องบินอาจเพิ่มขึ้นด้วยเช่นกันซึ่งสมการของตัวแบบการถดถอยเชิงเส้นของปริมาณการบริโภคน้ำมันเครื่องบินกับตัวแปรปัจจัยที่ส่งผลเป็นดังนี้

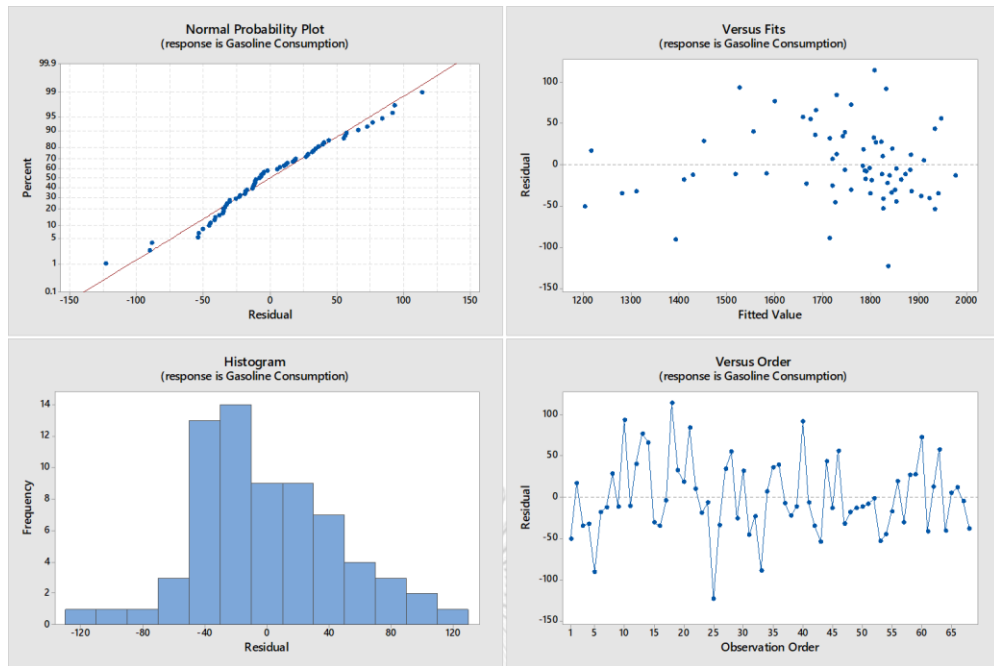
$$JP \text{ Consumption} = 181.0 + 0.000520 \text{ ExpeGDP} \quad (4.3)$$

ตัวแบบการถดถอยพหุคูณที่ได้มีค่าสัมประสิทธิ์การตัดสินใจปรับแต่ง  $R^2(\text{adj})$  เท่ากับ 90.17% แสดงให้เห็นว่าชุดข้อมูลฝึกสอนมีความเข้ากันกับตัวแบบการถดถอยเชิงเส้นพหุคูณที่ได้เป็นอย่างดี

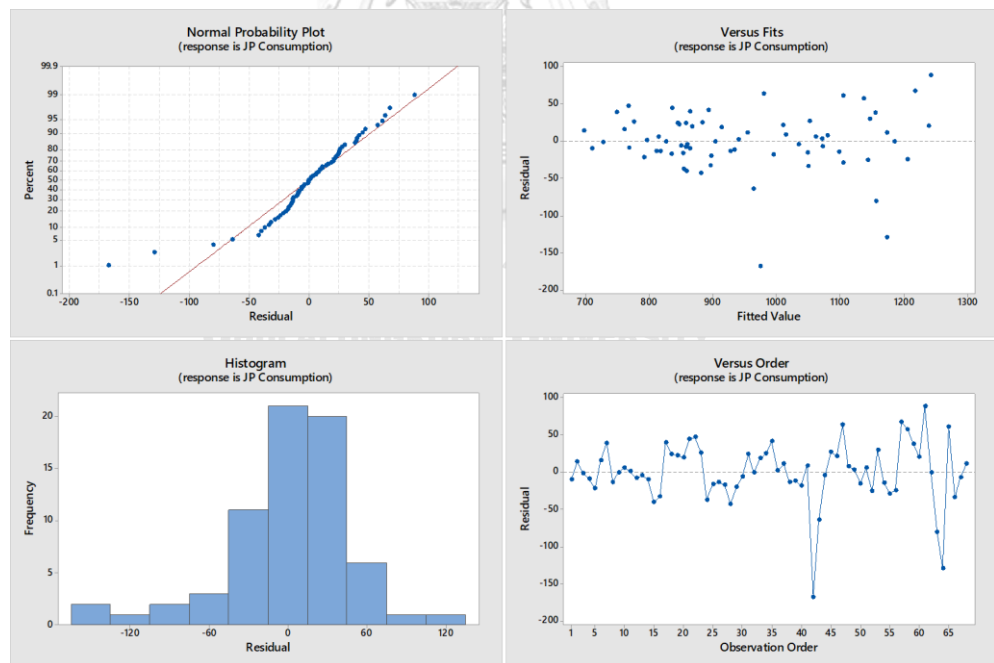
หลังจากได้ตัวแบบการถดถอยแล้วต่อไปเป็นการทดสอบสมมติฐานของการถดถอยเชิงเส้นกับตัวแบบที่สร้างขึ้น ซึ่งจะทดสอบด้วยการนำค่าความผิดพลาด (Residual) ไปสร้างกราฟดังแสดงในรูปที่ 48 49 และ 50 สรุปได้ว่าตัวแบบทั้ง 3 ตัวแบบสอดคล้องกับสมมติฐานของการถดถอย แต่ตัวแบบของน้ำมันเครื่องบินตรวจพบค่าผิดปกติ (Outlier) ที่อาจส่งผลกระทบต่อผลลัพธ์การพยากรณ์ได้



รูปที่ 48 ผลลัพธ์การทดสอบสมมติฐานการถดถอยของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง

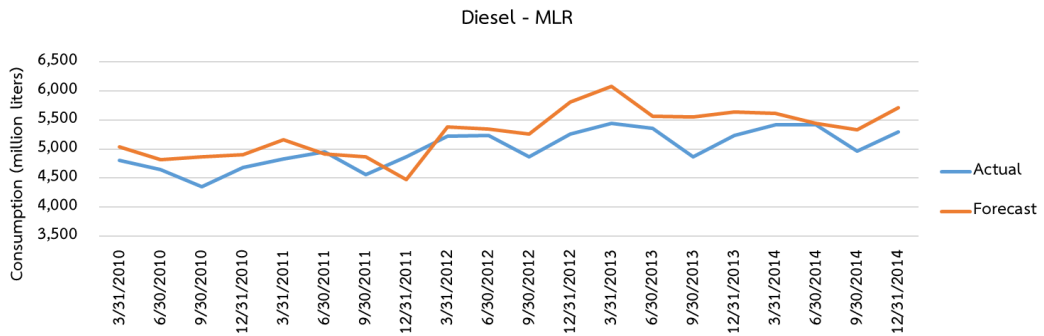


รูปที่ 49 ผลลัพธ์การทดสอบสมมติฐานการถดถอยของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง

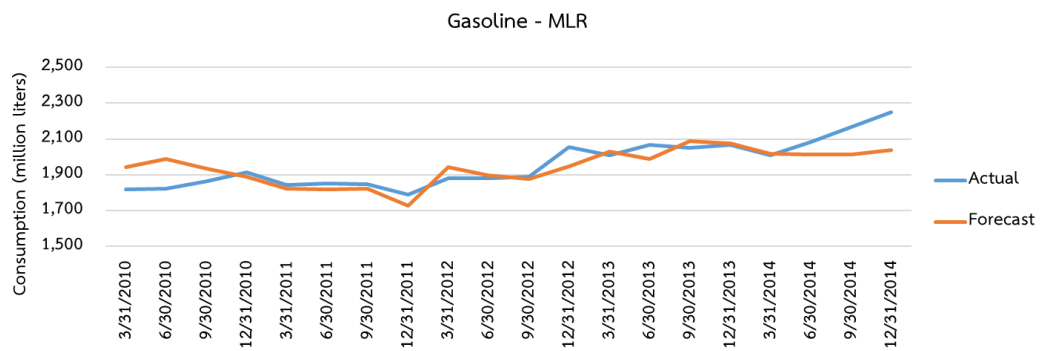


รูปที่ 50 ผลลัพธ์การทดสอบสมมติฐานการถดถอยของน้ำมันเครื่องบิน  
สำหรับการพยากรณ์ระยะกลาง

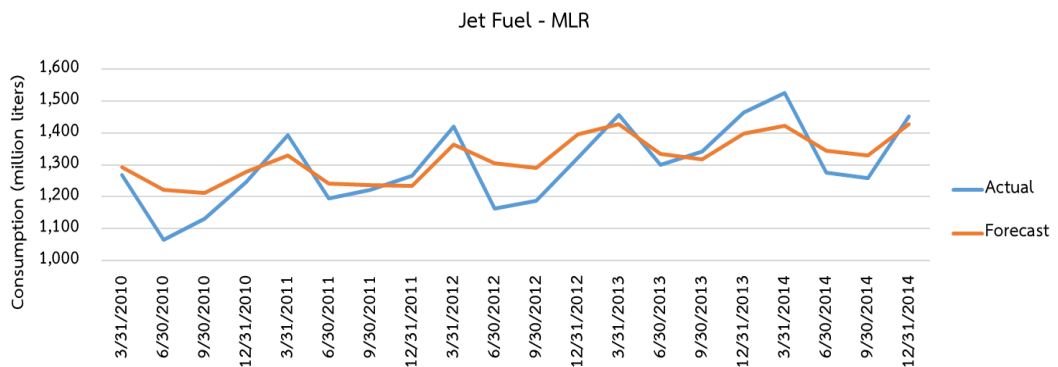
เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 51 52 และ 53



รูปที่ 51 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง



รูปที่ 52 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง



รูปที่ 53 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะกลาง

## 2) การพยากรณ์ระยะสั้น

ผลลัพธ์ของตัวแบบการถดถอยเชิงเส้นพหุคูณของน้ำมันดีเซลสำหรับการพยากรณ์ระยะสั้น ซึ่งใช้การเลื่อนชุดข้อมูลฝึกสอนไปข้างหน้าในการสร้างตัวแบบครั้งละ 4 ไตรมาส ได้ตัวแบบซึ่งเขียนเป็นสมการได้ดังนี้

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)

$$\begin{aligned} \text{Diesel} &= 5273 + 0.9569 \text{ Gasoline Production} - 53.64 \text{ Diesel Price} \\ \text{Consumption1} &+ 0.002085 \text{ ExpeGDP} + 0.001169 \text{ Total Vehicle} \\ &- 0.000095 \text{ Population} \end{aligned} \quad (4.4)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)

$$\begin{aligned} \text{Diesel} &= 6161 + 0.951 \text{ Gasoline Production} - 51.51 \text{ Diesel Price} \\ \text{Consumption2} &+ 0.001970 \text{ ExpeGDP} \\ &+ 0.001153 \text{ Total Vehicle} - 0.000107 \text{ Population} \end{aligned} \quad (4.5)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)

$$\begin{aligned} \text{Diesel} &= 12869 + 0.3896 \text{ Diesel Production} - 34.51 \text{ Diesel Price} \\ \text{Consumption3} &- 18.64 \text{ Gasoline Price} + 0.001904 \text{ ExpeGDP} \\ &+ 0.000775 \text{ Total Vehicle} \\ &- 0.000205 \text{ Population} \end{aligned} \quad (4.6)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)

$$\begin{aligned} \text{Diesel} &= 14543 + 0.3045 \text{ Diesel Production} - 34.92 \text{ Diesel Price} \\ \text{Consumption4} &- 14.82 \text{ Gasoline Price} + 0.001881 \text{ ExpeGDP} \\ &+ 0.000958 \text{ Total Vehicle} \\ &- 0.000227 \text{ Population} \end{aligned} \quad (4.7)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)

$$\begin{aligned} \text{Diesel} &= 18510 + 0.2024 \text{ Diesel Production} - 45.94 \text{ Diesel Price} \\ \text{Consumption5} &+ 0.001828 \text{ ExpeGDP} \\ &+ 0.001189 \text{ Total Vehicle} - 0.000286 \text{ Population} \end{aligned} \quad (4.8)$$

เนื่องจากชุดข้อมูลฝึกสอนสำหรับสมการการถดถอยเชิงเส้นพหุคูณทั้ง 5 ช่วงมีความแตกต่างกัน ตัวแบบจึงได้เลือกตัวแปรอิสระที่ส่งผลต่อปริมาณการบริโภคน้ำมันสำเร็จรูปแตกต่างกันบ้าง โดยในสมการที่ 4.5 ถึงสมการที่ 4.8 ตัวแบบได้เลือกปริมาณการผลิตน้ำมันดีเซลแทนปริมาณการผลิตน้ำมันเบนซิน ซึ่งผู้เขียนเห็นด้วย เพราะมีความเห็นว่าหากมีการผลิตสูงขึ้นก็จะนำเข้าน้ำมันสำเร็จรูป ลดลง ราคาของน้ำมันดีเซลอาจจะถูกลง ทำให้ผู้บริโภคมีกำลังซื้อน้ำมันดีเซลมากขึ้น ส่วนตัวแปรอิสระตัวอื่น ๆ ที่ถูกเลือกเข้ามาในสมการนั้นได้ตัวแปรอิสระเหมือนกันทุกสมการ จะแตกต่างกันที่ค่าคงที่ของแต่ละตัวแปรซึ่งเกิดจากความแตกต่างกันของชุดข้อมูลฝึกสอน

ต่อไปเป็นผลลัพธ์ของตัวแบบการถดถอยเชิงเส้นพหุคูณของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้น ได้ตัวแบบซึ่งเขียนเป็นสมการได้ดังนี้

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)

$$\begin{aligned} \text{Gasoline} &= -5172 + 0.0665 \text{ Diesel Production} \\ \text{Consumption1} &+ 0.2010 \text{ Gasoline Production} \\ &- 0.3292 \text{ LPG Production} - 10.45 \text{ Diesel Price} \\ &+ 0.000234 \text{ ExpeGDP} \\ &+ 0.000201 \text{ Total Vehicle} + 0.000104 \text{ Population} \end{aligned} \quad (4.9)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)

$$\begin{aligned} \text{Gasoline} &= -3792 + 0.2696 \text{ Gasoline Production} \\ \text{Consumption2} &- 0.1301 \text{ LPG Production} \\ &- 11.77 \text{ Diesel Price} + 0.000215 \text{ ExpeGDP} \\ &+ 0.000175 \text{ Total Vehicle} \\ &+ 0.000080 \text{ Population} \end{aligned} \quad (4.10)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)

$$\begin{aligned} \text{Gasoline Consumption}_3 &= 1272 + 0.1024 \text{ Diesel Production} - 11.46 \text{ Diesel Price} \\ &+ 16.29 \text{ LPG Price} \end{aligned} \quad (4.11)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)

$$\begin{aligned} \text{Gasoline Consumption}_4 &= 1251 + 0.0747 \text{ Diesel Production} - 13.57 \text{ Diesel Price} \\ &+ 13.36 \text{ LPG Price} + 0.000150 \text{ ExpeGDP} \end{aligned} \quad (4.12)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)

$$\begin{aligned} \text{Gasoline Consumption}_5 &= 1187 + 0.0912 \text{ Diesel Production} - 14.15 \text{ Diesel Price} \\ &+ 0.000277 \text{ ExpeGDP} \end{aligned} \quad (4.13)$$

ตัวแบบการถดถอยเชิงเส้นพหุคูณของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้นในแต่ละช่วงได้เลือกตัวอิสระแตกต่างกัน โดยมีตัวแบบในช่วงที่ 4 ไม่เลือกผลิตภัณฑ์มวลรวม ซึ่งผู้วิจัยไม่เห็นด้วย เนื่องจากเห็นว่าผลิตภัณฑ์มวลรวมเป็นตัวแปรอิสระที่ความสัมพันธ์กับปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลัก ส่วนสาเหตุของความแตกต่างของตัวแปรอิสระอื่น ๆ ที่ถูกเลือกจากการถดถอยแบบขั้นตอนของตัวแบบในแต่ละช่วง ผู้วิจัยมีความเห็นเช่นเดียวกับที่ได้กล่าวไว้ในตัวแบบก่อนหน้านี้

สำหรับผลลัพธ์ของตัวแบบการถดถอยเชิงเส้นพหุคูณของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะสั้น ได้ตัวแบบซึ่งเขียนเป็นสมการได้ดังนี้

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)

$$\text{JP Consumption}_1 = 181.0 + 0.000520 \text{ ExpeGDP} \quad (4.14)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)

$$\begin{aligned}
 \text{JP} &= 2340 + 0.1746 \text{ LPG Production} + 6.05 \text{ Diesel Price} \\
 \text{Consumption2} &- 8.90 \text{ Gasoline Price} \\
 &+ 0.000618 \text{ ExpeGDP} - 0.000151 \text{ Total Vehicle} \\
 &- 0.000039 \text{ Population}
 \end{aligned} \tag{4.15}$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)

$$\begin{aligned}
 \text{JP} &= 2809 + 0.2231 \text{ LPG Production} - 5.48 \text{ Gasoline Price} \\
 \text{Consumption3} &+ 0.000631 \text{ ExpeGDP} \\
 &- 0.000147 \text{ Total Vehicle} - 0.000048 \text{ Population}
 \end{aligned} \tag{4.16}$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)

$$\begin{aligned}
 \text{JP} &= 111.8 + 6.67 \text{ Diesel Price} - 7.65 \text{ Gasoline Price} \\
 \text{Consumption4} &+ 0.000595 \text{ ExpeGDP}
 \end{aligned} \tag{4.17}$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)

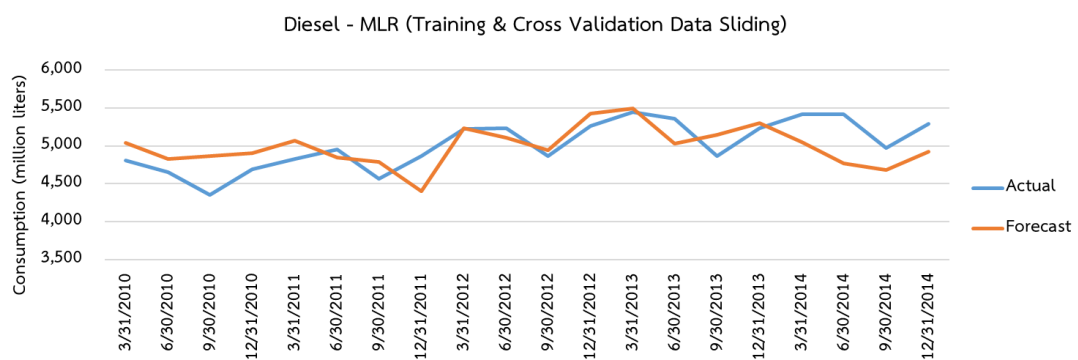
$$\begin{aligned}
 \text{JP} &= 2520 + 0.1014 \text{ LPG Production} + 0.1401 \text{ JP Production} \\
 \text{Consumption5} &- 6.32 \text{ Gasoline Price} \\
 &+ 0.000615 \text{ ExpeGDP} - 0.000043 \text{ Population}
 \end{aligned} \tag{4.18}$$

ตัวแบบการถดถอยเชิงเส้นพหุคูณของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้นได้เลือกตัวแปรอิสระที่แตกต่างกันในแต่ละช่วงเช่นเดียวกันกับน้ำมันสำเร็จรูปชนิดอื่น โดยผลิตภัณฑ์มวลรวมถูกเลือกทุกช่วง ซึ่งผู้วิจัยเห็นด้วย เพราะการที่ผลิตภัณฑ์มวลรวมสูงขึ้น แสดงถึงผู้บริโภคมีการจับจ่ายใช้สอยมากขึ้น กำลังซื้อเพิ่มขึ้น การเดินทางด้วยเครื่องบินอาจเพิ่มขึ้น ส่งผลทำให้การบริโภคน้ำมันเครื่องบินอาจเพิ่มขึ้น ในส่วนของตัวแปรกลุ่มปริมาณการผลิตก๊าซปิโตรเลียมเหลวและราคาน้ำมันเบนซินได้ถูกเลือกในบางช่วง ซึ่งผู้วิจัยไม่เห็นด้วย เพราะคิดว่าก๊าซปิโตรเลียมและน้ำมันเบนซินเป็น

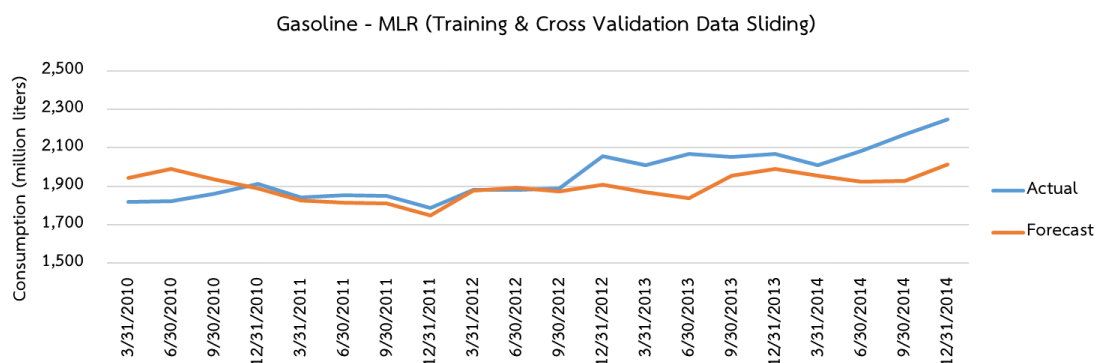


เชื้อเพลิงสำหรับการคมนาคมทางบก ส่วนน้ำมันเครื่องบินเป็นเชื้อเพลิงสำหรับการคมนาคมทางอากาศ ไม่น่าจะมีความเกี่ยวข้องกัน สำหรับจำนวนประชากรได้ถูกเลือกในบางช่วง แต่กลับมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันเครื่องบินในทิศทางตรงข้ามกัน ซึ่งผู้วิจัยไม่เห็นด้วยและคิดว่าจำนวนประชากรที่เพิ่มขึ้น ประชากรน่าจะมีโอกาสที่จะเดินทางด้วยเครื่องบินเพิ่มขึ้น ส่งผลให้ปริมาณการบริโภคน้ำมันเครื่องบินเพิ่มสูงขึ้น

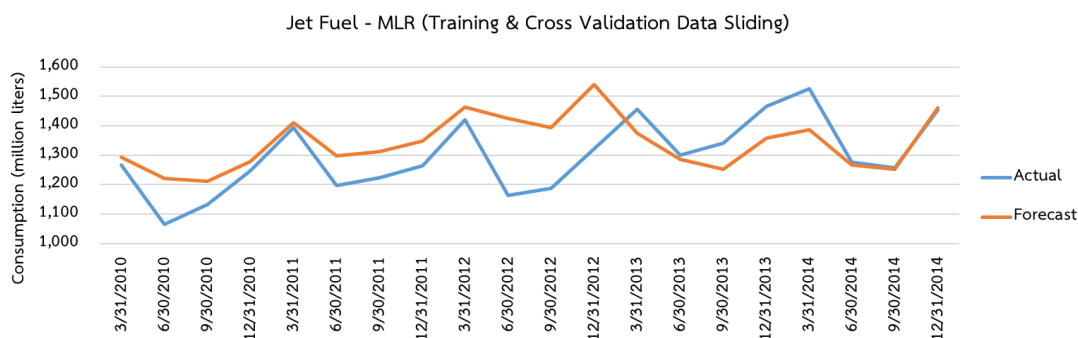
เมื่อนำตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับการพยากรณ์ระยะสั้นทั้งหมดไปทดสอบสมมติฐานของการถดถอยเชิงเส้นกับตัวแบบที่สร้างขึ้น สรุปได้ว่าตัวแบบทั้งหมดสอดคล้องกับสมมติฐานของการถดถอย จากนั้นทำการพยากรณ์ไปล่วงหน้า 4 ไตรมาสหรือจำนวน 1 ปี โดยมีการเลื่อน (Sliding) การพยากรณ์ไปครั้งละ 4 ไตรมาส จำนวน 5 ครั้ง จนครบ 20 ไตรมาส ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 54 55 และ 56



รูปที่ 54 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะสั้น



รูปที่ 55 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้น



รูปที่ 56 การพยากรณ์ด้วยตัวแบบ MLR ของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะสั้น

#### 4.4 ตัวแบบ Random Sample Consensus Regression

หลังจากที่ได้ตัวแปรอิสระที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันสำเร็จรูปทั้ง 3 ชนิด ผู้วิจัยได้นำตัวแปรอิสระดังกล่าวไปใช้กับตัวแบบ Random Sample Consensus Regression ซึ่งจากนี้ไปจะเรียกว่าตัวแบบ RANSAC-REG โดยมีผลลัพธ์ของการพยากรณ์ดังนี้

##### 1) การพยากรณ์ระยะกลาง

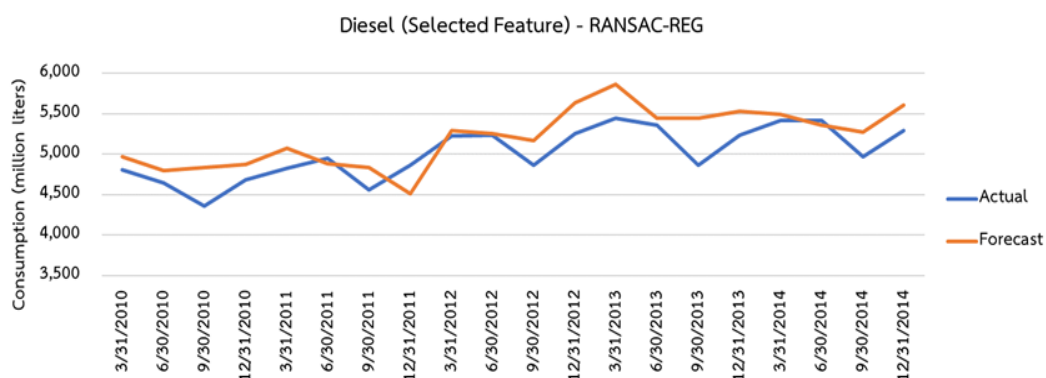
ผลลัพธ์ของการสร้างตัวแบบ RANSAC-REG สำหรับการพยากรณ์ระยะกลางจากชุดข้อมูลฝึกสอนของปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน มาจากการกำหนดค่า random\_state ก่อนการสร้างตัวแบบ เมื่อสร้างตัวแบบแล้ววัดผลลัพธ์ด้วยค่า MAPE กับชุดข้อมูลตรวจสอบพบว่า random\_state ที่ให้ค่า MAPE ต่ำที่สุดสำหรับการพยากรณ์ระยะกลางของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินคือ 123 12345 และ 7891011 ตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 57 58 และ 59 และได้สมการของ RANSAC-REG ของการพยากรณ์น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ดังสมการที่ 4.19 4.20 และ 4.21 ตามลำดับ

$$\begin{aligned}
 \text{Diesel Consumption} &= 3368 + 0.8508 \text{ Gasoline Production} - 39.23 \text{ Diesel Price} \\
 &+ 0.001750 \text{ ExpeGDP} + 0.000927 \text{ Total Vehicle} \\
 &- 0.000055 \text{ Population}
 \end{aligned} \tag{4.19}$$

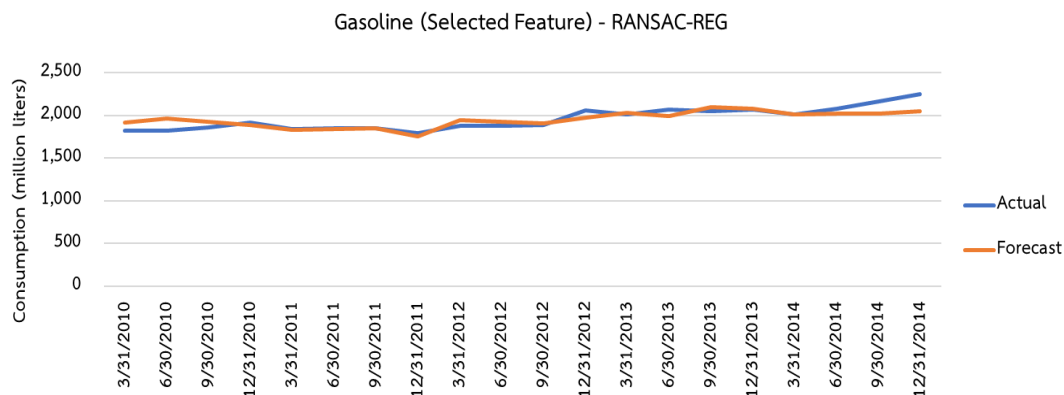
$$\begin{aligned}
 \text{Gasoline Consumption} &= -5250 + 0.07040 \text{ Diesel Production} \\
 &+ 0.1788 \text{ Gasoline Production} \\
 &- 0.2452 \text{ LPG Production} - 13.83 \text{ Diesel Price} \\
 &+ 0.000175 \text{ ExpeGDP} \\
 &+ 0.000195 \text{ Total Vehicle} + 0.000106 \text{ Population} \quad (4.20)
 \end{aligned}$$

$$\text{JP Consumption} = 217.6 + 0.000495 \text{ ExpeGDP} \quad (4.21)$$

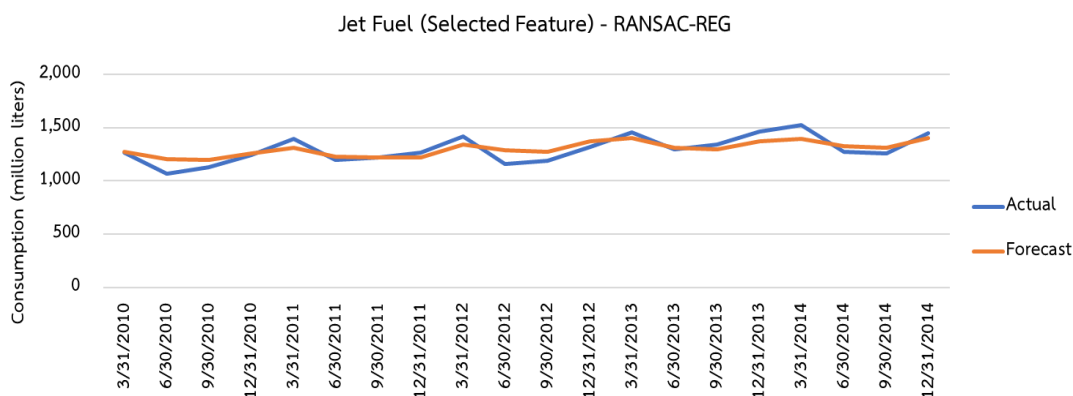
เมื่อพิจารณาสมการที่ 4.19 4.20 และ 4.21 ซึ่งเป็นสมการของตัวแบบ RANSAC-REG ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินตามลำดับ เปรียบเทียบกับสมการที่ 4.1 4.2 และ 4.3 ซึ่งเป็นสมการของตัวแบบ MLR ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินตามลำดับ พบว่าแต่ละตัวแปรอิสระมีทิศทางของสัมประสิทธิ์เหมือนกันทั้งหมด แต่มีขนาดของสัมประสิทธิ์ที่แตกต่างกัน ซึ่งผู้วิจัยคิดว่าขนาดของสัมประสิทธิ์ที่มีค่าแตกต่างกันน่าจะเป็นเพราะตัวแบบ RANSAC-REG จะไม่นำข้อมูลป้อนเข้าที่มีค่าผิดปกติ (Outlier) มาใช้ในการประมาณค่าสัมประสิทธิ์ของตัวแบบ และผู้วิจัยมีความเห็นด้วยและไม่เห็นกับทิศทางของสัมประสิทธิ์ของแต่ละตัวแปรอิสระที่ได้จากตัวแบบ RANSAC-REG เช่นเดียวกันกับตัวแบบ MLR เนื่องจากมีทิศทางของสัมประสิทธิ์ของตัวแปรอิสระเป็นทิศทางเดียวกัน



รูปที่ 57 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง



รูปที่ 58 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง



รูปที่ 59 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันเครื่องบิน  
สำหรับการพยากรณ์ระยะกลาง

## 2) การพยากรณ์ระยะสั้น

ผลลัพธ์ของการสร้างตัวแบบ RANSAC-REG สำหรับการพยากรณ์ระยะสั้นจากชุดข้อมูลฝึกสอนของปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน มาจากการกำหนดค่า `random_state` ก่อนการสร้างตัวแบบ เมื่อสร้างตัวแบบแล้ววัดผลลัพธ์ด้วยค่า MAPE กับชุดข้อมูลตรวจสอบพบว่า `random_state` ที่ให้ค่า MAPE ต่ำที่สุดสำหรับการพยากรณ์ระยะสั้นของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินคือ 123 123 และ 7891011 ตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 4 ไตรมาสหรือจำนวน 1 ปี โดยมีการเลื่อน (Sliding) การพยากรณ์ไปครั้งละ 4 ไตรมาสจำนวน 5 ครั้ง จนครบ 20 ไตรมาส ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 60 61 และ 62 โดยแต่ละตัวแบบ RANSAC-REG สำหรับการพยากรณ์น้ำมันดีเซลแบบระยะสั้นได้สมการดังนี้

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)

$$\begin{aligned}
 \text{Diesel} &= 3368 + 0.8508 \text{ Gasoline Production} \\
 \text{Consumption1} &- 39.23 \text{ Diesel Price} \\
 &+ 0.001750 \text{ ExpeGDP} + 0.000927 \text{ Total Vehicle} \\
 &- 0.000055 \text{ Population}
 \end{aligned} \tag{4.22}$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)

$$\begin{aligned}
 \text{Diesel} &= 3688 + 0.8419 \text{ Gasoline Production} - 37.53 \text{ Diesel Price} \\
 \text{Consumption2} &+ 0.001637 \text{ ExpeGDP} \\
 &+ 0.000962 \text{ Total Vehicle} - 0.000058 \text{ Population}
 \end{aligned} \tag{4.23}$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)

$$\begin{aligned}
 \text{Diesel} &= 9808 + 0.3685 \text{ Diesel Production} - 58.32 \text{ Diesel Price} \\
 \text{Consumption3} &- 9.690 \text{ Gasoline Price} + 0.002027 \text{ ExpeGDP} \\
 &+ 0.000639 \text{ Total Vehicle} \\
 &- 0.000150 \text{ Population}
 \end{aligned} \tag{4.24}$$


สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)

$$\begin{aligned}
 \text{Diesel} &= 14215 + 0.2823 \text{ Diesel Production} - 33.92 \text{ Diesel Price} \\
 \text{Consumption4} &- 16.14 \text{ Gasoline Price} + 0.001996 \text{ ExpeGDP} \\
 &+ 0.000873 \text{ Total Vehicle} \\
 &- 0.000220 \text{ Population}
 \end{aligned} \tag{4.25}$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)


$$\begin{aligned} \text{Diesel Consumption5} &= 16300 + 0.3045 \text{ Diesel Production} - 50.12 \text{ Diesel Price} \\ &+ 0.001670 \text{ ExpeGDP} \\ &+ 0.001095 \text{ Total Vehicle} - 0.000250 \text{ Population} \end{aligned} \quad (4.26)$$

สำหรับการพยากรณ์ระยะสั้นด้วยตัวแบบ RANSAC-REG น้ำมันเบนซินได้สมการดังนี้  
สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)



$$\begin{aligned} \text{Gasoline Consumption1} &= -4445 + 0.0739 \text{ Diesel Production} \\ &+ 0.2082 \text{ Gasoline Production} \\ &- 0.2462 \text{ LPG Production} - 9.82 \text{ Diesel Price} \\ &+ 0.000214 \text{ ExpeGDP} \\ &+ 0.000021 \text{ Total Vehicle} + 0.000092 \text{ Population} \end{aligned} \quad (4.27)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)



$$\begin{aligned} \text{Gasoline Consumption2} &= -3740 + 0.2332 \text{ Gasoline Production} \\ &- 0.1170 \text{ LPG Production} \\ &- 10.57 \text{ Diesel Price} + 0.000277 \text{ ExpeGDP} \\ &+ 0.000045 \text{ Total Vehicle} \\ &+ 0.000080 \text{ Population} \end{aligned} \quad (4.28)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)

$$\begin{aligned} \text{Gasoline Consumption3} &= 1333 + 0.0461 \text{ Diesel Production} - 11.06 \text{ Diesel Price} \\ &+ 29.36 \text{ LPG Price} \end{aligned} \quad (4.29)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)

$$\begin{aligned} \text{Gasoline Consumption}_4 &= 1347 + 0.0684 \text{ Diesel Production} - 10.34 \text{ Diesel Price} \\ &+ 11.36 \text{ LPG Price} + 0.000089 \text{ ExpeGDP} \end{aligned} \quad (4.30)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)

$$\begin{aligned} \text{Gasoline Consumption}_5 &= 1001 + 0.1309 \text{ Diesel Production} - 21.63 \text{ Diesel Price} \\ &+ 0.000355 \text{ ExpeGDP} \end{aligned} \quad (4.31)$$

สำหรับการพยากรณ์ระยะสั้นด้วยตัวแบบ RANSAC-REG น้ำมันเครื่องบินได้สมการดังนี้  
สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)

$$\text{JP Consumption}_1 = 217.6 + 0.000495 \text{ ExpeGDP} \quad (4.32)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)

$$\begin{aligned} \text{JP Consumption}_2 &= 859.6 + 0.0882 \text{ LPG Production} + 5.46 \text{ Diesel Price} \\ &- 8.28 \text{ Gasoline Price} \\ &+ 0.000626 \text{ ExpeGDP} - 0.000098 \text{ Total Vehicle} \\ &- 0.000013 \text{ Population} \end{aligned} \quad (4.33)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)

$$\begin{aligned} \text{JP Consumption}_3 &= 2724 + 0.2334 \text{ LPG Production} - 5.15 \text{ Gasoline Price} \\ &+ 0.000551 \text{ ExpeGDP} \\ &- 0.000140 \text{ Total Vehicle} - 0.000045 \text{ Population} \end{aligned} \quad (4.34)$$

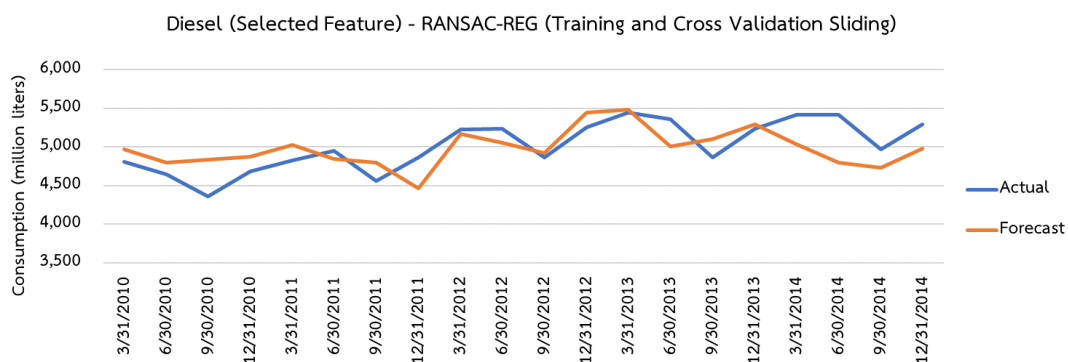
สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)

$$\begin{aligned} \text{JP} &= 119.5 + 8.58 \text{ Diesel Price} - 9.92 \text{ Gasoline Price} \\ \text{Consumption4} &+ 0.000601 \text{ ExpeGDP} \end{aligned} \quad (4.35)$$

สมการสำหรับชุดข้อมูลฝึกสอนช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)

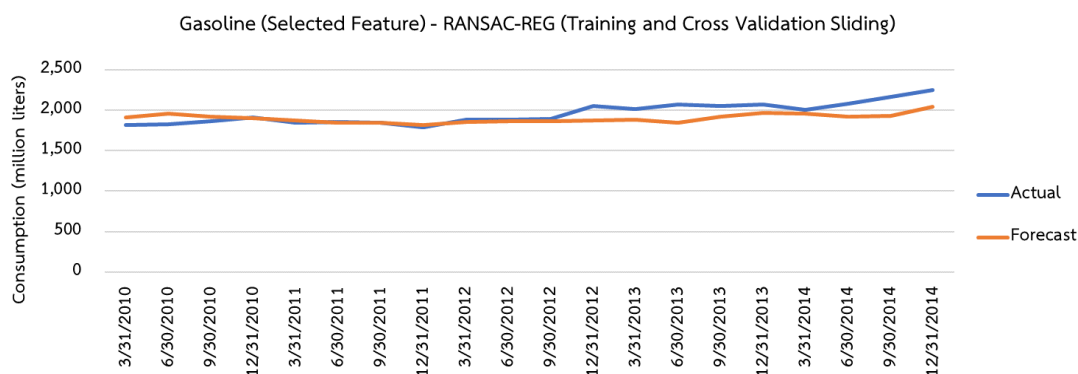
$$\begin{aligned} \text{JP} &= 2084 + 0.1220 \text{ LPG Production} + 0.0590 \text{ JP Production} \\ \text{Consumption5} &- 6.04 \text{ Gasoline Price} \\ &+ 0.000606 \text{ ExpeGDP} - 0.000035 \text{ Population} \end{aligned} \quad (4.36)$$

เมื่อพิจารณาสมการการพยากรณ์ระยะสั้นของตัวแบบ RANSAC-REG เปรียบเทียบกับสมการการพยากรณ์ระยะสั้นของตัวแบบ MLR ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินพบว่าแต่ละตัวแปรอิสระมีทิศทางของสัมประสิทธิ์เหมือนกันทั้งหมด แต่มีขนาดของสัมประสิทธิ์ที่แตกต่างกัน ซึ่งผู้วิจัยมีความเห็นเช่นเดียวกันกับดังที่ได้กล่าวถึงในการพยากรณ์ RANSAC-REG แบบระยะกลาง

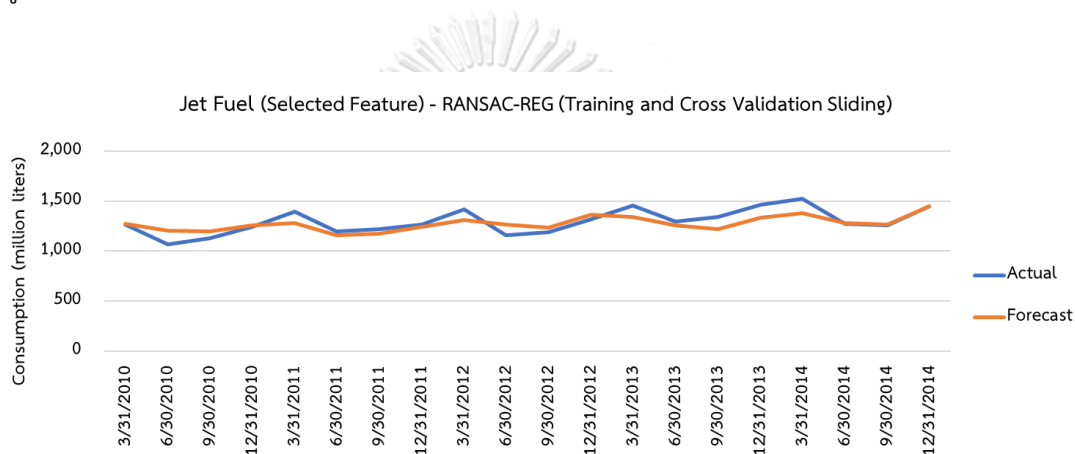


รูปที่ 60 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะสั้น





รูปที่ 61 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้น



รูปที่ 62 การพยากรณ์ด้วยตัวแบบ RANSAC-REG ของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะสั้น

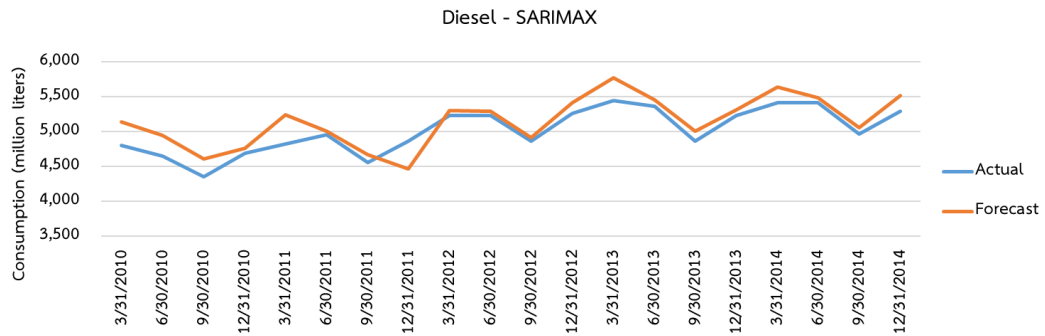
#### 4.5 ตัวแบบ SARIMA กับตัวแปรอิสระภายนอก (SARIMAX)

หลังจากที่ได้ตัวแปรอิสระที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันสำเร็จรูปทั้ง 3 ชนิด ผู้วิจัยได้นำตัวแปรอิสระดังกล่าวไปใช้กับตัวแบบบอซ-เจนกินส์ที่มีตัวแปรอิสระภายนอก ซึ่งจากนี้ไป จะเรียกว่าตัวแบบ SARIMAX โดยวิจัยเลือกใช้ Python Library: pmdarima.arima เพื่อหาตัวแบบ SARIMAX และพยากรณ์น้ำมันสำเร็จรูปทั้ง 3 ชนิดในรูปแบบการพยากรณ์ระยะสั้นและระยะยาว

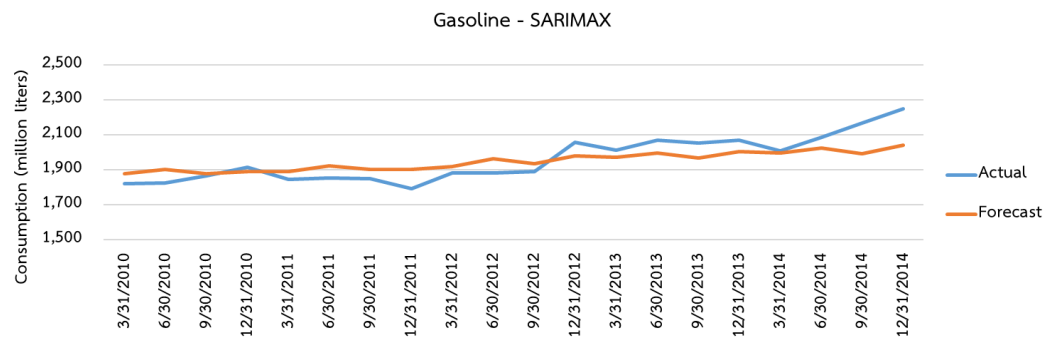
##### 1) การพยากรณ์ระยะกลาง

ผลลัพธ์ของการสร้างตัวแบบ SARIMAX สำหรับการพยากรณ์ระยะกลางจากชุดข้อมูล ฝึกสอนของปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน จะถูกเลือกจากค่าที่ทำให้ Akaike Information Criterion (AIC) มีค่าต่ำที่สุดเช่นเดียวกับตัวแบบ SARIMA โดยผลลัพธ์ของ น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน สามารถเขียนเป็นตัวแบบอย่างง่ายได้ดังนี้ SARIMAX(0,1,0)(0,1,1)<sub>4</sub> SARIMAX(0,1,1)(0,1,1)<sub>4</sub> และ SARIMAX(0,1,0)(0,1,1)<sub>4</sub> ตามลำดับ เมื่อทำ

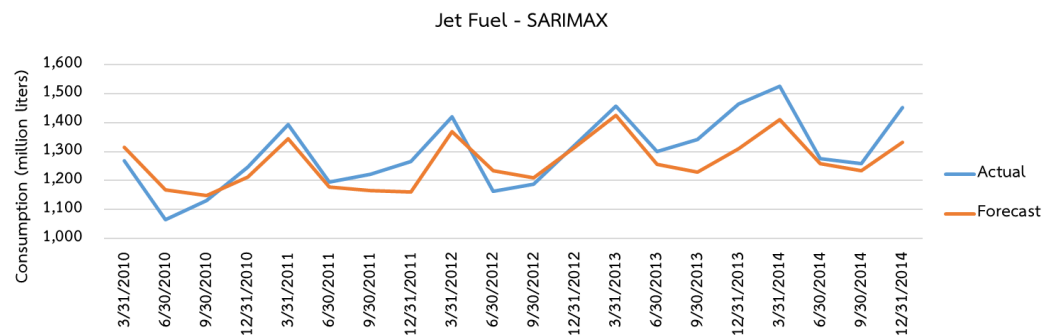
การตรวจสอบรูปแบบของตัวแบบ SARIMAX เช่นเดียวกับกับตัวแบบ SARIMA ซึ่งผลลัพธ์ที่ได้พบว่ามี ความคล้อยกับเงื่อนไขดังกล่าว จากนั้นทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 63 64 และ 65



รูปที่ 63 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะกลาง



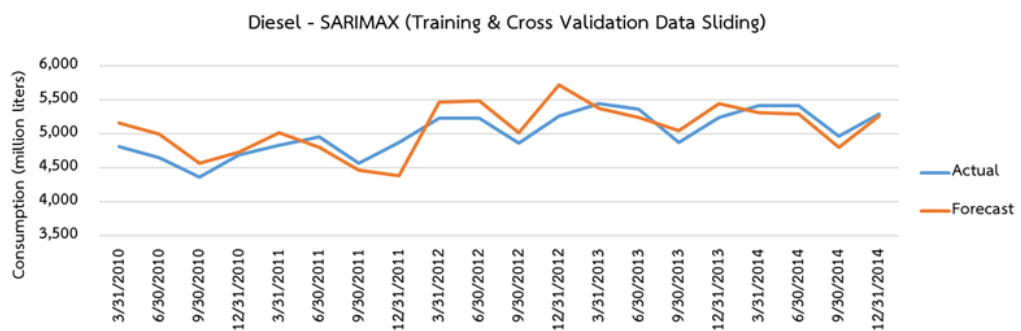
รูปที่ 64 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง



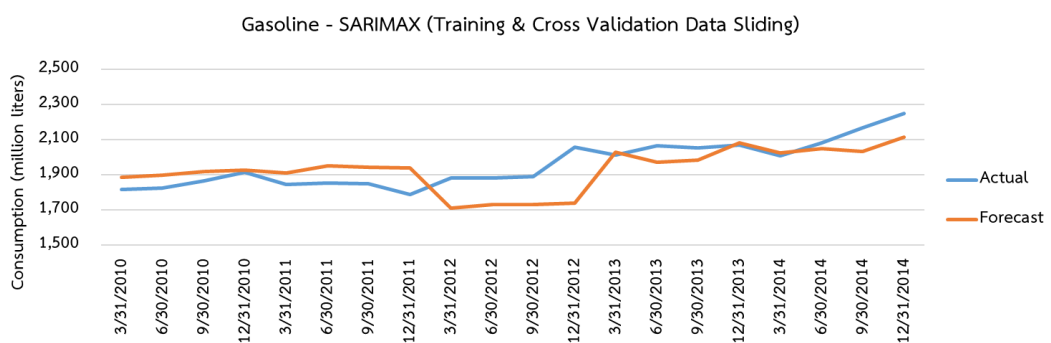
รูปที่ 65 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะกลาง

## 2) การพยากรณ์ระยะสั้น

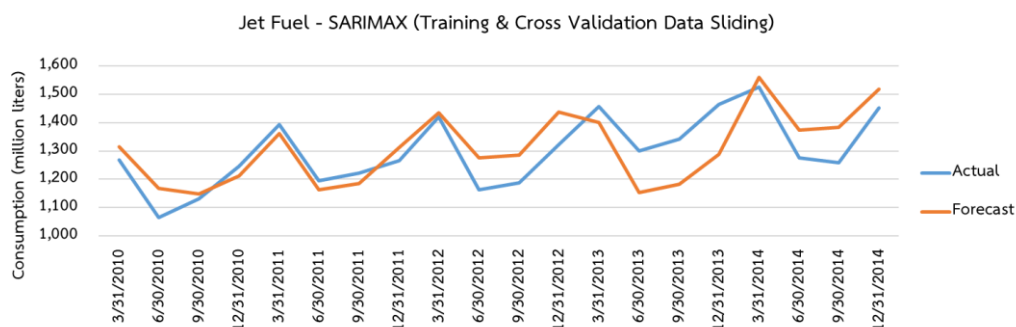
ผลลัพธ์ของตัวแบบ SARIMAX ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน สำหรับการพยากรณ์ระยะสั้นซึ่งใช้การเลื่อนชุดข้อมูลฝึกสอนไปข้างหน้าในการสร้างตัวแบบครั้งละ 4 ไตรมาส ได้ตัวแบบ SARIMAX ดังแสดงในตารางที่ 11 จะเห็นได้ว่าตัวแบบ SARIMAX ของน้ำมันดีเซล และน้ำมันเครื่องบินในแต่ละช่วงข้อมูลฝึกสอนที่ใช้กันในตัวแบบที่เหมือนกันทั้ง 5 ช่วง แต่ตัวแบบ SARIMAX ของน้ำมันดีเซลได้ตัวแบบที่แตกต่างกัน เมื่อทำการตรวจสอบรูปแบบของตัวแบบ SARIMAX เช่นเดียวกันกับตัวแบบ SARIMA ซึ่งผลลัพธ์ที่ได้พบว่ามีคล้อยกับเงื่อนไขดังกล่าว จากนั้นทำการพยากรณ์ไปล่วงหน้า 4 ไตรมาสหรือจำนวน 1 ปี โดยมีการเลื่อน (Sliding) การพยากรณ์ไปครั้งละ 4 ไตรมาส จำนวน 5 ครั้ง จนครบ 20 ไตรมาส ผลลัพธ์ของการพยากรณ์แสดงในรูปแบบที่ 66 67 และ 68



รูปที่ 66 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันดีเซลสำหรับการพยากรณ์ระยะสั้น



รูปที่ 67 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันเบนซินสำหรับการพยากรณ์ระยะสั้น



รูปที่ 68 การพยากรณ์ด้วยตัวแบบ SARIMAX ของน้ำมันเครื่องบินสำหรับการพยากรณ์ระยะสั้น

ตารางที่ 11 ตัวแบบ SARIMAX ของน้ำมันสำเร็จรูปที่ได้จากการสร้างตัวแบบด้วยชุดข้อมูลฝึกสอน สำหรับการพยากรณ์ระยะสั้น จำนวน 5 ช่วงเวลา

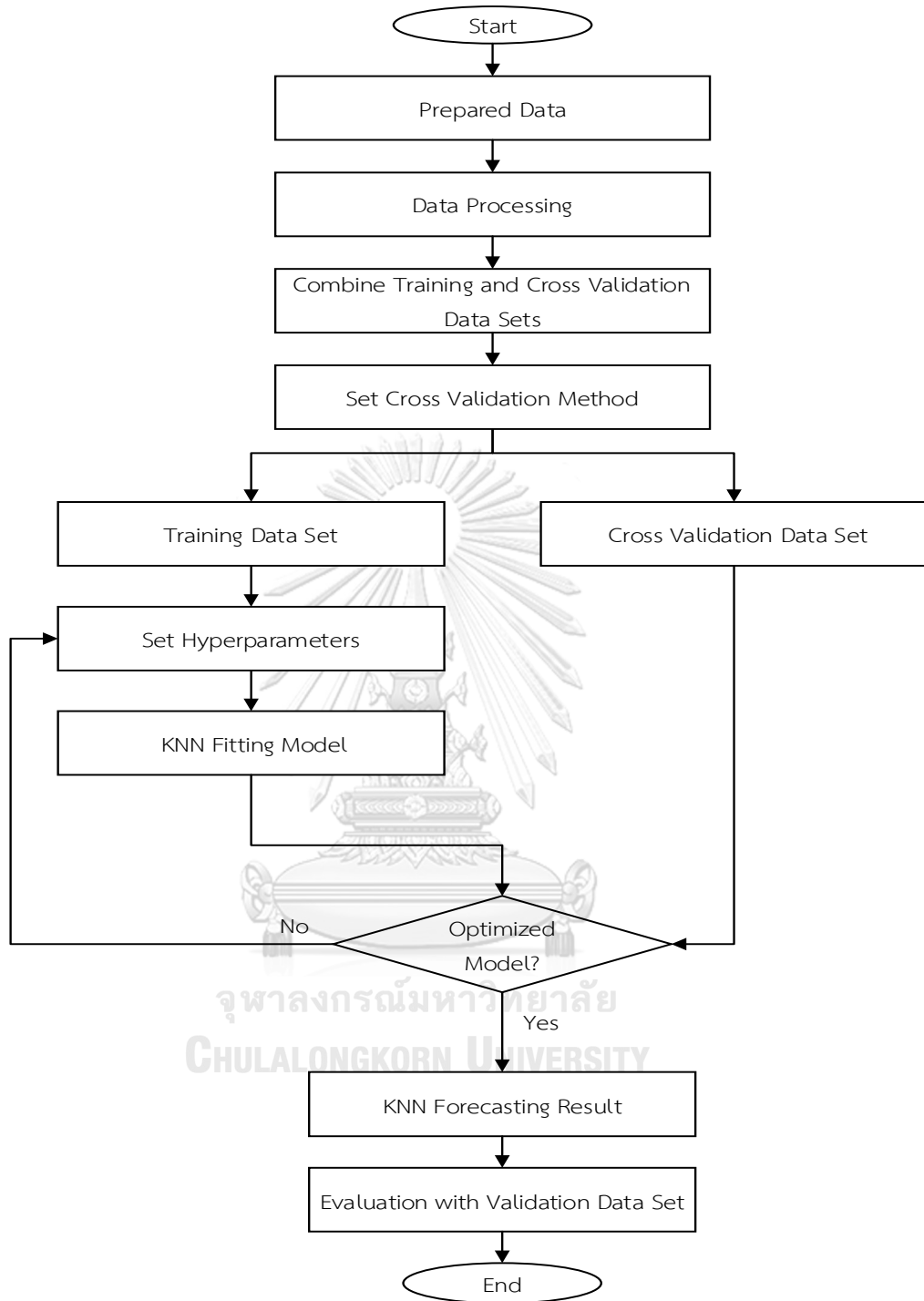
ชนิดน้ำมันสำเร็จรูป	ช่วงข้อมูลที่ใช้สร้างตัวแบบ	ตัวแบบ SARIMAX ที่ได้
ดีเซล	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
เบนซิน	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	SARIMAX(0,1,0)(1,1,0) <sub>4</sub>
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	SARIMAX(0,1,0)(1,1,0) <sub>4</sub>
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	SARIMAX(0,1,1)(0,1,1) <sub>4</sub>
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	SARIMAX(0,1,2)(0,1,2) <sub>4</sub>
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
เครื่องบิน	ช่วงที่ 1 (พ.ศ. 2536 - พ.ศ. 2553)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 2 (พ.ศ. 2537 - พ.ศ. 2554)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 3 (พ.ศ. 2538 - พ.ศ. 2555)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 4 (พ.ศ. 2539 - พ.ศ. 2556)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>
	ช่วงที่ 5 (พ.ศ. 2540 - พ.ศ. 2557)	SARIMAX(0,1,0)(0,1,1) <sub>4</sub>

จากตารางที่ 11 จะเห็นได้ว่าตัวแบบ SARIMAX ของน้ำมันดีเซลและน้ำมันเครื่องบินในแต่ละช่วงเวลาได้ผลลัพธ์เป็นตัวแบบเดียวกันทั้งหมด ส่วนของตัวแบบ SARIMAX ของน้ำมันเบนซินในช่วงที่ 3 และ 4 ได้ผลลัพธ์แตกต่างจากช่วงที่ 1 2 และ 5 ส่วนที่แตกต่างของตัวแบบคือเทอมของ  $q$  หรือเรียกว่าอันดับของตัวแบบ Moving Average ที่ไม่มีฤดูกาล ซึ่งผู้วิจัยมีความเห็นว่าน่าจะเกิดจากสาเหตุเดียวกับที่ที่เคยกล่าวไว้ในผลลัพธ์ของตัวแบบ SARIMA สำหรับการพยากรณ์ระยะสั้น คือ ข้อมูลปริมาณการบริโภคน้ำมันเบนซินในช่วงปี พ.ศ. 2536 ถึง พ.ศ. 2538 (ค.ศ. 1993 ถึง ค.ศ. 1995) นั้นมีรูปแบบแนวโน้มในลักษณะชันขึ้นมาก แต่หลังจากปี พ.ศ. 2539 จนถึง พ.ศ. 2557 (ค.ศ. 1996 ถึง ค.ศ. 2014) พบว่าข้อมูลมีรูปแบบค่อนข้างคงที่ ดังแสดงในรูปที่ 13 เมื่อทำสร้างตัวแบบการพยากรณ์ระยะสั้นแบบการเลื่อนตำแหน่งช่วงเวลา ในช่วงที่ 3 ตัวแบบครอบคลุมบางส่วนของช่วงที่มีแนวโน้มชันขึ้นมาก ในช่วงที่ 4 ตัวแบบไม่ครอบคลุมในช่วงที่มีแนวโน้มชันขึ้นมาก ซึ่งจะครอบคลุมช่วงที่ค่อนข้างคงที่เป็นส่วนใหญ่ และเนื่องจากตัวแบบ SARIMAX มีการนำตัวแปรอิสระภายนอกที่ได้มาจากการคัดเลือกด้วยเทคนิคการถดถอยแบบขั้นตอน ซึ่งตัวแปรอิสระภายนอกอาจมีการเปลี่ยนแปลงบางช่วงที่ไม่ขึ้นกับรูปแบบอนุกรมเวลาหรืออาจมีความสัมพันธ์กันเอง ผลลัพธ์ของการสร้างตัวแบบน้ำมันเบนซินช่วงที่ 3 และ 4 จึงได้ผลลัพธ์ที่แตกต่างไปจากช่วงอื่นและบางช่วงได้ผลลัพธ์ที่แตกต่างไปจากผลลัพธ์ที่ได้จากตัวแบบ SARIMA

#### 4.6 ตัวแบบ K-nearest Neighbors

ตัวแบบ K-nearest Neighbors เป็นวิธีการแบบนอนพารามิเตอร์ (Nonparametric Method) ที่ใช้ในการพยากรณ์ได้ มีหลักการคือใช้การประมาณความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตามโดยการหาค่าเฉลี่ยจากการสังเกตในละแวกเดียวกัน (เพื่อนบ้านใกล้สุด) ซึ่งจะใช้ค่า  $k$  เป็นไฮเปอร์พารามิเตอร์ที่กำหนดขนาดของเพื่อนบ้านใกล้สุด  $k$  ตัว ในการฝึกสอนตัวแบบ KNN จะใช้ชุดข้อมูลฝึกสอนที่ผ่านกระบวนการปรับสเกลของตัวแปรอิสระแล้ว จากนั้นจึงดำเนินการสร้างตัวแบบตามขั้นตอนที่กล่าวในหัวข้อที่ 3.4.2 โดยกระบวนการฝึกสอนตัวแบบ KNN แสดงดังรูปที่ 69

ผลของการฝึกสอนตัวแบบ K-nearest Neighbors โดยใช้ตัวแปรอิสระที่ได้จากการคัดเลือกด้วยเทคนิคการถดถอยแบบขั้นตอน จากการปรับไฮเปอร์พารามิเตอร์ด้วย GridSearchCV ของการสร้างตัวแบบ KNN ไฮเปอร์พารามิเตอร์ของตัวแบบ KNN ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ซึ่งให้ค่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำที่สุดในทุกการรวมกันของแต่ละค่าพารามิเตอร์ที่เป็นไปได้แสดงในตารางที่ 12

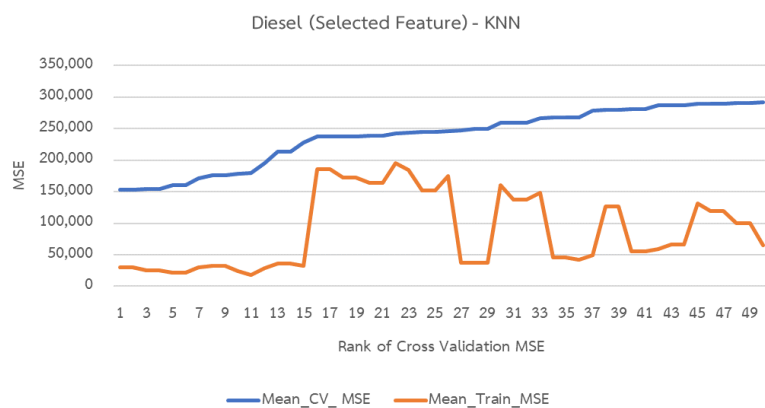


รูปที่ 69 กระบวนการสร้างตัวแบบ KNN

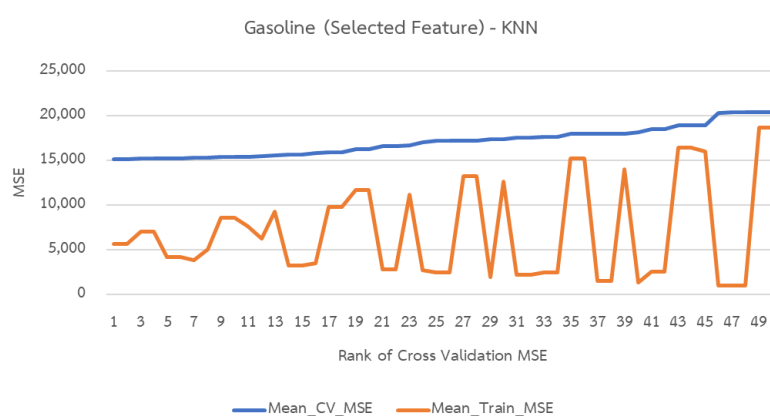
ตารางที่ 12 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ KNN

ชนิดน้ำมัน	ไฮเปอร์พารามิเตอร์	ค่ากำหนดที่ใช้ใน GridSearchCV	ค่าที่เหมาะสม
ดีเซล	n_neighbors (k)	1, 2, 3, 4, ..., 22	4
	metric	'euclidean', 'manhattan', 'minkowski'	'euclidean'
เบนซิน	n_neighbors (k)	1, 2, 3, 4, ..., 22	10
	metric	'euclidean', 'manhattan', 'minkowski'	'euclidean'
เครื่องบิน	n_neighbors (k)	1, 2, 3, 4, ..., 22	2
	metric	'euclidean', 'manhattan', 'minkowski'	'euclidean'

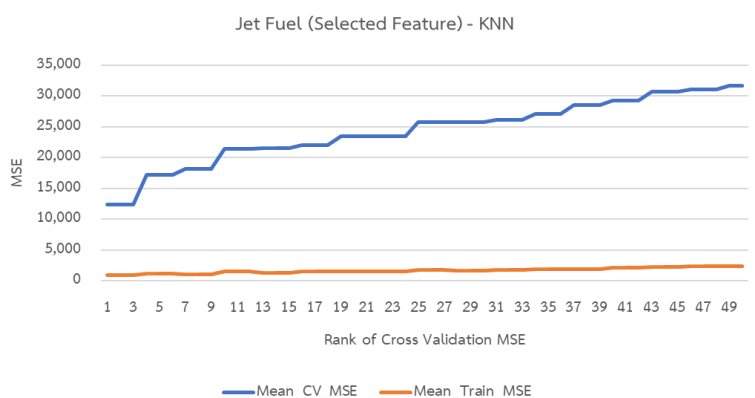
สำหรับการฝึกสอนตัวแบบ KNN แต่ละชนิดน้ำมันสำเร็จรูปใช้เวลาฝึกสอนตัวแบบจำนวน 180 ตัวแบบดังนี้ น้ำมันดีเซลใช้เวลา 1.2 วินาที น้ำมันเบนซินใช้เวลา 1.1 วินาที และน้ำมันเครื่องบินใช้เวลา 1.6 วินาที จากตารางผนวกที่ 1 2 และ 3 ค่า Mean\_CV\_MSE คือ ค่าเฉลี่ยกลุ่ม Time Series Split ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของชุดข้อมูลตรวจสอบ Cross Validation Set) ทั้ง 3 กลุ่ม และ Mean\_Train\_MSE คือ ค่าเฉลี่ยกลุ่ม Time Series Split ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของชุดข้อมูลสร้างตัวแบบ (Training Set) ทั้ง 3 กลุ่ม จะเห็นได้ว่าตัวแบบ KNN ของน้ำมันสำเร็จรูปทั้ง 3 ชนิด มีค่า Mean\_CV\_MSE สูงกว่า Mean\_Train\_MSE ค่อนข้างมาก โดยค่าสัดส่วนของ Mean\_CV\_MSE ต่อค่า Mean\_Train\_MSE ของตัวแบบ KNN ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินที่ดีที่สุดที่ได้จากค่าไฮเปอร์พารามิเตอร์ที่เหมาะสม มีค่าสัดส่วนเป็น 5.18 2.68 และ 13.73 ตามลำดับ แสดงให้เห็นว่าตัวแบบ KNN นี้ยังพบปัญหาความเข้ากันเกินไป (Overfitting) ค่อนข้างสูง จากรูปที่ 70 71 และ 72 แสดงกราฟ Mean\_CV\_MSE และ Mean\_Train\_MSE โดยมีแกนนอนเป็นลำดับของค่า Mean\_CV\_MSE ที่เรียงลำดับจากค่าน้อยไปค่ามากของตัวแบบที่ได้จากการทำ GridSearchCV เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 73 74 และ 75



รูปที่ 70 ค่า Mean CV MSE และ Mean Train MSE ของตัวแบบ KNN ของน้ำมันดีเซล

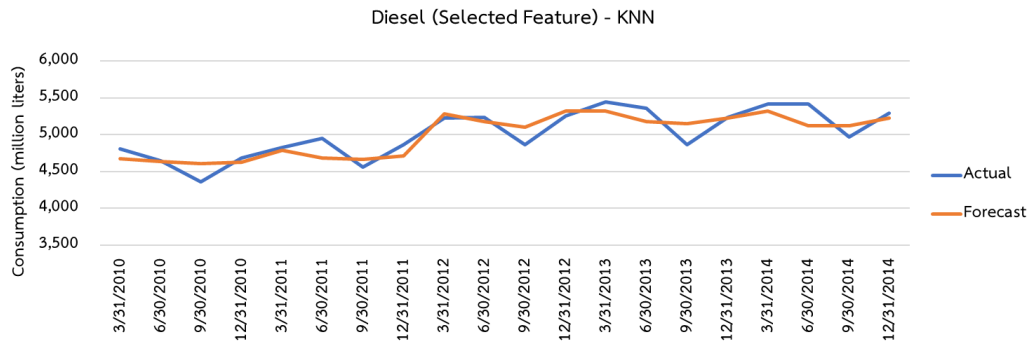


รูปที่ 71 ค่า Mean CV MSE และ Mean Train MSE ของตัวแบบ KNN ของน้ำมันเบนซิน

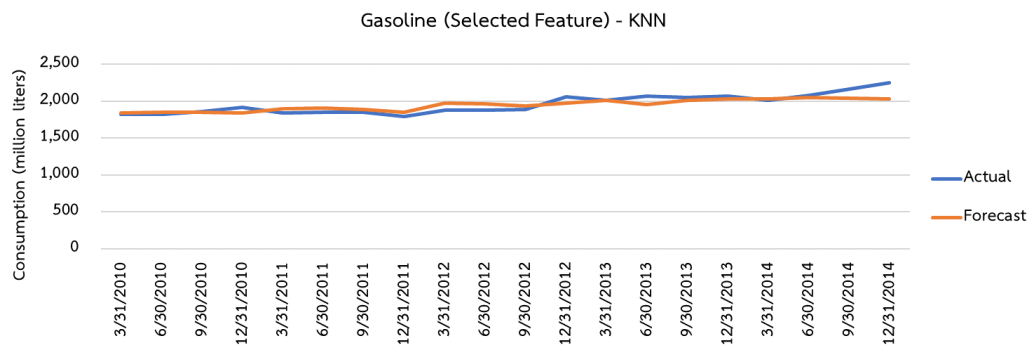


รูปที่ 72 ค่า Mean CV MSE และ Mean Train MSE ของตัวแบบ KNN ของน้ำมันเครื่องบิน

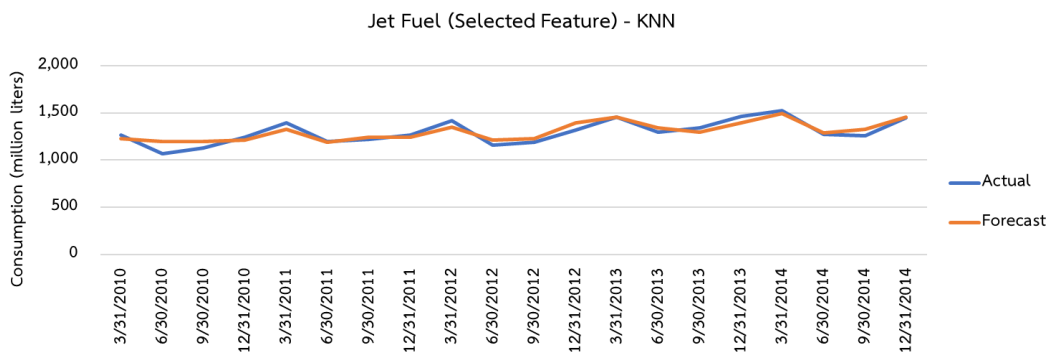




รูปที่ 73 การพยากรณ์ด้วยตัวแบบ KNN ของน้ำมันดีเซล



รูปที่ 74 การพยากรณ์ด้วยตัวแบบ KNN ของน้ำมันเบนซิน

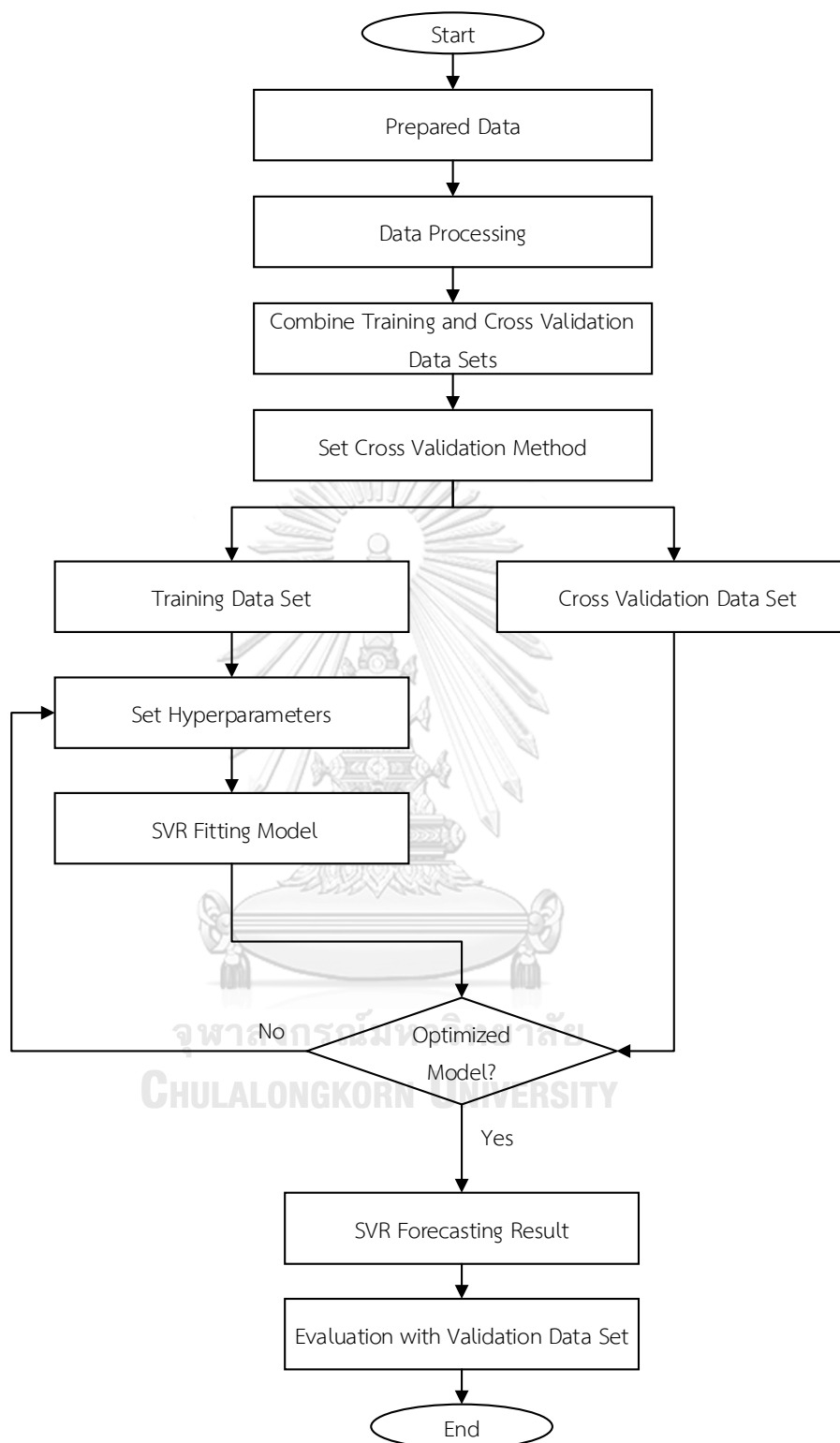


รูปที่ 75 การพยากรณ์ด้วยตัวแบบ KNN ของน้ำมันเครื่องบิน

#### 4.7 ตัวแบบ Support Vector Regression

ตัวแบบ Support Vector Regression เป็นหนึ่งในตัวแบบการเรียนรู้ของเครื่องที่มีความสามารถจับรูปแบบข้อมูลที่เป็นฟังก์ชันไม่เชิงเส้นได้โดยการเลือก Kernel Function ให้เหมาะสมกับข้อมูล ในการฝึกสอนตัวแบบ SVR จะใช้ชุดข้อมูลฝึกสอนที่ผ่านกระบวนการปรับสเกลของตัวแปรอิสระแล้ว จากนั้นจึงดำเนินการสร้างตัวแบบตามขั้นตอนที่กล่าวในหัวข้อที่ 3.4.3 โดยกระบวนการฝึกสอนตัวแบบ SVR แสดงดังรูปที่ 76 ผลของการฝึกสอนตัวแบบ SVR ที่ใช้ตัวแปรอิสระที่ได้จากการคัดเลือกด้วยเทคนิคการถดถอยแบบขั้นตอน จากการปรับไฮเปอร์พารามิเตอร์ด้วย GridSearchCV ของการสร้างตัวแบบ SVR ไฮเปอร์พารามิเตอร์ของตัวแบบ SVR ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ซึ่งให้ค่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำที่สุดในทุกการรวมกันของแต่ละค่าพารามิเตอร์ที่เป็นไปได้แสดงในตารางที่ 13 จะเห็นได้ว่าตัวแบบที่ดีที่สุดของน้ำมันดีเซลและน้ำมันเครื่องบินที่ได้มาจากการปรับไฮเปอร์พารามิเตอร์นั้นได้เลือก Kernel Function เป็น Radial Basis Function Kernel และของน้ำมันเบนซินได้เลือก Sigmoid Kernel ซึ่งเป็นฟังก์ชันไม่เชิงเส้นทั้งคู่ แสดงว่าข้อมูลน้ำมันสำเร็จรูปทั้ง 3 ชนิดนั้นเหมาะกับสมกับตัวแบบพยากรณ์ที่เป็นฟังก์ชันไม่เชิงเส้นมากกว่าฟังก์ชันเชิงเส้น

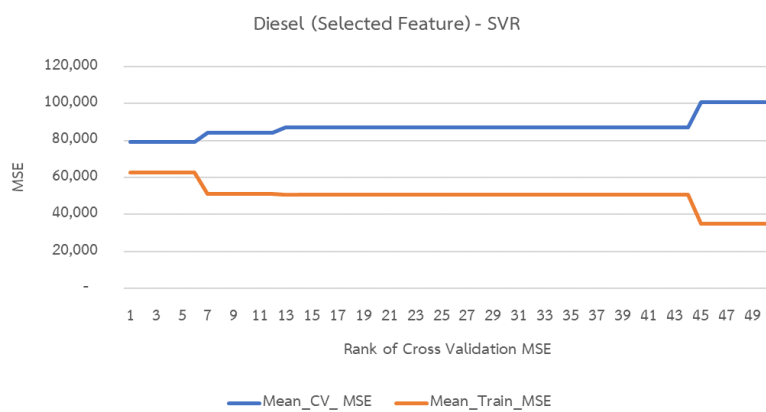
สำหรับการฝึกสอนตัวแบบ SVR แต่ละชนิดน้ำมันสำเร็จรูปใช้เวลาฝึกสอนตัวแบบจำนวน 1,800 ตัวแบบดังนี้ น้ำมันดีเซลใช้เวลา 2.5 วินาที น้ำมันเบนซินใช้เวลา 3.0 วินาที และน้ำมันเครื่องบินใช้เวลา 2.6 วินาที จากตารางผนวกที่ 4 5 และ 6 จะเห็นได้ว่าตัวแบบ SVR ของน้ำมันสำเร็จรูปทั้ง 3 ชนิด มีค่า Mean\_CV\_MSE สูงกว่า Mean\_Train\_MSE ในสัดส่วนที่น้อยกว่าของตัวแบบ KNN ดังแสดงในรูปที่ 77 78 และ 79 โดยค่าสัดส่วนของ Mean\_CV\_MSE ต่อค่า Mean\_Train\_MSE ของตัวแบบ SVR ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินที่ดีที่สุดที่ได้จากค่าไฮเปอร์พารามิเตอร์ที่เหมาะสม มีค่าสัดส่วนเป็น 1.26 1.03 และ 2.94 ตามลำดับ แสดงให้เห็นว่าตัวแบบ SVR ที่ได้มีความเข้ากันมากเกินไปของตัวแบบน้อยกว่าตัวแบบ KNN เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 80 81 และ 82



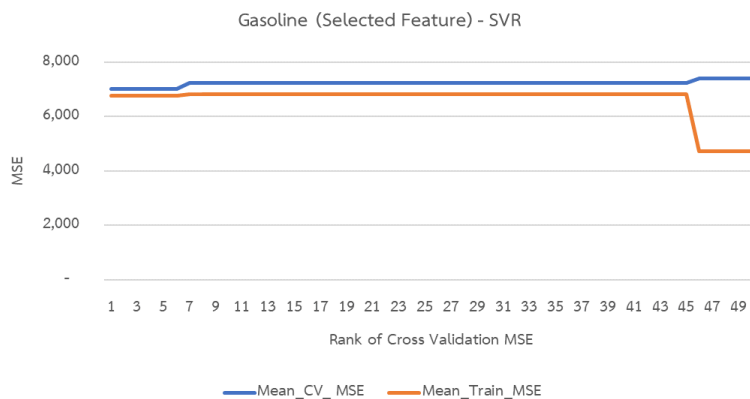
รูปที่ 76 กระบวนการสร้างตัวแบบ SVR

ตารางที่ 13 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ SVR

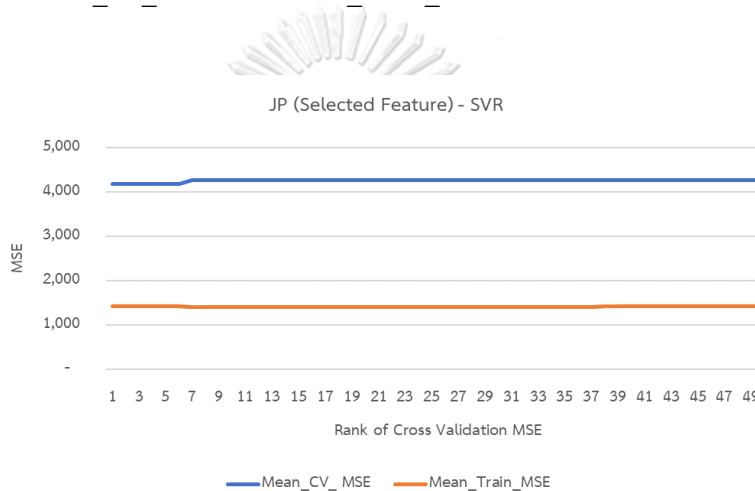
ชนิดน้ำมัน	ไฮเปอร์พารามิเตอร์	ค่ากำหนดที่ใช้ใน GridSearchCV	ค่าที่เหมาะสม
ดีเซล	C	0.1, 1, 100, 1000, 3000, 10000	3000
	kernel	'rbf', 'poly', 'sigmoid', 'linear'	'rbf'
	degree	1, 2, 3, 4, 5	1
	gamma	1, 0.1, 0.01, 0.001, 0.0001	0.1
เบนซิน	C	0.1, 1, 100, 1000, 3000, 10000	1000
	kernel	'rbf', 'poly', 'sigmoid', 'linear'	'sigmoid'
	degree	1, 2, 3, 4, 5	1
	gamma	1, 0.1, 0.01, 0.001, 0.0001	0.1
เครื่องบิน	C	0.1, 1, 100, 1000, 3000, 10000	10000
	kernel	'rbf', 'poly', 'sigmoid', 'linear'	'rbf'
	degree	1, 2, 3, 4, 5	1
	gamma	1, 0.1, 0.01, 0.001, 0.0001	0.1



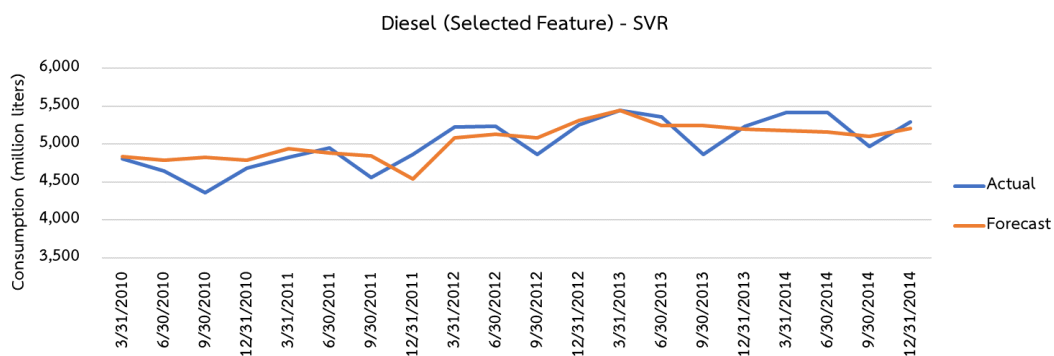
รูปที่ 77 ค่า Mean CV MSE และ Mean Train MSE ของตัวแบบ SVR ของน้ำมันดีเซล



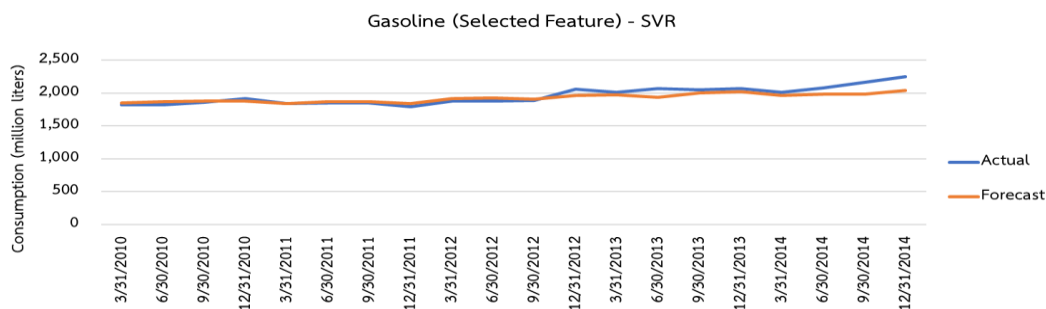
รูปที่ 78 ค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ SVR ของน้ำมันเบนซิน



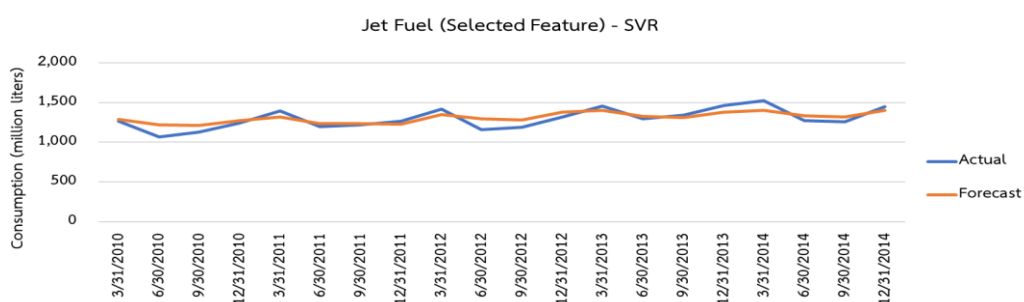
รูปที่ 79 ค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ SVR ของน้ำมันเครื่องบิน



รูปที่ 80 การพยากรณ์ด้วยตัวแบบ SVR ของน้ำมันดีเซล



รูปที่ 81 การพยากรณ์ด้วยตัวแบบ SVR ของน้ำมันเบนซิน



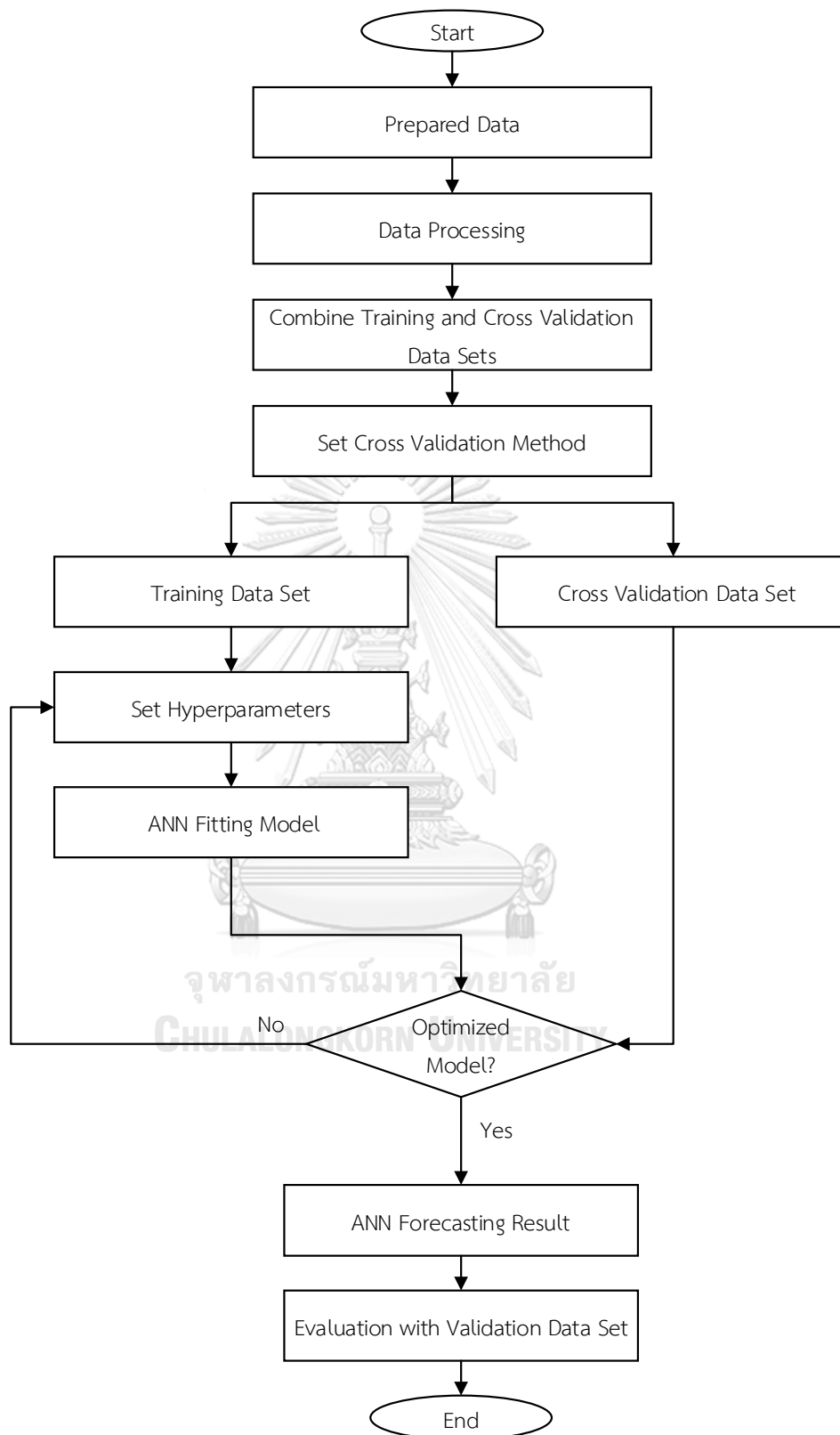
รูปที่ 82 การพยากรณ์ด้วยตัวแบบ SVR ของน้ำมันเครื่องบิน

#### 4.8 ตัวแบบ Artificial Neural Network

ตัวแบบ Artificial Neural Network มีความสามารถในการหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันที่ซับซ้อนได้ดี ซึ่งสามารถเพิ่มความแม่นยำในการพยากรณ์ให้ดีขึ้นได้ แต่ก็ขึ้นอยู่กับ การปรับไฮเปอร์พารามิเตอร์ของงานวิจัยว่าจะสามารถปรับไฮเปอร์พารามิเตอร์จนได้ค่าใช้ความแม่นยำที่ดีที่สุดเป็นเท่าใด ซึ่งเวลาที่ใช้ในการฝึกสอนตัวแบบจะเพิ่มมากขึ้นไปตามความซับซ้อนของตัวแบบ ยิ่งมีการฝึกสอนให้ตัวแบบเรียนรู้จากค่าของแต่ละไฮเปอร์พารามิเตอร์ยิ่งมาก เวลาที่ใช้ในการฝึกสอนยิ่งเพิ่มขึ้น การนำผลการพยากรณ์ไปใช้ในสถานการณ์จริงจะยิ่งเป็นการใช้ทรัพยากรสูง มีค่าใช้จ่ายมากขึ้น ดังนั้นผู้วิจัยจึงเลือกปรับไฮเปอร์พารามิเตอร์และใช้วิธีการสร้างตัวแบบ ANN ตามวิธีการในหัวข้อที่ 3.4.4 ซึ่งใช้เวลาและทรัพยากรในการสร้างตัวแบบไม่มากนัก โดยกระบวนการในการสร้างตัวแบบ ANN แสดงในรูปที่ 83 ผลของการฝึกสอนตัวแบบ ANN ที่ใช้ตัวแปรอิสระที่ได้จากการคัดเลือกด้วยเทคนิคการถดถอยแบบขั้นตอน จากการปรับไฮเปอร์พารามิเตอร์ด้วย GridSearchCV ของการสร้างตัวแบบ ANN ไฮเปอร์พารามิเตอร์ของตัวแบบ ANN ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ที่ให้ค่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำที่สุดในทุกการรวมกันของแต่ละค่าพารามิเตอร์ที่เป็นไปได้แสดงในตารางที่ 14 โดยผู้วิจัยมีความเห็นว่าการที่ตัวแบบ ANN ของน้ำมันเครื่องบินได้ไฮเปอร์พารามิเตอร์ numLayers ที่เหมาะสมเป็น 1 แต่ตัวแบบ ANN ของน้ำมันดีเซล

และน้ำมันเบนซินได้ numLayers ที่เหมาะสมเป็น 2 น่าจะเป็นเพราะว่าในการสร้างตัวแบบ ANN ของน้ำมันเครื่องบินมีตัวแปรอิสระที่ถูกคัดเลือกจากเทคนิคการถอยถอยแบบขั้นตอนเพียงหนึ่งตัวแปร ดังนั้นตัวแบบจึงอาจไม่จำเป็นต้องเรียนรู้ความสัมพันธ์ระหว่างตัวแปรอิสระที่ซับซ้อนมากขึ้น การใช้ numLayers ที่มีค่าเป็น 1 จึงมีความเหมาะสมแล้ว ในขณะที่ตัวแปรอิสระที่ถูกใช้สร้างตัวแบบ ANN ของน้ำมันดีเซลและน้ำมันเบนซินมีจำนวน 5 และ 7 ตัวแปร การใช้ numLayers ที่มีค่าเป็น 2 จึงมีความเหมาะสม เพราะว่าจะส่งผลให้ตัวแบบเรียนรู้ความสัมพันธ์ระหว่างตัวแปรอิสระได้ซับซ้อนมากยิ่งขึ้นและความแม่นยำในการพยากรณ์อาจเพิ่มขึ้นด้วย

สำหรับการฝึกสอนตัวแบบ ANN ของแต่ละชนิดน้ำมันสำเร็จรูปใช้เวลาฝึกสอนตัวแบบจำนวน 24,300 ตัวแบบดังนี้ น้ำมันดีเซลใช้เวลา 36.2 นาที น้ำมันเบนซินใช้เวลา 35.3 นาที และน้ำมันเครื่องบินใช้เวลา 33.3 นาที เนื่องจากการฝึกสอนตัวแบบ ANN และหาไฮเปอร์พารามิเตอร์ที่เหมาะสมด้วยเทคนิค GridSearchCV ต้องทำการปรับสเกลของทั้งตัวแปรต้นและตัวแปรตามเพื่อให้ใช้งาน Library ของ GridSearchCV CV กับตัวแบบ ANN ได้อย่างไม่มีปัญหา โดยปรับสเกลด้วยวิธีการแปลงเป็นค่าสูงสุด-ต่ำสุดให้อยู่ในช่วง 0.01 – 0.99 ดังนั้นค่า MSE ของตัวแบบ ANN ที่ได้จึงมีค่าน้อยมากเมื่อเทียบกับตัวแบบอื่นๆ แต่เมื่อได้ค่าน้ำหนักของ Input แต่ละโหนดจากการเรียนรู้จนได้ตัวแบบที่เหมาะสมแล้ว เมื่อทำการพยากรณ์แล้วต้องปรับสเกลค่าพยากรณ์คืนกลับไปเป็นแบบสเกลเริ่มต้น และใช้ค่าน้ำหนักเดิมของตัวแบบที่เหมาะสมที่ได้จากสร้างตัวแบบมาใช้ในการประเมินประสิทธิภาพของตัวแบบต่อไป จากตารางผนวกที่ 7 8 และ 9 พบว่าค่าสัดส่วนของ Mean\_CV\_MSE ต่อ Mean\_Train\_MSE น้อยกว่าของตัวแบบ KNN และมีค่าใกล้เคียงกับของ SVR โดยค่าสัดส่วนของ Mean\_CV\_MSE ต่อค่า Mean\_Train\_MSE ของตัวแบบ ANN ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินที่ดีที่สุดที่ได้จากค่าไฮเปอร์พารามิเตอร์ที่เหมาะสม มีค่าสัดส่วนเป็น 0.91 0.84 และ 2.31 ตามลำดับ แสดงให้เห็นว่าตัวแบบ ANN ที่ได้นี้มีปัญหาเรื่องความเข้ากันเกินไปของตัวแบบน้อยกว่าของตัวแบบ KNN ซึ่งผู้วิจัยคิดว่าเกิดจากการที่ตัวแบบ ANN มีความสามารถในการเรียนรู้ความสัมพันธ์ที่ซับซ้อนของข้อมูลได้ดีกว่าตัวแบบ KNN และเมื่อพิจารณาค่าสัดส่วน Mean\_CV\_MSE ต่อ Mean\_Train\_MSE ของน้ำมันเบนซินจะเห็นว่าสูงกว่าของน้ำมันดีเซลและน้ำมันเบนซิน ผู้วิจัยคิดว่าสาเหตุเป็นจำนวนตัวแปรอิสระที่ใช้สร้างตัวแบบ ANN ของน้ำมันเครื่องบินนั้นมีน้อยกว่าจำนวนตัวแปรอิสระที่ใช้สร้างตัวแบบ ANN ของน้ำมันดีเซลและน้ำมันเบนซิน ซึ่งน้ำมันเครื่องบินมี 1 ตัวแปรอิสระ ในขณะที่น้ำมันดีเซลและน้ำมันเบนซินมีตัวแปรอิสระจำนวน 5 และ 7 ตัวแปรตามลำดับ โดยสามารถดูค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ ANN จากรูปที่ 84 85 และ 86 เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 87 88 และ 89

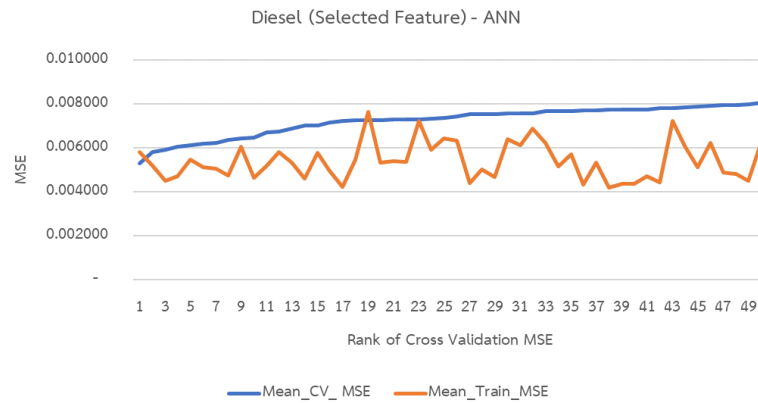


รูปที่ 83 กระบวนการสร้างตัวแบบ ANN

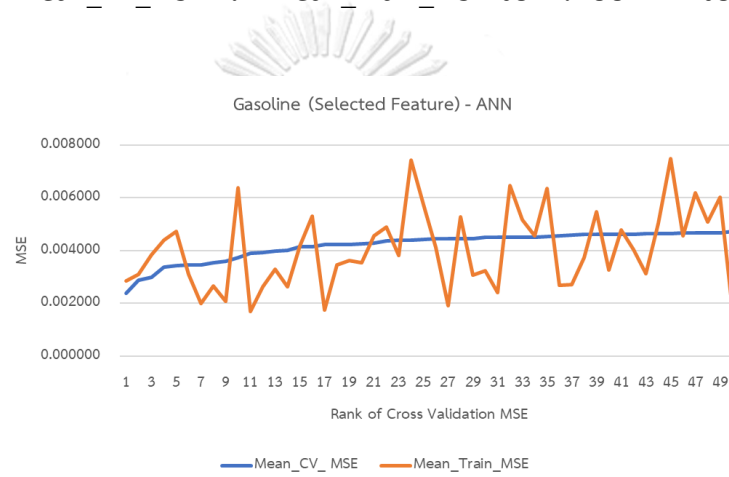


ตารางที่ 14 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ ANN

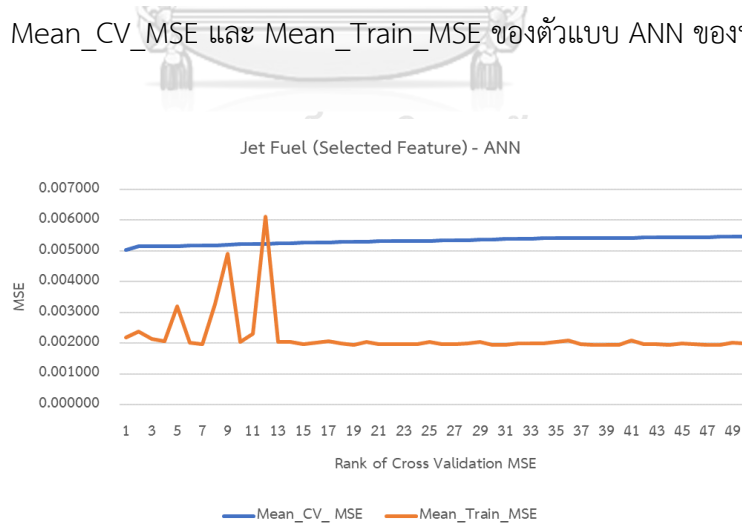
ชนิดน้ำมัน	ไฮเปอร์พารามิเตอร์	ค่ากำหนดที่ใช้ใน GridSearchCV	ค่าที่เหมาะสม
ดีเซล	batch_size	24, 48, 88	48
	epochs	25, 50, 100, 500, 1000	100
	optimizer	'adam', 'sgd'	'adam'
	units	3, 5, 15, 30, 50	30
	numLayers	1, 2	2
	activation	'relu', 'elu', 'selu'	'selu'
	dropout	0.0, 0.2, 0.4	0.2
	seed	1, 23, 456	23
เบนซิน	batch_size	24, 48, 88	24
	epochs	25, 50, 100, 500, 1000	1000
	optimizer	'adam', 'sgd'	'adam'
	units	3, 5, 15, 30, 50	5
	numLayers	1, 2	2
	activation	'relu', 'elu', 'selu'	'elu'
	dropout	0.0, 0.2, 0.4	0.2
	seed	1, 23, 456	1
เครื่องบิน	batch_size	24, 48, 88	88
	epochs	25, 50, 100, 500, 1000	500
	optimizer	'adam', 'sgd'	'adam'
	units	3, 5, 15, 30, 50	3
	numLayers	1, 2	1
	activation	'relu', 'elu', 'selu'	'elu'
	dropout	0.0, 0.2, 0.4	0.4
	seed	1, 23, 456	456



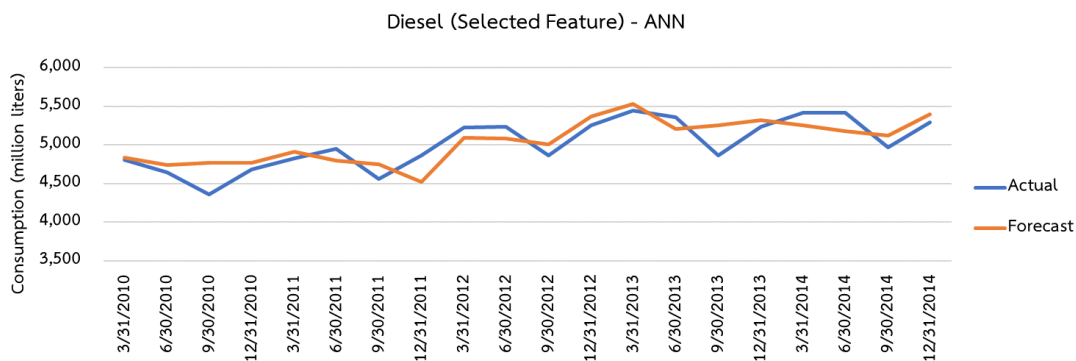
รูปที่ 84 ค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ ANN ของน้ำมันดีเซล



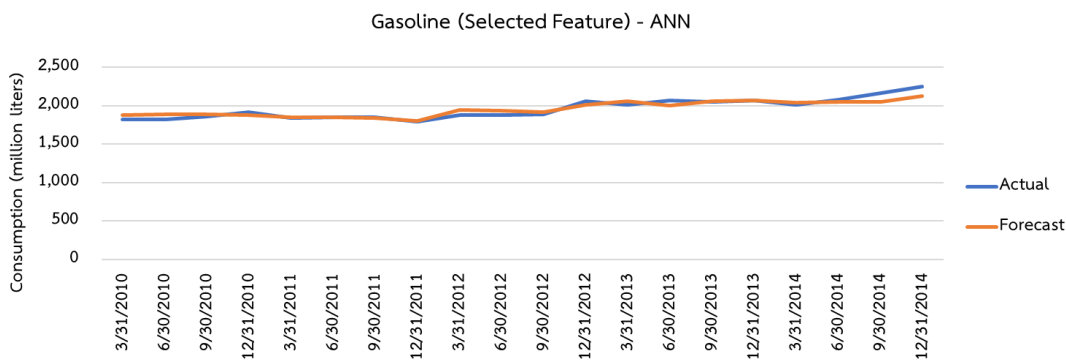
รูปที่ 85 ค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ ANN ของน้ำมันเบนซิน



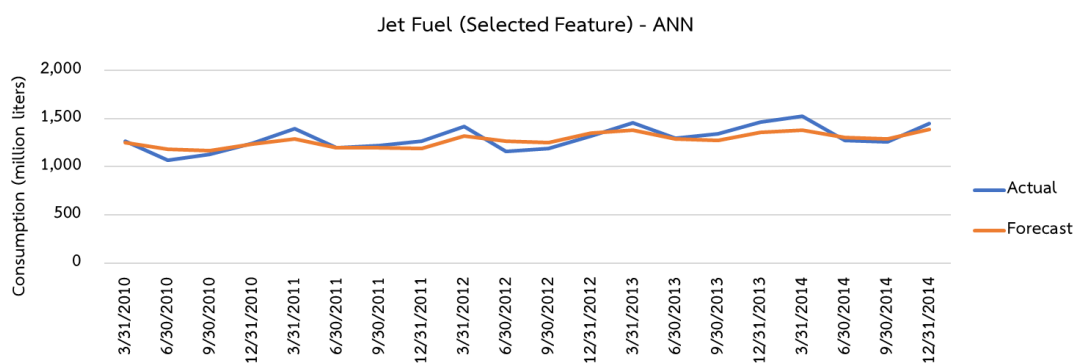
รูปที่ 86 ค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ ANN ของน้ำมันเครื่องบิน



รูปที่ 87 การพยากรณ์ด้วยตัวแบบ ANN ของน้ำมันดีเซล



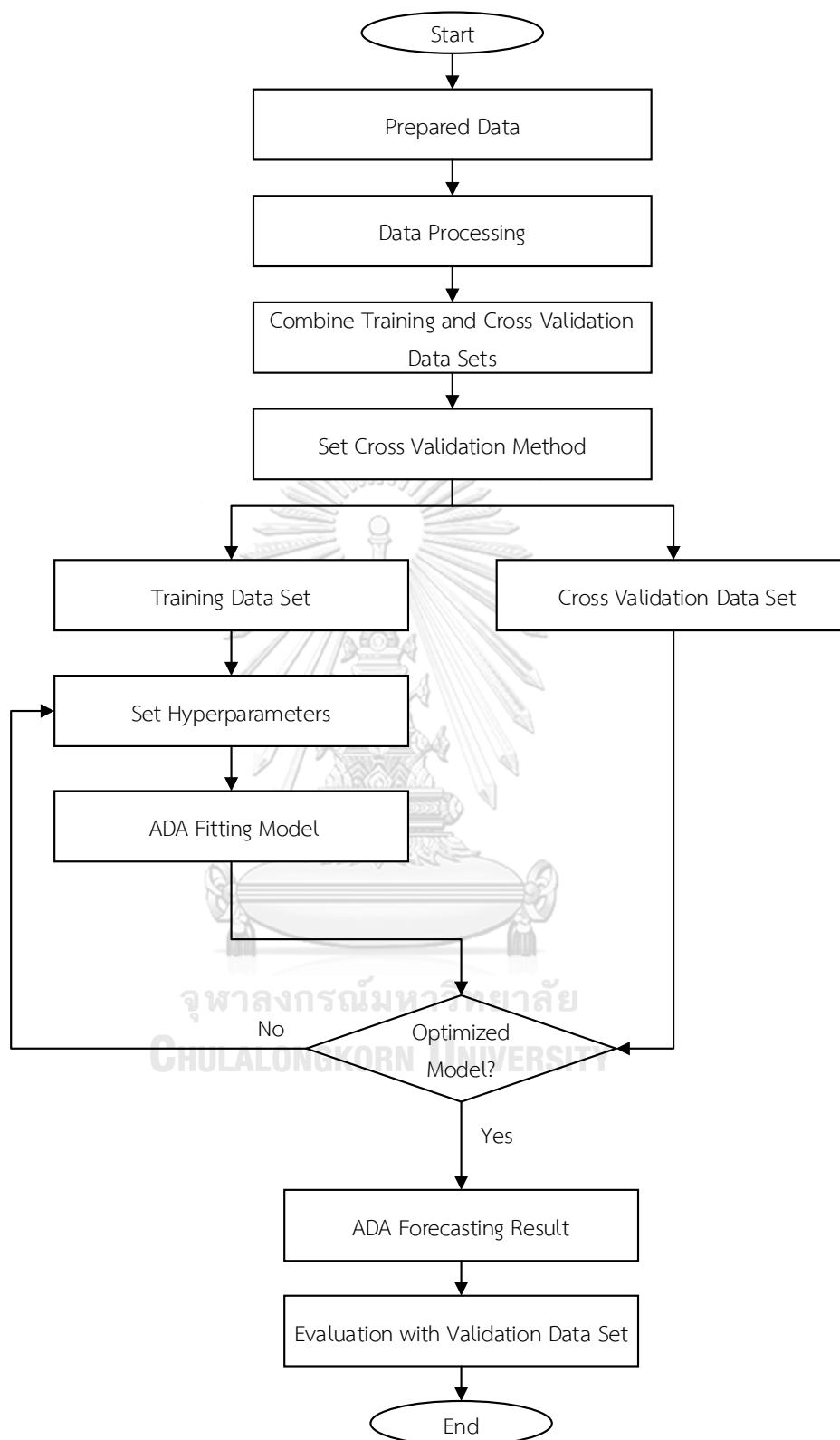
รูปที่ 88 การพยากรณ์ด้วยตัวแบบ ANN ของน้ำมันเบนซิน



รูปที่ 89 การพยากรณ์ด้วยตัวแบบ ANN ของน้ำมันเครื่องบิน

#### 4.9 ตัวแบบ Adaboost

ตัวแบบ Adaboost เป็นหนึ่งในตัวแบบกลุ่ม Ensemble โดยผู้วิจัยใช้ตัว Regression Tree เป็น Weak Learner (หมายถึงมีความแม่นยำที่ต่ำกว่าการเดาสุ่มเพียงเล็กน้อยเท่านั้น) เมื่อนำ Weak Learner หลาย ๆ ตัวแบบไปรวมกลุ่มด้วยวิธีการ Ensemble ในรูปแบบ Boosting ด้วยตัวแบบ Adaboost จะได้เป็น Strong Learner (ได้ความแม่นยำสูง) หลักการของตัวแบบ ADA คือ ฝึกสอนตัวแบบ Weak Learner ด้วยชุดข้อมูลฝึกสอน เมื่อได้ผลลัพธ์จากการฝึกสอนในครั้งแรกแล้ว จะนำชุดข้อมูลฝึกสอนที่ถูกถ่วงน้ำหนักด้วยค่าน้ำหนักสัมพัทธ์ไปใช้ในฝึกสอนตัวแบบ Weak Learner ใหม่อีกครั้ง โดยค่าน้ำหนักสัมพัทธ์จะถูกเปลี่ยนไปตามความผิดพลาดของผลลัพธ์ที่แบบตัวแบบในรอบนั้น ๆ มีความผิดพลาดต่อข้อมูลที่จะพยากรณ์ ซึ่งค่าน้ำหนักสัมพัทธ์จะถูกเปลี่ยนแปลงไปทุกครั้งตามจำนวนครั้งที่ใช้ Weak Learner ในการฝึกสอน ซึ่งเป็นหนึ่งในไฮเปอร์พารามิเตอร์ที่ต้องกำหนดนั้นคือ  $n\_estimators$  ผู้วิจัยเลือกปรับไฮเปอร์พารามิเตอร์และใช้วิธีการสร้างตัวแบบ ADA ตามวิธีการในหัวข้อที่ 3.4.5 โดยกระบวนการในการสร้างตัวแบบ ADA แสดงในรูปที่ 90 ผลของการฝึกสอนตัวแบบ ADA ที่ใช้ตัวแปรอิสระที่ได้จากการคัดเลือกด้วยเทคนิคการถดถอยแบบขั้นตอน จากการปรับไฮเปอร์พารามิเตอร์ด้วย GridSearchCV ของการสร้างตัวแบบ ADA ไฮเปอร์พารามิเตอร์ของตัวแบบ ADA ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำที่สุดในทุกการรวมกันของแต่ละค่าพารามิเตอร์ที่เป็นไปได้แสดงในตารางที่ 15

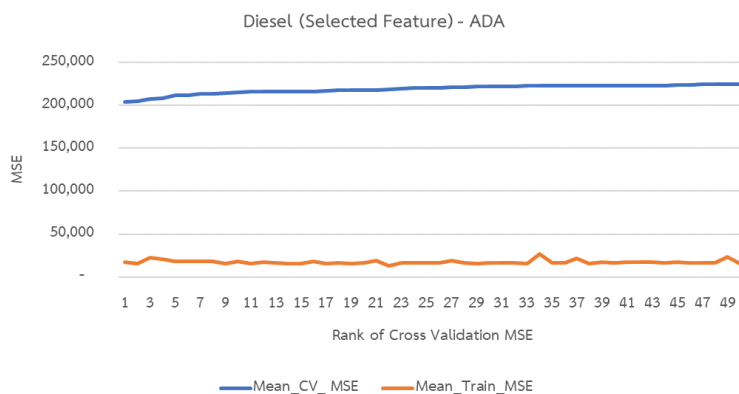


รูปที่ 90 กระบวนการสร้างตัวแบบ ADA

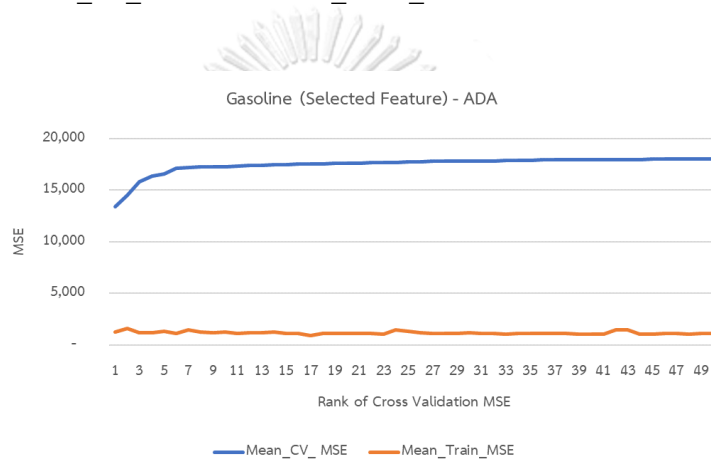
ตารางที่ 15 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ ADA

ชนิดน้ำมัน	ไฮเปอร์พารามิเตอร์	ค่ากำหนดที่ใช้ใน GridSearchCV	ค่าที่เหมาะสม
ดีเซล	n_estimators	10, 50, 100, 500, 1000	50
	learning_rate	0.0001, 0.001, 0.01, 0.1, 1.0	0.01
	random_state	1, 23, 456, 7890, 111213	1
เบนซิน	n_estimators	10, 50, 100, 500, 1000	10
	learning_rate	0.0001, 0.001, 0.01, 0.1, 1.0	0.1
	random_state	1, 23, 456, 7890, 111213	23
เครื่องบิน	n_estimators	10, 50, 100, 500, 1000	100
	learning_rate	0.0001, 0.001, 0.01, 0.1, 1.0	0.01
	random_state	1, 23, 456, 7890, 111213	23

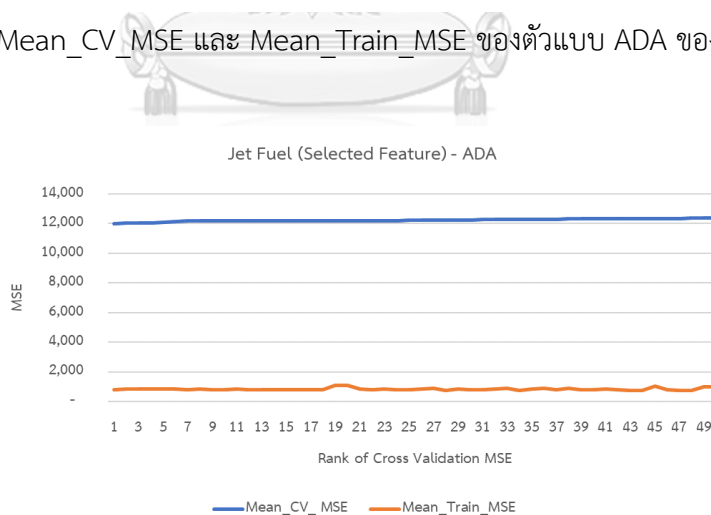
สำหรับการฝึกสอนตัวแบบ ADA แต่ละชนิดน้ำมันสำเร็จรูปใช้เวลาฝึกสอนตัวแบบจำนวน 375 ตัวแบบดังนี้ น้ำมันดีเซลใช้เวลา 8.5 วินาที น้ำมันเบนซินใช้เวลา 9.3 วินาที และน้ำมันเครื่องบินใช้เวลา 6.4 วินาที จากรูปที่ 91 92 และ 93 แสดงให้เห็นว่าตัวแบบ ADA นั้นมีปัญหาความเข้ากันเกินไปของตัวแบบค่อนข้างมาก โดยสามารถดูจากตารางผนวกที่ 10 11 และ 12 จะเห็นได้ว่าตัวแบบ ADA ของน้ำมันสำเร็จรูปทั้ง 3 ชนิด มีค่า Mean\_CV\_MSE สูงกว่า Mean\_Train\_MSE ในสัดส่วนที่สูง โดยค่าสัดส่วนของ Mean\_CV\_MSE ต่อค่า Mean\_Train\_MSE ของตัวแบบ ADA ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินที่ดีที่สุดที่ได้จากค่าไฮเปอร์พารามิเตอร์ที่เหมาะสม มีค่าสัดส่วนเป็น 11.92 10.58 และ 14.98 ตามลำดับ ซึ่งมีสัดส่วนที่สูงกว่าของตัวแบบ KNN SVR และ ANN แสดงให้เห็นว่าตัวแบบ ADA ที่ได้มีความเข้ากันมากเกินไปสูง เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 94 95 และ 96



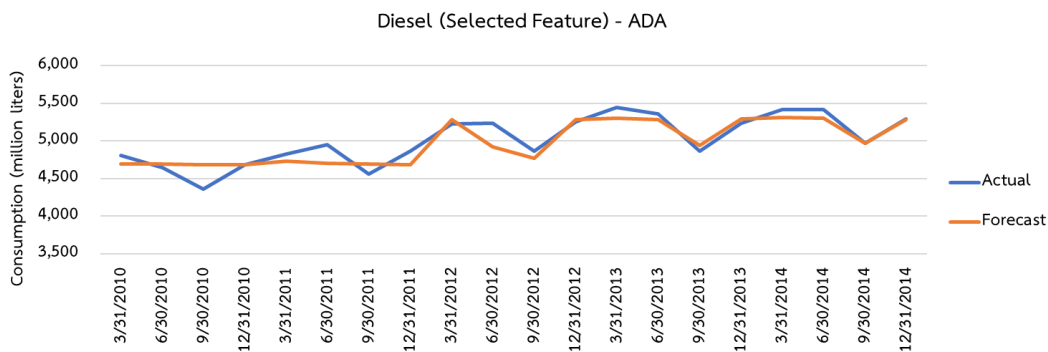
รูปที่ 91 ค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ ADA ของน้ำมันดีเซล



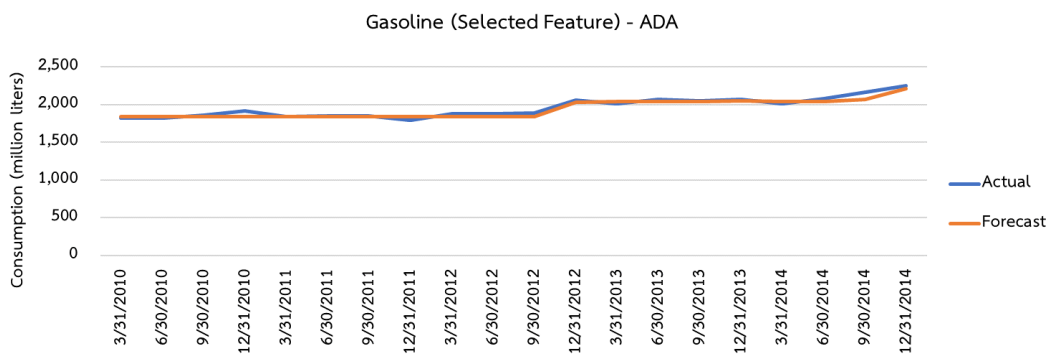
รูปที่ 92 ค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ ADA ของน้ำมันเบนซิน



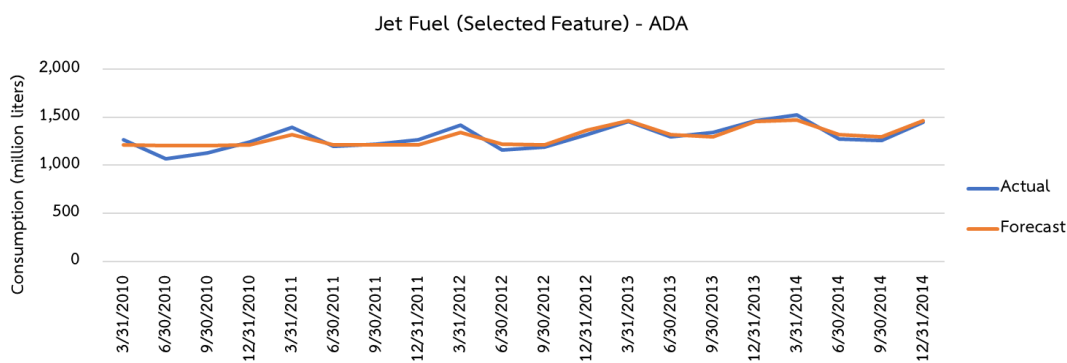
รูปที่ 93 ค่า Mean\_CV\_MSE และ Mean\_Train\_MSE ของตัวแบบ ADA ของน้ำมันเครื่องบิน



รูปที่ 94 การพยากรณ์ด้วยตัวแบบ ADA ของน้ำมันดีเซล



รูปที่ 95 การพยากรณ์ด้วยตัวแบบ ADA ของน้ำมันเบนซิน



รูปที่ 96 การพยากรณ์ด้วยตัวแบบ ADA ของน้ำมันเครื่องบิน



#### 4.10 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ในกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติและตัวแบบการถดถอยเชิงเส้นพหุคูณ

เมื่อนำชุดข้อมูลสำหรับตรวจสอบ (Cross Validation Data) กับผลการพยากรณ์ในช่วงเวลาปี พ.ศ. 2553 ถึงปี พ.ศ. 2557 สำหรับการพยากรณ์ระยะสั้นและการพยากรณ์ระยะกลางที่ได้จากแต่ละตัวแบบมาทำการวัดประสิทธิภาพของตัวแบบด้วย MAPE ได้ผลลัพธ์ดังตารางที่ 16 17 และ 18 ซึ่งในการพยากรณ์ระยะสั้น ตัวแบบการพยากรณ์ปริมาณการบริโภคน้ำมันดีเซลที่แม่นยำที่สุดจาก 5 ตัวแบบคือ SARIMA มีค่า MAPE คือ 2.5700% ตัวแบบสำหรับน้ำมันเบนซินที่แม่นยำที่สุดคือ Holt-Winters มีค่า MAPE คือ 3.6667% และในส่วนของน้ำมันเครื่องบินตัวแบบที่แม่นยำที่สุดคือ Holt-Winters มีค่า MAPE คือ 3.9985% ในส่วนของการพยากรณ์ระยะกลาง ตัวแบบการพยากรณ์ปริมาณการบริโภคน้ำมันดีเซลที่แม่นยำที่สุดจาก 5 ตัวแบบคือ SARIMAX มีค่า MAPE คือ 3.4629% ตัวแบบสำหรับน้ำมันเบนซินที่แม่นยำที่สุดคือ RANSAC-REG มีค่า MAPE คือ 2.8895% และในส่วนของน้ำมันเครื่องบินตัวแบบที่แม่นยำที่สุดคือ RANSAC-REG มีค่า MAPE คือ 4.7085%

จากการเปรียบเทียบการพยากรณ์ระยะสั้นและระยะกลางด้วยค่า MAPE เฉลี่ย 5 ปี หรือจำนวน 20 ไตรมาส พบว่าสำหรับการพยากรณ์ระยะสั้น ตัวแบบอนุกรมเวลาเชิงสถิติที่ไม่มีตัวแปรอิสระภายนอก (Holt-Winters และ SARIMA) จะมีความแม่นยำสูงกว่าตัวแบบที่มีตัวแปรอิสระภายนอก (MLR RANSAC-REG และ SARIMAX) นอกจากนี้เมื่อเปรียบเทียบกราฟการพยากรณ์ระยะสั้นและระยะกลางของตัวแบบอนุกรมเวลาเชิงสถิติของน้ำมันสำเร็จรูปทั้ง 3 ชนิดจากรูปที่ 24 ถึงรูปที่ 29 สำหรับตัวแบบ Holt-Winters และรูปที่ 39 ถึงรูปที่ 44 สำหรับตัวแบบ SARIMA พบว่าส่วนใหญ่ค่าพยากรณ์ที่ได้จะมีค่าต่ำกว่าค่าจริงโดยเฉพาะในช่วงกลางจนถึงช่วงท้ายของกราฟ ซึ่งดูเหมือนตัวแบบจะวิ่งตามแนวโน้มได้ไม่ดี แต่กราฟของการพยากรณ์ระยะสั้นของตัวแบบอนุกรมเวลาดูเหมือนจะให้ค่าพยากรณ์ที่ตามแนวโน้มได้ดี การพยากรณ์ระยะสั้นจึงพบปัญหาค่าพยากรณ์มีค่าต่ำกว่าค่าจริงน้อยกว่าการพยากรณ์ระยะกลาง สำหรับการพยากรณ์ระยะกลางนั้นพบว่าตัวแบบที่มีตัวแปรอิสระภายนอก (MLR RANSAC-REG และ SARIMAX) เมื่อพิจารณาจากค่า MAPE แล้วจะให้ผลการพยากรณ์ที่แม่นยำกว่าตัวแบบอนุกรมเวลาเชิงสถิติที่ไม่มีตัวแปรอิสระภายนอก (Holt-Winters และ SARIMA) นอกจากนี้ตัวแบบ ตัวแบบอนุกรมเวลาเชิงสถิติที่ไม่มีตัวแปรอิสระภายนอก (Holt-Winters และ SARIMA) จะให้ผลการพยากรณ์ระยะสั้นที่นำมาเฉลี่ย 5 ปีแม่นยำกว่าการพยากรณ์ระยะกลาง ส่วนตัวแบบที่มีตัวแปรอิสระภายนอก (MLR RANSAC-REG และ SARIMAX) นั้นมีทั้งที่ผลการพยากรณ์ระยะกลางแม่นยำกว่าระยะสั้น และผลการพยากรณ์ระยะสั้นแม่นยำกว่าระยะกลาง ผู้วิจัยคิดว่าน่าจะเป็นผลมาจากปัจจัยภายนอกที่อาจมีการเปลี่ยนแปลงบางช่วงที่ไม่ขึ้นกับรูปแบบอนุกรมเวลาหรืออาจมีความสัมพันธ์กันเอง และเมื่อพิจารณาการเปรียบเทียบความแม่นยำระหว่างกลุ่มตัว

แบบพยากรณ์เชิงสาเหตุด้วยกันคือตัวแบบ MLR และตัวแบบ RANSAC-REG พบว่าตัวแบบ RANSAC-REG ให้ความแม่นยำในการพยากรณ์ดีกว่าตัวแบบ MLR ในการพยากรณ์ทั้งแบบระยะสั้นและแบบระยะยาว ซึ่งผู้วิจัยคิดที่น่าจะเป็นเพราะตัวแบบ RANSAC-REG ที่มีความแกร่ง (Robust) ต่อค่าผิดปกติของข้อมูลป้อนเข้า (Outlier ของตัวแปรอิสระ) ในขณะที่ตัวแบบ MLR นั้นอ่อนไหวต่อค่าผิดปกติของข้อมูลป้อนเข้า

ตารางที่ 16 ค่า MAPE ของแต่ละตัวแบบการพยากรณ์ระยะสั้นและระยะกลางของน้ำมันดีเซลสำหรับการตรวจสอบตัวแบบในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557

ชนิดน้ำมัน	ตัวแบบ	ระยะการพยากรณ์	MAPE (%) เฉลี่ย 5 ปี (20 ไตรมาส)
ดีเซล	Holt-Winters	ระยะกลาง	4.4917
		ระยะสั้น	2.8315
	SARIMA	ระยะกลาง	5.5853
		ระยะสั้น	2.5700
	MLR	ระยะกลาง	6.2489
		ระยะสั้น	6.5021
	RANSAC-REG	ระยะกลาง	4.8850
		ระยะสั้น	6.4263
	SARIMAX	ระยะกลาง	3.4629
		ระยะสั้น	3.9273

ตารางที่ 17 ค่า MAPE ของแต่ละตัวแบบการพยากรณ์ระยะสั้นและระยะกลางของน้ำมันเบนซิน  
สำหรับการตรวจสอบตัวแบบในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557

ชนิดน้ำมัน	ตัวแบบ	ระยะการพยากรณ์	MAPE (%) เฉลี่ย 5 ปี (20 ไตรมาส)
เบนซิน	Holt-Winters	ระยะกลาง	4.9547
		ระยะสั้น	3.6667
	SARIMA	ระยะกลาง	7.1490
		ระยะสั้น	4.4888
	MLR	ระยะกลาง	3.2701
		ระยะสั้น	5.8355
	RANSAC-REG	ระยะกลาง	2.8895
		ระยะสั้น	5.7378
	SARIMAX	ระยะกลาง	3.5834
		ระยะสั้น	5.1656

ตารางที่ 18 ค่า MAPE ของแต่ละตัวแบบการพยากรณ์ระยะสั้นและระยะกลางของน้ำมันเครื่องบิน  
สำหรับการตรวจสอบตัวแบบในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557

ชนิดน้ำมัน	ตัวแบบ	ระยะการพยากรณ์	MAPE (%) เฉลี่ย 5 ปี (20 ไตรมาส)
เครื่องบิน	Holt-Winters	ระยะกลาง	5.4177
		ระยะสั้น	3.9985
	SARIMA	ระยะกลาง	5.5589
		ระยะสั้น	4.1978
	MLR	ระยะกลาง	4.9555
		ระยะสั้น	8.8966
	RANSAC-REG	ระยะกลาง	4.7085
		ระยะสั้น	8.6805
	SARIMAX	ระยะกลาง	4.7452
		ระยะสั้น	6.0321

#### 4.11 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ในกลุ่มตัวแบบการเรียนรู้ของเครื่อง

เมื่อนำชุดข้อมูลสำหรับตรวจสอบ (Cross Validation Data) กับผลการพยากรณ์ในช่วงเวลาปี พ.ศ. 2553 ถึงปี พ.ศ. 2557 สำหรับการพยากรณ์กลุ่มตัวแบบการเรียนรู้ของเครื่องแบบใช้ตัวแปรอิสระที่คัดเลือกด้วยเทคนิคการถดถอยแบบขั้นตอนมาทำการวัดประสิทธิภาพของตัวแบบด้วย MAPE ได้ผลลัพธ์ดังตารางที่ 19 ซึ่งผลลัพธ์การวัดประสิทธิภาพตัวแบบของน้ำมันดีเซลพบว่า ตัวแบบที่แม่นยำที่สุดแบบเรียงตามอันดับคือ ตัวแบบ ADA ตัวแบบ KNN ตัวแบบ ANN และตัวแบบ SVR ซึ่งมีค่า MAPE คือ 2.2348% 2.6534% 3.3364% และ 3.3988% ตามลำดับ แต่ตัวแบบ KNN และ ADA นั้นพบปัญหาการเข้ากันเกินไปสูงกว่าตัวแบบ SVR และ ANN มาก เนื่องจากค่า MAPE ของทุกตัวแบบมีค่าต่ำกว่า 5% ซึ่งถือว่ามีความแม่นยำค่อนข้างสูง ดังนั้นผู้วิจัยจึงเลือกตัวแบบ SVR และ ANN ให้เป็นตัวแบบที่เหมาะสมในกลุ่มตัวแบบการเรียนรู้ของเครื่องของน้ำมันดีเซลที่จะนำไปใช้ในการผสมตัวแบบต่อไป สำหรับผลลัพธ์การวัดประสิทธิภาพตัวแบบของน้ำมันเบนซินพบว่า ตัวแบบที่แม่นยำที่สุดแบบเรียงตามอันดับคือ ตัวแบบ ADA ตัวแบบ ANN ตัวแบบ SVR และตัวแบบ KNN มีค่า MAPE คือ 1.6357% 2.0596% 2.9376% และ 3.0597% ตามลำดับ ซึ่งเป็นเช่นเดียวกับตัวแบบของน้ำมันเบนซินที่ตัวแบบ ADA และ KNN มีปัญหาความเข้ากันเกินไปสูง ดังนั้นผู้วิจัยจึงเลือกตัวแบบ SVR และ ANN เป็นตัวแบบที่เหมาะสมในกลุ่มตัวแบบการเรียนรู้ของเครื่องของน้ำมันเบนซินที่จะนำไปใช้ในการผสมตัวแบบต่อไป สำหรับผลลัพธ์การวัดประสิทธิภาพตัวแบบของน้ำมันเครื่องบินพบว่า ตัวแบบที่แม่นยำที่สุด 5 อันดับแรกคือ ตัวแบบ ADA ตัวแบบ KNN ตัวแบบ ANN และตัวแบบ SVR มีค่า MAPE คือ 3.5401% 3.5516% 4.5870% และ 5.0404% ตามลำดับ ซึ่งเป็นเช่นเดียวกับตัวแบบของน้ำมันดีเซลและน้ำมันเบนซินที่ตัวแบบ ADA และ KNN มีปัญหาความเข้ากันเกินไปสูง ดังนั้นผู้วิจัยจึงเลือกตัวแบบ SVR และ ANN เป็นตัวแบบที่เหมาะสมในกลุ่มตัวแบบการเรียนรู้ของเครื่องของน้ำมันเครื่องบินที่จะนำไปใช้ในการผสมตัวแบบต่อไป

ตารางที่ 19 ค่า MAPE ของแต่ละตัวแบบการเรียนรู้ของเครื่องสำหรับการตรวจสอบตัวแบบของน้ำมันสำเร็จรูปกลุ่มหลัก ในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557

ชนิดน้ำมัน	ตัวแบบ	MAPE (%) เฉลี่ย 5 ปี (20 ไตรมาส)	สัดส่วน Mean_CV_MSE ต่อ Mean_Train_MSE
ดีเซล	KNN	2.6534	5.18
	SVR	3.3988	1.26
	ANN	3.3364	0.91
	ADA	2.2348	11.92
เบนซิน	KNN	3.0597	2.68
	SVR	2.9376	1.03
	ANN	2.0596	0.84
	ADA	1.6357	10.58
เครื่องบิน	KNN	3.5516	13.73
	SVR	5.0404	2.94
	ANN	4.5870	2.31
	ADA	3.5401	14.98

#### 4.12 ตัวแบบผสม (Hybrid Model)

ผู้วิจัยศึกษาตัวแบบผสมเพื่อปรับปรุงความแม่นยำของตัวแบบเดี่ยวที่ใช้พยากรณ์ปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินให้ดีขึ้น จากผลลัพธ์ของตัวแบบเดี่ยวที่ได้กล่าวถึงในหัวข้อที่ผ่านมา เพื่อให้ตัวแบบผสมมีความสามารถในการตรวจจับรูปแบบข้อมูลให้ได้ทั้งรูปแบบในฟังก์ชันเชิงเส้นและรูปแบบฟังก์ชันไม่เชิงเส้น และการสร้างตัวแบบที่ครอบคลุม ผู้วิจัยจึงเลือกตัวแบบการพยากรณ์จากกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติเป็นตัวแบบ SARIMAX ที่ใช้ในการพยากรณ์ระยะกลาง ซึ่งให้ค่าความแม่นยำที่ค่อนข้างสูงเมื่อเทียบกับตัวอื่น และยังสามารถใช้ทดแทนการพยากรณ์ระยะสั้นได้เนื่องจากเมื่อพยากรณ์ไปข้างหน้าจำนวน 20 ไตรมาส ยังให้ค่า MAPE ที่ค่อนข้างต่ำคือไม่เกิน 5% โดยจะใช้ตัวแบบ SARIMAX ไปพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้น ส่วนตัวแบบในกลุ่มการเรียนรู้ของเครื่องที่เหมาะสมสำหรับนำไปพยากรณ์ข้อมูลในส่วนของฟังก์ชันไม่เชิงเส้น ผู้วิจัยเลือกใช้ตัวแบบ SVR หรือ ANN แบบใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอน เพื่อให้สามารถผสมกับตัวแบบ SARIMAX ที่ใช้ตัวแปรอิสระภายนอกจากการคัดเลือกจากการถดถอยแบบขั้นตอนได้แบบใช้ตัวแปรอิสระตัวเดียวกันในการสร้างตัวแบบผสม โดยตัวแบบ SVR และ ANN ที่เลือกใช้มีความแม่นยำสูงและมีปัญหาการเข้ากันเกินไปน้อยกว่าตัวแบบ ADA และ KNN ตัวแบบ SVR และ ANN จึงเหมาะสมที่จะนำมาใช้ในการพยากรณ์ข้อมูลในส่วนที่เป็นฟังก์ชันไม่เชิงเส้น จากผลลัพธ์ของการฝึกสอนตัวแบบผสมมีดังนี้

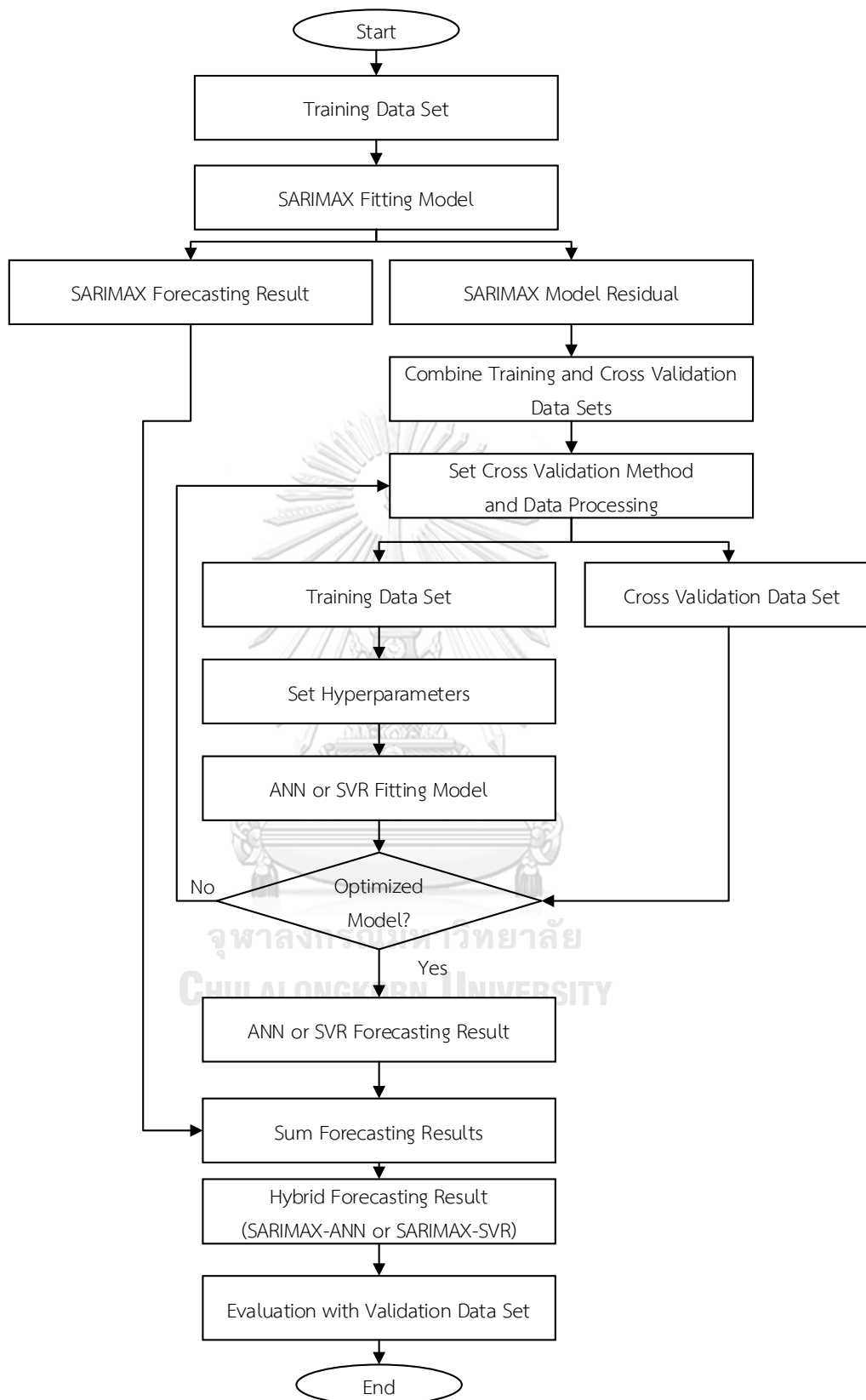
##### 4.12.1 ตัวแบบ SARIMAX-ANN

ตัวแบบ SARIMAX-ANN ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินเป็นตัวแบบที่มีโครงสร้างการผสมแบบอนุกรม ซึ่งมีกระบวนการสร้างตัวแบบดังแสดงในรูปที่ 97 โดยเริ่มจากนำตัวแบบ SARIMAX ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ที่ได้ผลลัพธ์การสร้างตัวแบบดังที่ได้กล่าวในหัวข้อที่ 4.5 มาพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้น แล้วจึงนำค่าส่วนเหลือ (Residual) ของตัวแบบ SARIMAX ไปพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันไม่เชิงเส้นด้วยตัวแบบ ANN โดยมีการปรับไฮเปอร์พารามิเตอร์ด้วยวิธี GridSearchCV เช่นเดียวกับกับตัว ANN แบบเดี่ยว แต่เปลี่ยนไฮเปอร์พารามิเตอร์ของฟังก์ชันกระตุ้นของชั้นซ่อนเป็นฟังก์ชัน Hyperbolic Tangent ('activation': 'tanh') และของปัจจัยนำออก (Output) เป็นฟังก์ชันเชิงเส้น ('activation': 'linear') เพื่อให้รองรับข้อมูลของค่าส่วนเหลือที่มีค่าเป็นลบได้ เนื่องจากฟังก์ชันกระตุ้นที่ใช้กับตัวแบบ ANN แบบเดี่ยวที่ใช้พยากรณ์น้ำมันสำเร็จรูปไม่มีค่าพยากรณ์ที่เป็นลบ จึงสามารถใช้ฟังก์ชันกระตุ้นจำพวกฟังก์ชัน Rectified Linear Unit (relu) ฟังก์ชัน Exponential Linear Unit (elu) และฟังก์ชัน Scaled Exponential Linear Unit (selu) ได้ ผลของการฝึกสอนตัวแบบ ANN ที่ใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอนและใช้ค่าส่วนเหลือของตัวแบบ SARIMAX เป็นตัวแปรตาม

ของตัวแบบ ANN เริ่มจากการปรับไฮเปอร์พารามิเตอร์ด้วย GridSearchCV ไฮเปอร์พารามิเตอร์ของตัวแบบ ANN ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ที่ให้ค่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำที่สุดในทุกการรวมกันของแต่ละค่าพารามิเตอร์ที่เป็นไปได้ ซึ่งได้ไฮเปอร์พารามิเตอร์ที่เหมาะสมแสดงในตารางที่ 20 จากนั้นจึงรวมผลการพยากรณ์ในส่วนที่เป็นฟังก์ชันเชิงเส้นและฟังก์ชันไม่เชิงเส้นเข้าด้วยกัน ตัวแบบ SARIAMX-ANN ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินใช้เวลาในการฝึกสอนตัวแบบทั้งหมด 15.5 นาที 12.9 นาที และ 14.1 นาที ตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 98 99 และ 100



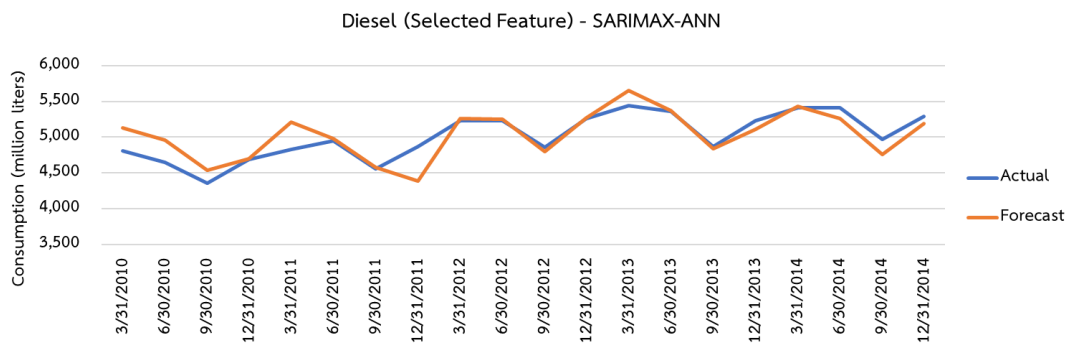




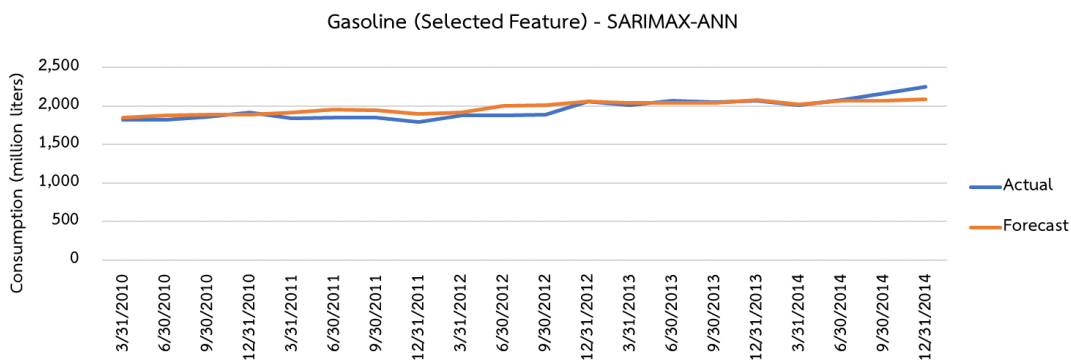
รูปที่ 97 กระบวนการสร้างตัวแบบ SARIMAX-ANN และตัวแบบ SARIMAX-SVR

ตารางที่ 20 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ ANN ที่ใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอนและใช้ค่าส่วนเหลือของตัวแบบ SARIMAX เป็นตัวแปรตาม

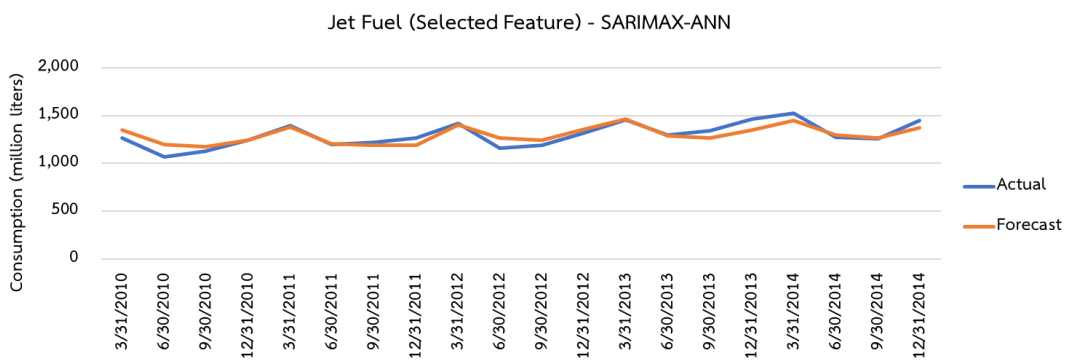
ชนิดน้ำมัน	ไฮเปอร์พารามิเตอร์	ค่ากำหนดที่ใช้ใน GridSearchCV	ค่าที่เหมาะสม
ดีเซล	batch_size	24, 48, 88	24
	epochs	25, 50, 100, 500, 1000	500
	optimizer	'adam', 'sgd'	'adam'
	units	3, 5, 15, 30, 50	50
	numLayers	1, 2	1
	activation	'tanh'	'tanh'
	dropout	0.0, 0.2, 0.4	0.0
	seed	1, 23, 456	456
เบนซิน	batch_size	24, 48, 88	24
	epochs	25, 50, 100, 500, 1000	1000
	optimizer	'adam', 'sgd'	'adam'
	units	3, 5, 15, 30, 50	30
	numLayers	1, 2	2
	activation	'tanh'	'tanh'
	dropout	0.0, 0.2, 0.4	0.2
	seed	1, 23, 456	456
เครื่องบิน	batch_size	24, 48, 88	24
	epochs	25, 50, 100, 500, 1000	100
	optimizer	'adam', 'sgd'	'sgd'
	units	3, 5, 15, 30, 50	50
	numLayers	1, 2	2
	activation	'tanh'	'tanh'
	dropout	0.0, 0.2, 0.4	0.4
	seed	1, 23, 456	456



รูปที่ 98 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN ของน้ำมันดีเซล



รูปที่ 99 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN ของน้ำมันเบนซิน



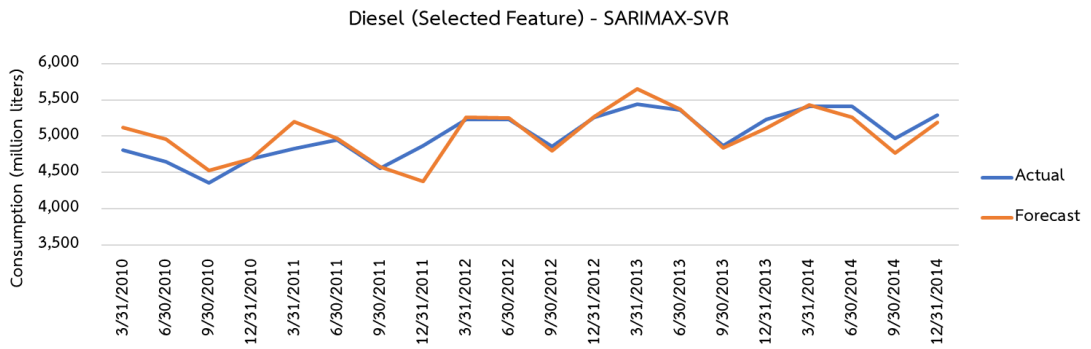
รูปที่ 100 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN ของน้ำมันเครื่องบิน

#### 4.12.2 ตัวแบบ SARIMAX-SVR

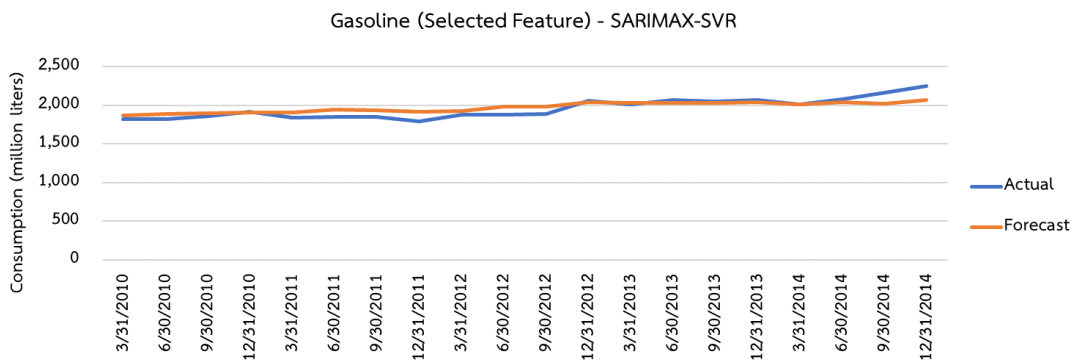
ตัวแบบ SARIMAX-SVR ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินเป็นตัวแบบที่มีโครงสร้างการผสมแบบอนุกรม ซึ่งมีกระบวนการสร้างตัวแบบดังแสดงในรูปที่ 97 โดยนำตัวแบบ SARIMAX ที่ได้ผลลัพธ์การสร้างตัวแบบดังที่ได้กล่าวในหัวข้อที่ 4.5 มาพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้น แล้วจึงนำค่าส่วนเหลือ (Residual) ของตัวแบบ SARIMAX ไปพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันไม่เชิงเส้นด้วยตัวแบบ SVR ที่ใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอน และใช้ค่าส่วนเหลือของตัวแบบ SARIMAX เป็นตัวแปรตาม โดยมีการปรับไฮเปอร์พารามิเตอร์ด้วยวิธี GridSearchCV เช่นเดียวกันกับตัวแบบ SVR แบบเดียว ซึ่งได้ไฮเปอร์พารามิเตอร์ที่เหมาะสมดังแสดงในตารางที่ 21 จากนั้นจึงรวมผลการพยากรณ์ในส่วนที่เป็นฟังก์ชันเชิงเส้นและฟังก์ชันไม่เชิงเส้นเข้าด้วยกัน ตัวแบบ SARIAMX-SVR ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินใช้เวลาในการฝึกสอนตัวแบบทั้งหมด 16.7 วินาที 12.9 วินาที และ 4.3 วินาที ตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 101 102 และ 103

ตารางที่ 21 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ SVR ที่ใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอนและใช้ค่าส่วนเหลือของตัวแบบ SARIMAX เป็นตัวแปรตาม

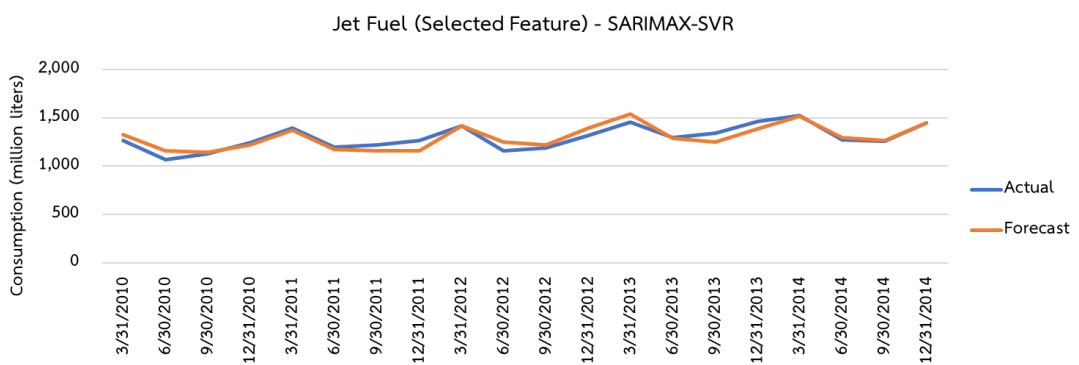
ชนิดน้ำมัน	ไฮเปอร์พารามิเตอร์	ค่ากำหนดที่ใช้ใน GridSearchCV	ค่าที่เหมาะสม
ดีเซล	C	0.1, 1, 100, 1000, 3000, 10000	100
	kernel	'rbf', 'poly', 'sigmoid', 'linear'	'poly'
	degree	1, 2, 3, 4, 5	2
	gamma	1, 0.1, 0.01, 0.001, 0.0001	0.01
เบนซิน	C	0.1, 1, 100, 1000, 3000, 10000	1
	kernel	'rbf', 'poly', 'sigmoid', 'linear'	'poly'
	degree	1, 2, 3, 4, 5	2
	gamma	1, 0.1, 0.01, 0.001, 0.0001	1
เครื่องบิน	C	0.1, 1, 100, 1000, 3000, 10000	10000
	kernel	'rbf', 'poly', 'sigmoid', 'linear'	'poly'
	degree	1, 2, 3, 4, 5	4
	gamma	1, 0.1, 0.01, 0.001, 0.0001	1



รูปที่ 101 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR ของน้ำมันดีเซล



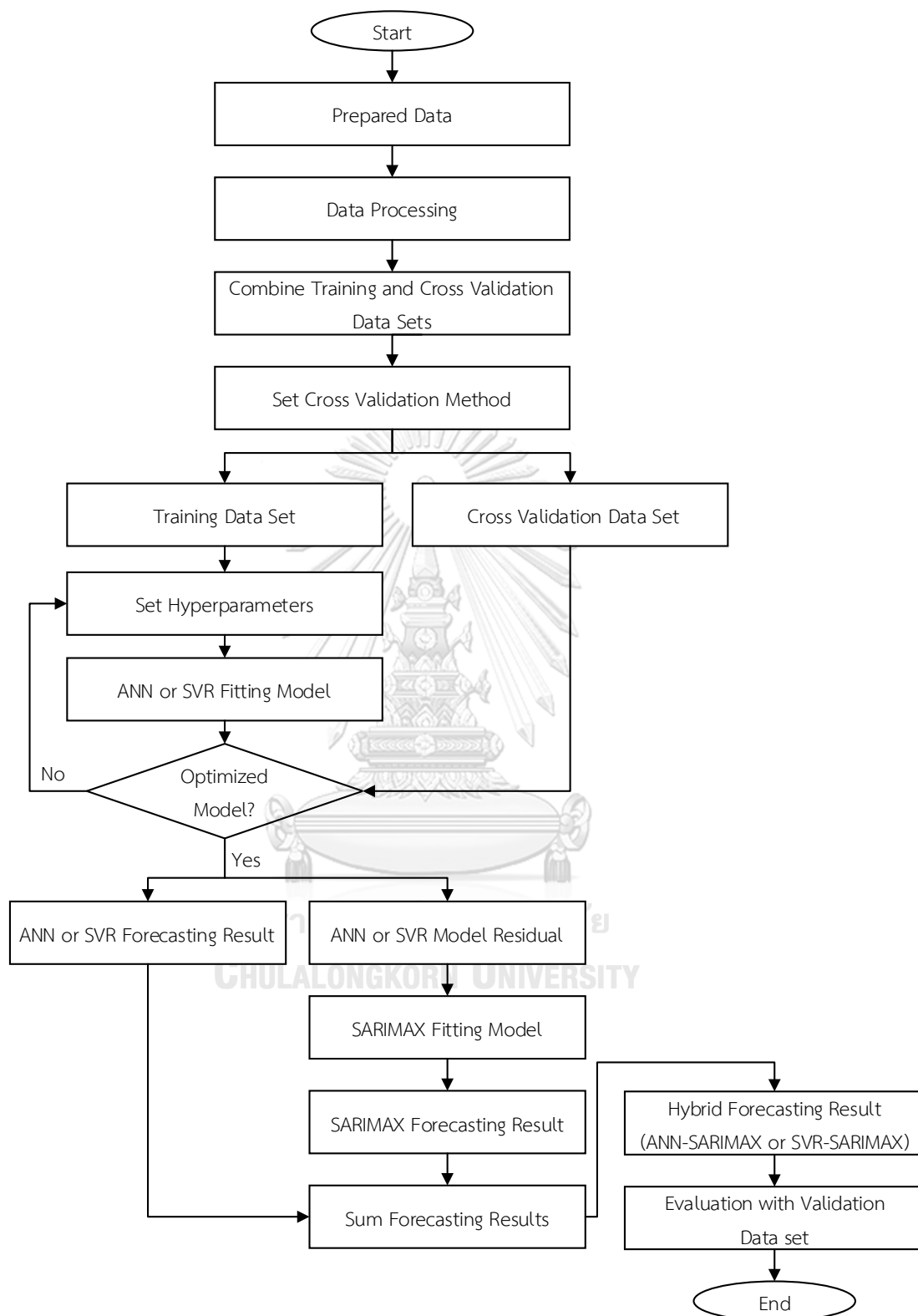
รูปที่ 102 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR ของน้ำมันเบนซิน



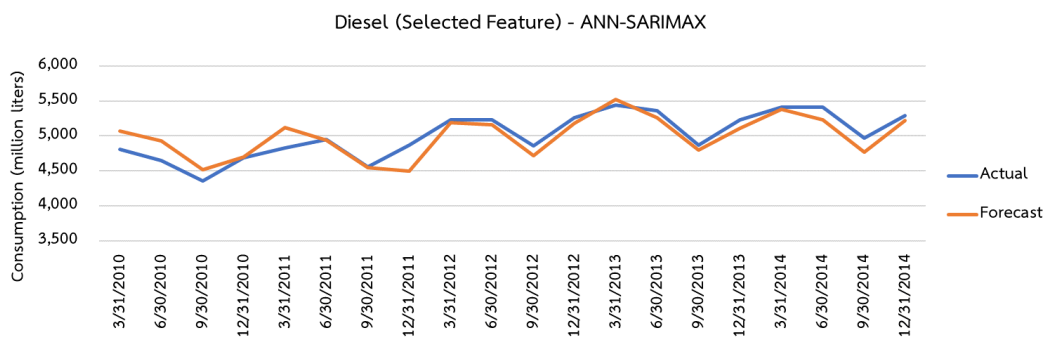
รูปที่ 103 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR ของน้ำมันเครื่องบิน

#### 4.12.3 ตัวแบบ ANN-SARIMAX

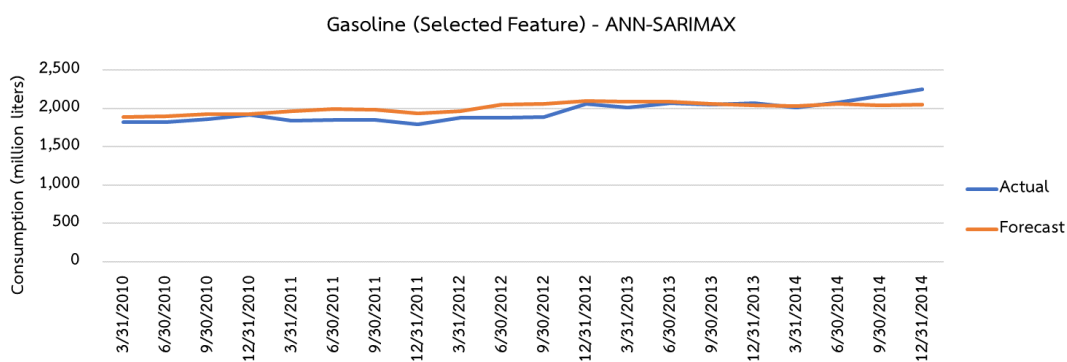
ตัวแบบ ANN-SARIMAX ของน้ำมันดีเซลเป็นตัวแบบที่มีโครงสร้างการผสมแบบอนุกรม มีกระบวนการสร้างตัวแบบดังแสดงในรูปที่ 104 ซึ่งนำตัวแบบ ANN มาพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันไม่เชิงเส้นก่อน ซึ่งได้ผลลัพธ์ดังแสดงในหัวข้อที่ 4.8 จากนั้นจึงนำค่าส่วนเหลือที่ได้จากตัวแบบ ANN ไปพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นด้วยตัวแบบ SARIMAX ที่ใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอนและใช้ค่าส่วนเหลือของตัวแบบ ANN เป็นตัวแปรตาม ซึ่งตัวแบบของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินได้ผลลัพธ์เป็นตัวแบบ SARIMAX(0,1,0)(0,1,1)<sub>4</sub> SARIMAX(0,1,0)(1,1,0)<sub>4</sub> SARIMAX(0,1,0)(0,1,1)<sub>4</sub> ตามลำดับ ตัวแบบ SARIMAX ที่ได้ใช้ในการจับข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้น จากนั้นจึงรวมผลการพยากรณ์ในส่วนที่เป็นฟังก์ชันเชิงเส้นและฟังก์ชันไม่เชิงเส้นเข้าด้วยกัน ตัวแบบ ANN-SARIMAX ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินใช้เวลาในการฝึกสอนตัวแบบทั้งหมด 36.3 นาที 35.4 นาที และ 33.3 นาที ตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 105 106 และ 107



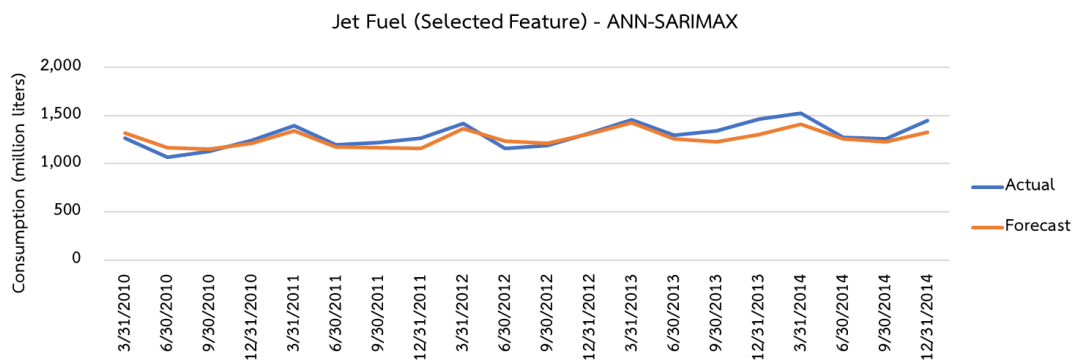
รูปที่ 104 กระบวนการสร้างตัวแบบ ANN-SARIMAX และตัวแบบ SVR-SARIMAX



รูปที่ 105 การพยากรณ์ด้วยตัวแบบ ANN-SARIMAX ของน้ำมันดีเซล



รูปที่ 106 การพยากรณ์ด้วยตัวแบบ ANN-SARIMAX ของน้ำมันเบนซิน

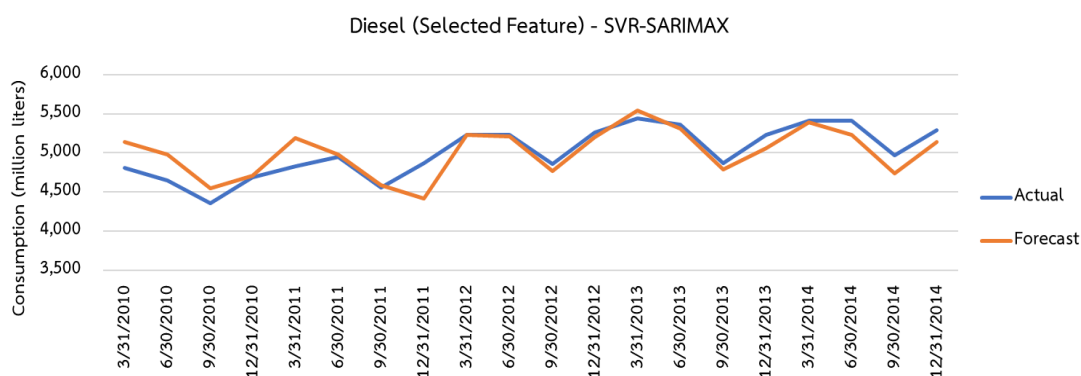


รูปที่ 107 การพยากรณ์ด้วยตัวแบบ ANN-SARIMAX ของน้ำมันเครื่องบิน

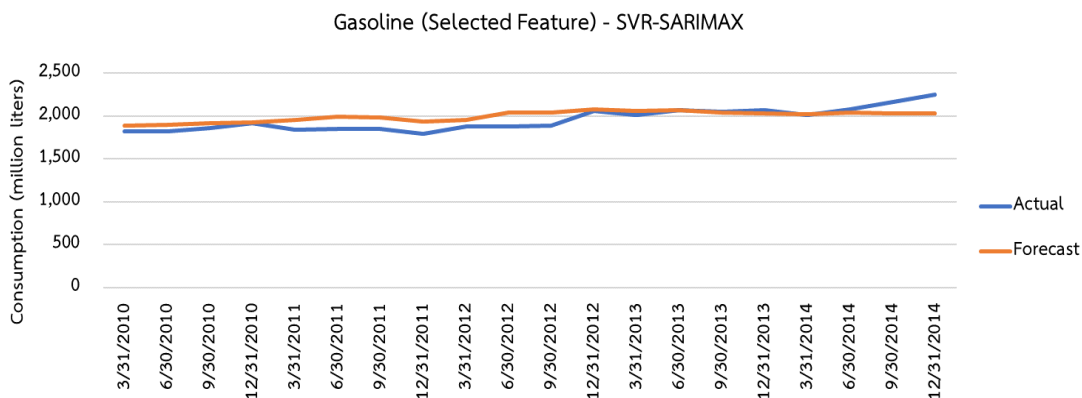


#### 4.12.4 ตัวแบบ SVR-SARIMAX

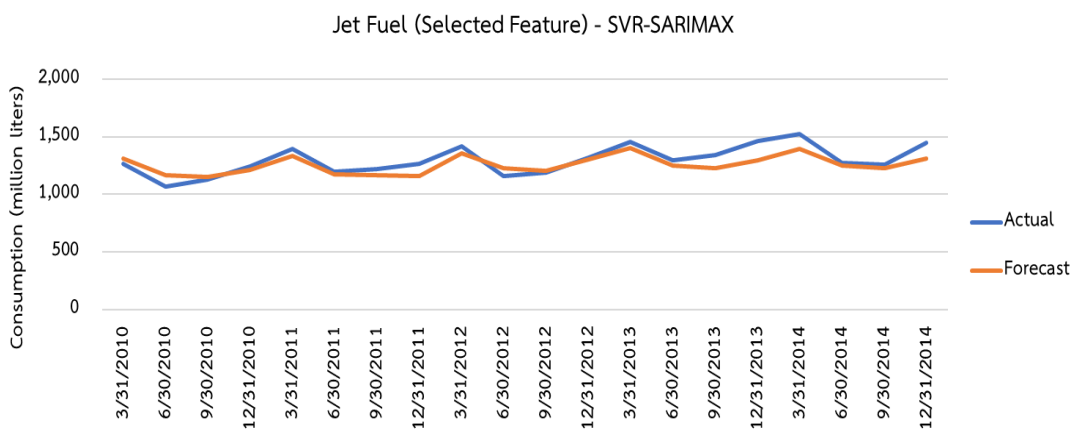
ตัวแบบ SVR-SARIMAX ของน้ำมันดีเซลเป็นตัวแบบที่มีโครงสร้างการผสมแบบอนุกรม มีกระบวนการสร้างตัวแบบดังแสดงในรูปที่ 104 ซึ่งนำตัวแบบ SVR มาพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันไม่เชิงเส้นก่อน ซึ่งได้ผลลัพธ์ดังแสดงในหัวข้อที่ 4.7 จากนั้นจึงนำค่าส่วนเหลือที่ได้จากตัวแบบ SVR ไปพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นด้วยตัวแบบ SARIMAX ที่ใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอนและใช้ค่าส่วนเหลือของตัวแบบ SVR เป็นตัวแปรตาม ซึ่งตัวแบบของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินได้ผลลัพธ์เป็นตัวแบบ SARIMAX(0,1,0)(0,1,1)<sub>4</sub> SARIMAX(0,1,0)(1,1,0)<sub>4</sub> และ SARIMAX(0,1,0)(0,1,1)<sub>4</sub> ตามลำดับ ตัวแบบ SARIMAX ใช้ในการจับข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้น จากนั้นจึงรวมผลการพยากรณ์ในส่วนที่เป็นฟังก์ชันเชิงเส้นและฟังก์ชันไม่เชิงเส้นเข้าด้วยกัน ตัวแบบ SVR-SARIMAX ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินใช้เวลาในการฝึกสอนตัวแบบทั้งหมด 11.5 วินาที 9.0 วินาที และ 4.5 วินาที ตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 108 109 และ 110



รูปที่ 108 การพยากรณ์ด้วยตัวแบบ SVR-SARIMAX ของน้ำมันดีเซล



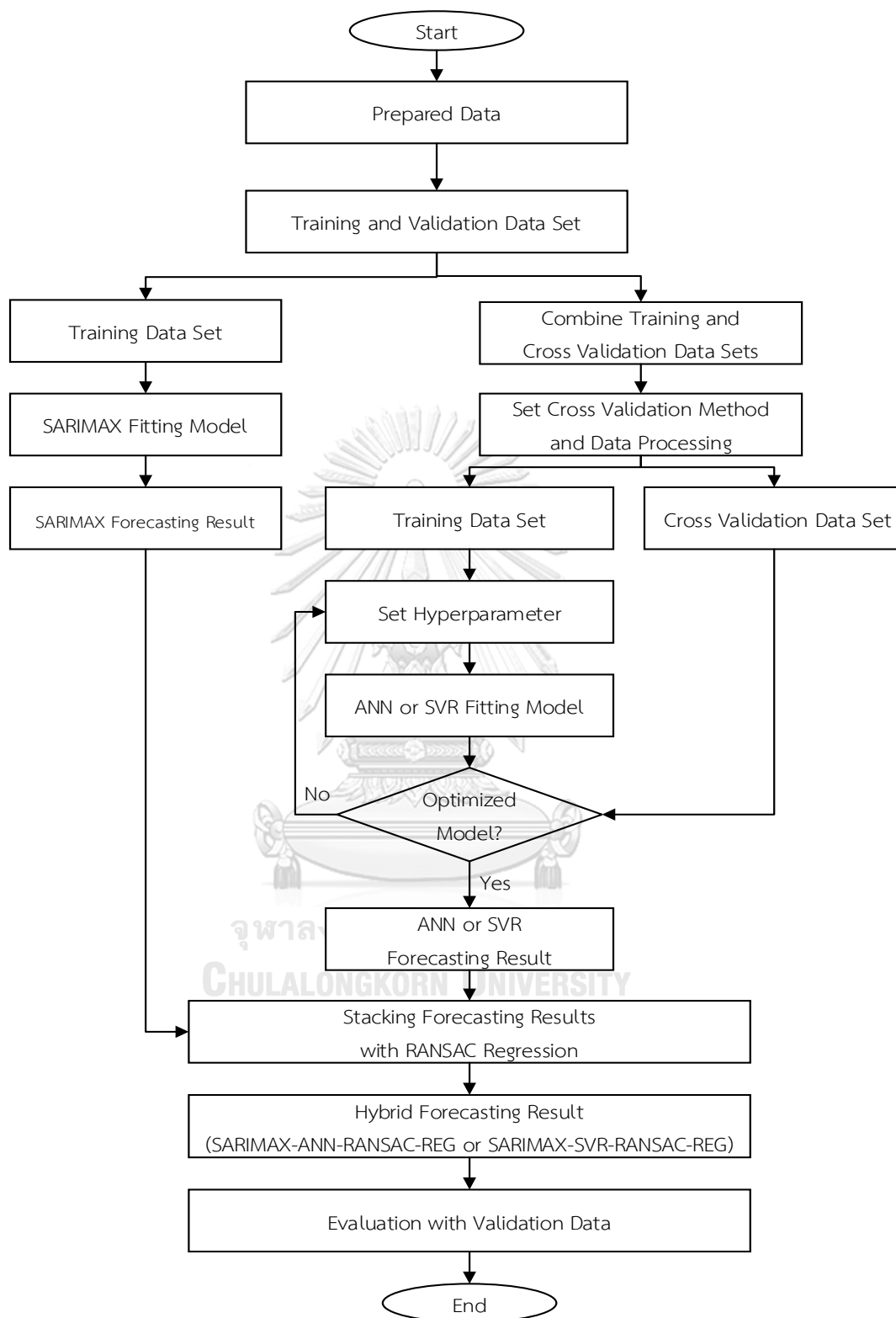
รูปที่ 109 การพยากรณ์ด้วยตัวแบบ SVR-SARIMAX ของน้ำมันเบนซิน



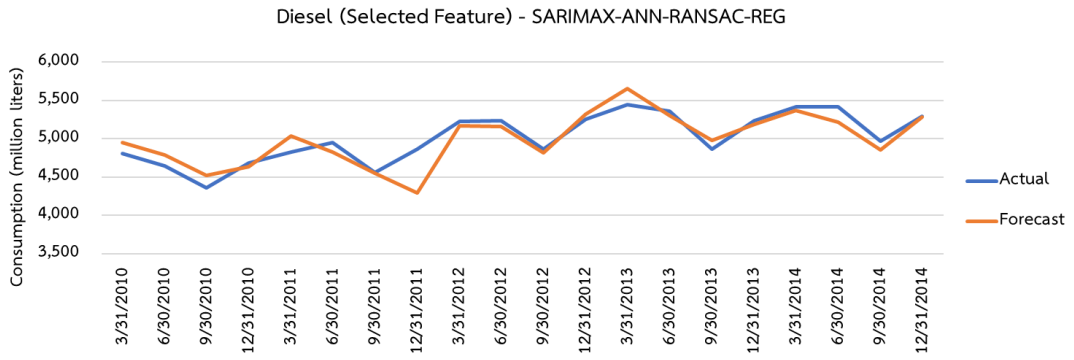
รูปที่ 110 การพยากรณ์ด้วยตัวแบบ SVR-SARIMAX ของน้ำมันเครื่องบิน

#### 4.12.5 ตัวแบบ SARIMAX-ANN-RANSAC-REG

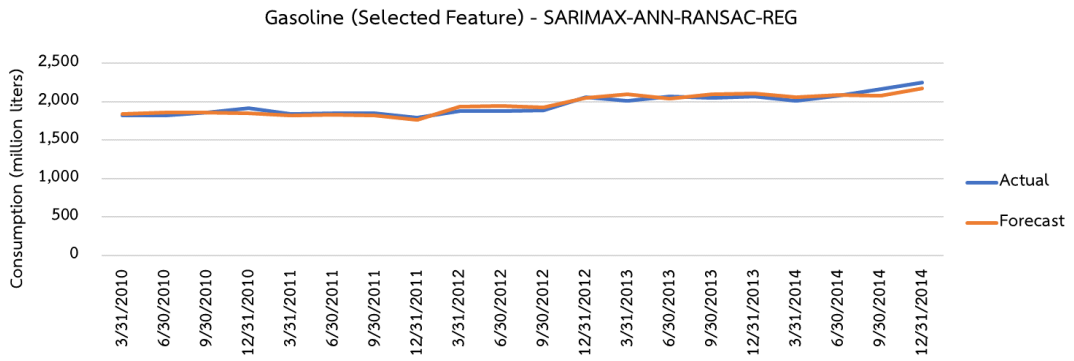
ตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันดีเซลเป็นตัวแบบที่มีโครงสร้างการผสมแบบขนาน มีกระบวนการสร้างตัวแบบดังแสดงในรูปที่ 111 ซึ่งนำตัวแบบเดี่ยว SARIMAX ซึ่งได้ผลลัพธ์ดังแสดงในหัวข้อที่ 4.5 และตัวแบบเดี่ยว ANN ซึ่งได้ผลลัพธ์ดังแสดงในหัวข้อที่ 4.8 จากนั้นจึงรวมผลการพยากรณ์ที่ได้จากตัวแบบทั้งสองด้วยฟังก์ชัน RANSAC Regression หรือที่เรียกว่า Stacking Model with RANSAC Regression สาเหตุที่ผู้วิจัยเลือกใช้ฟังก์ชันการรวมผลการพยากรณ์ด้วยฟังก์ชัน RANSAC Regression แทนฟังก์ชันการถดถอยเชิงเส้นพหุคูณนั้น เนื่องจากในการพยากรณ์ทั้งระยะกลางและระยะสั้นของน้ำมันสำเร็จรูปทั้ง 3 ชนิด ตัวแบบ RANSAC Regression ให้ผลลัพธ์การพยากรณ์ที่แม่นยำกว่าตัวแบบการถดถอยเชิงเส้นพหุคูณทั้งหมด โดยตัวแบบ RANSAC Regression มีความแกร่งต่อความผิดปกติ (Outlier) ของข้อมูลป้อนเข้า ผู้วิจัยจึงคาดว่าตัวแบบผสมที่ใช้ฟังก์ชัน RANSAC Regression รวมผลการพยากรณ์จะมีความแกร่งต่อความผิดปกติ (Outlier) ของข้อมูลป้อนเข้า อาจส่งผลให้ตัวแบบผสมมีความแม่นยำเพิ่มขึ้นและมีความแกร่ง (Robust) เพิ่มขึ้น ซึ่งตัวแบบผสมทุกตัวแบบที่มีการรวมผลการพยากรณ์ของตัวแบบอื่นด้วยฟังก์ชัน RANSAC Regression จะใช้ข้อมูลปริมาณน้ำมันสำเร็จรูปในช่วงตรวจสอบเป็นตัวแปรตามในการประมาณค่าพารามิเตอร์ของฟังก์ชัน RANSAC Regression ภายใต้สมมติฐานคือมีข้อมูลผลเฉลยของชุดข้อมูลตรวจสอบ (Validation Data Set) แล้ว ตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินใช้เวลาในการฝึกสอนตัวแบบทั้งหมด 36.4 นาที 35.5 นาที และ 33.4 นาทีตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 112 113 และ 114



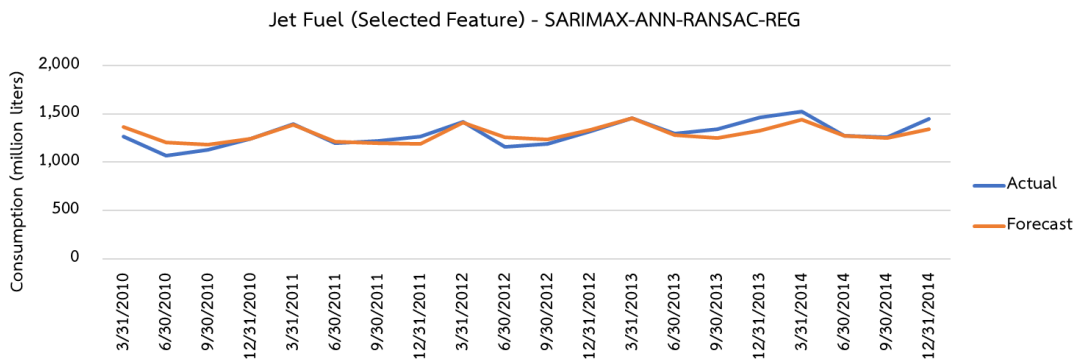
รูปที่ 111 กระบวนการสร้างตัวแบบ SARIMAX-ANN- RANSAC-REG และตัวแบบ SARIMAX-SVR- RANSAC-REG



รูปที่ 112 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันดีเซล



รูปที่ 113 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันเบนซิน



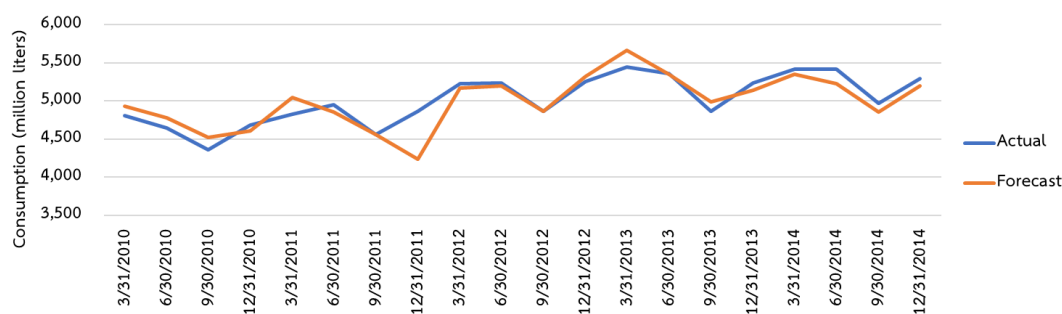
รูปที่ 114 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันเครื่องบิน

#### 4.12.6 ตัวแบบ SARIMAX-SVR-RANSAC-REG

ตัวแบบ SARIMAX-SVR-RANSAC-REG ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน เป็นตัวแบบที่มีโครงสร้างการผสมแบบขนาน มีกระบวนการสร้างตัวแบบดังแสดงในรูปที่ 111 ซึ่งนำตัวแบบเดี่ยว SARIMAX ซึ่งได้ผลลัพธ์ดังแสดงในหัวข้อที่ 4.5 และตัวแบบเดี่ยว SVR ซึ่งได้ผลลัพธ์ดังแสดงในหัวข้อที่ 4.7 จากนั้นจึงรวมผลการพยากรณ์ที่ได้จากตัวแบบทั้งสองด้วยฟังก์ชันการถดถอยเชิงเส้น หรือที่เรียกว่า Stacking Model with RANSAC Regression ตัวแบบ SARIMAX-SVR-RANSAC-REG ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ใช้เวลาในการฝึกสอนตัวแบบทั้งหมด 11.6 วินาที 9.1 วินาที และ 4.6 วินาที ตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 115 116 และ 117



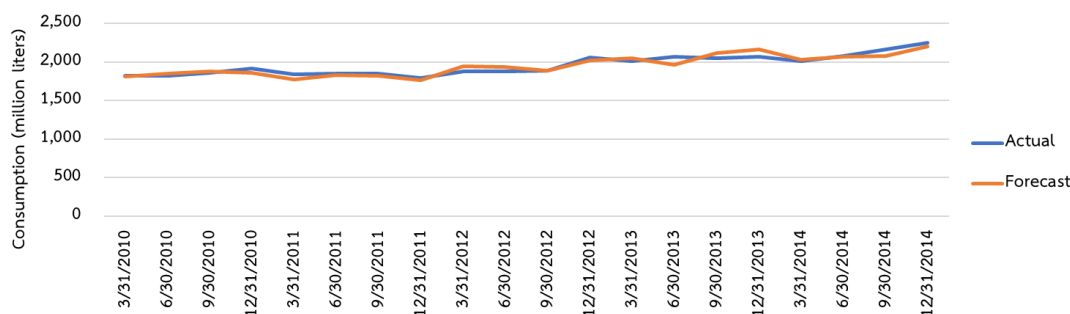
Diesel (Selected Feature) - SARIMAX-SVR-RANSAC-REG



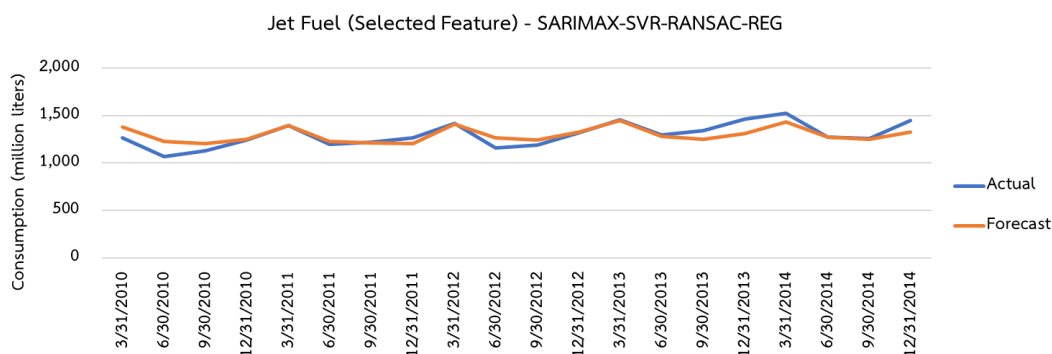
รูปที่ 115 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR-RANSAC-REG ของน้ำมันดีเซล

จุฬาลงกรณ์มหาวิทยาลัย

Gasoline (Selected Feature) - SARIMAX-SVR-RANSAC-REG



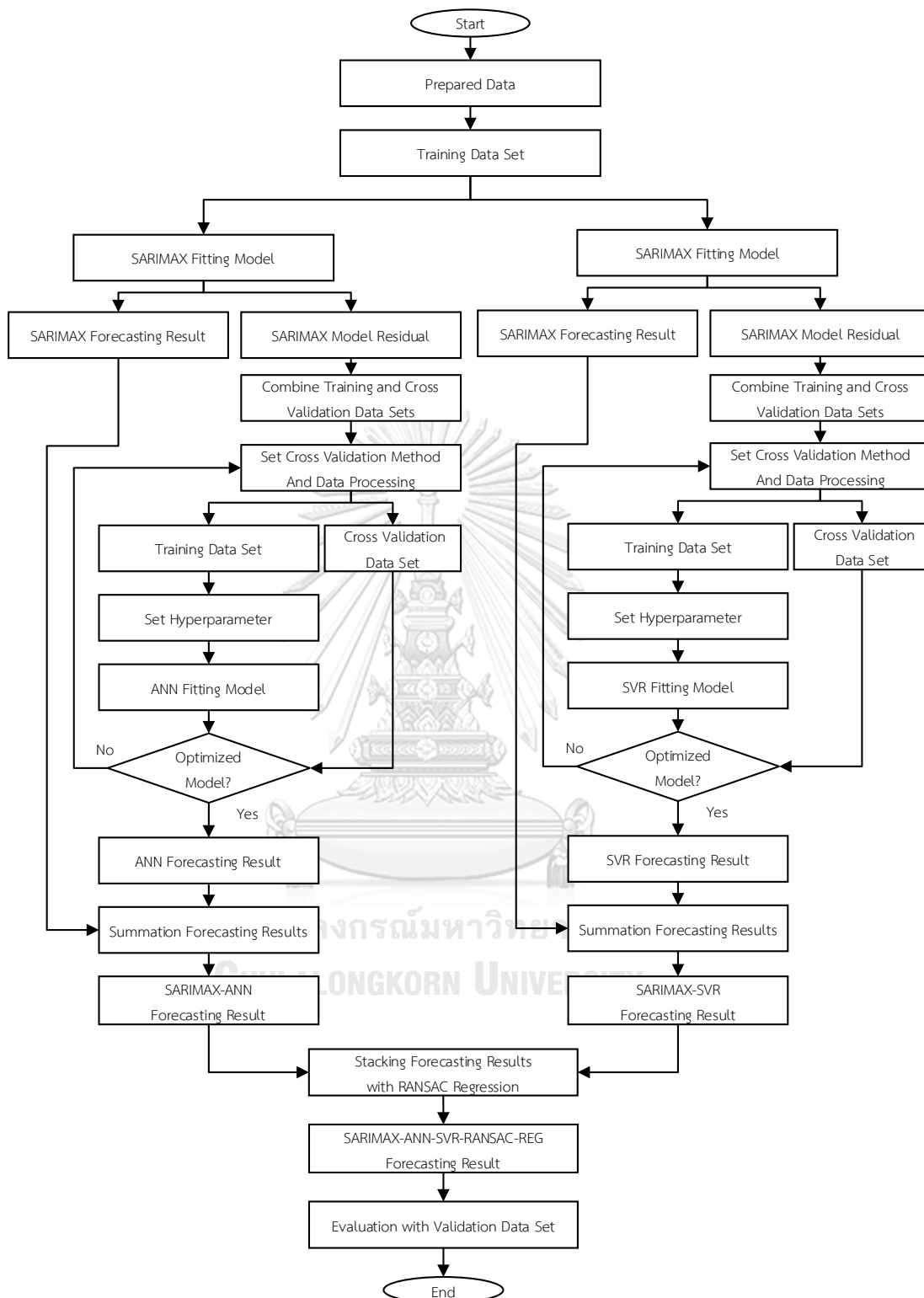
รูปที่ 116 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR-RANSAC-REG ของน้ำมันเบนซิน



รูปที่ 117 การพยากรณ์ด้วยตัวแบบ SARIMAX-SVR-RANSAC-REG ของน้ำมันเครื่องบิน

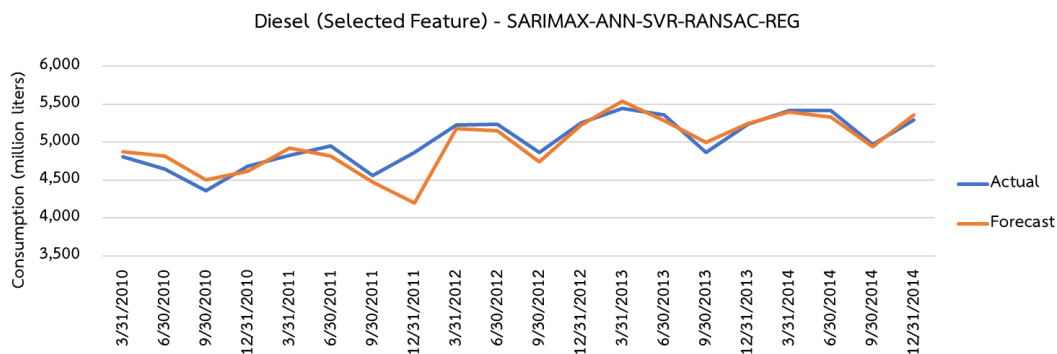
#### 4.12.7 ตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG

ตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG เป็นตัวแบบที่เกิดจากการนำผลการพยากรณ์ของตัวแบบ SARIMAX-ANN ที่ได้จากหัวข้อ 4.12.1 และของตัวแบบ SARIMAX-SVR ที่ได้จากหัวข้อ 4.12.2 มารวมผลการพยากรณ์ทั้งสองด้วยวิธีฟังก์ชันการถดถอยเชิงเส้น มีกระบวนการสร้างตัวแบบดังแสดงในรูปที่ 118 ผู้วิจัยเลือกผสมตัวแบบ SARIMAX-ANN กับตัวแบบ SARIMAX-SVR ของน้ำมันสำเร็จรูปทั้ง 3 ชนิด ซึ่งจากผลการพยากรณ์ของผู้วิจัยส่วนใหญ่มีค่าความแม่นยำดีกว่าการผสมตัวแบบระหว่างตัวแบบ ANN-SARIMAX กับตัวแบบ SVR-SARIMAX และตัวแบบ SARIMAX-ANN กับตัวแบบ SARIMAX-SVR เป็นที่นิยมในการใช้พยากรณ์มากกว่า โดยตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ใช้เวลาในการฝึกสอนตัวแบบทั้งหมด 15.8 นาที 13.1 นาที และ 14.2 นาที ตามลำดับ เมื่อทำการพยากรณ์ไปล่วงหน้า 20 ไตรมาสหรือจำนวน 5 ปี ตั้งแต่ พ.ศ. 2553 ถึง พ.ศ. 2557 (ค.ศ. 2010 ถึง ค.ศ. 2014) ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 119 120 และ 121

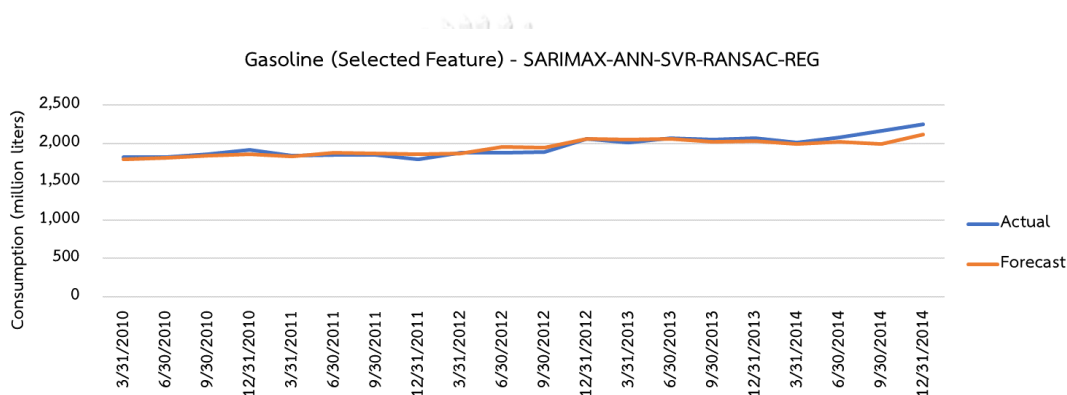


รูปที่ 118 กระบวนการสร้างตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG

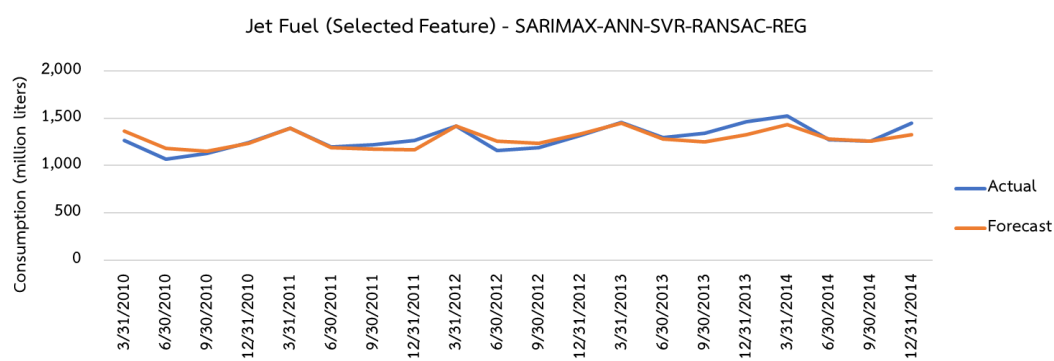




รูปที่ 119 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันดีเซล



รูปที่ 120 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันเบนซิน



รูปที่ 121 การพยากรณ์ด้วยตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันเครื่องบิน

#### 4.13 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ในกลุ่มตัวแบบผสม

เมื่อนำชุดข้อมูลสำหรับตรวจสอบ (Cross Validation Data) กับผลการพยากรณ์ในช่วงเวลา ปี พ.ศ. 2553 ถึงปี พ.ศ. 2557 สำหรับการพยากรณ์กลุ่มตัวแบบผสมแบบใช้ตัวแปรอิสระที่คัดเลือก ด้วยเทคนิคการถดถอยแบบขั้นตอนมาทำการวัดประสิทธิภาพของตัวแบบด้วย MAPE ได้ผลลัพธ์ดัง ตารางที่ 22 23 และ 24 ซึ่งผลลัพธ์การวัดประสิทธิภาพตัวแบบผสมของน้ำมันดีเซลพบว่าตัวแบบที่ แม่นยำที่สุดคือตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG มีค่า MAPE เท่ากับ 2.2785% ตัวแบบ นี้สามารถเพิ่มความแม่นยำจากการพยากรณ์ด้วยตัวแบบเดี่ยว SVR และ ANN โดยมีความแม่นยำ เพิ่มขึ้น 32.96% และ 31.71% ตามลำดับ ในส่วนของตัวแบบผสมของน้ำมันเบนซินพบว่าตัวแบบที่ แม่นยำที่สุดคือตัวแบบ SARIMAX-ANN-RANSAC-REG มีค่า MAPE เท่ากับ 1.9966% ตัวแบบนี้ สามารถเพิ่มความแม่นยำจากการพยากรณ์ด้วยตัวแบบเดี่ยว SVR และ ANN โดยมีความแม่นยำ เพิ่มขึ้น 32.03% และ 3.06% ตามลำดับ สำหรับตัวแบบผสมของน้ำมันเครื่องบินพบว่าตัวแบบที่ แม่นยำที่สุดคือตัวแบบ SARIMAX-SVR มีค่า MAPE เท่ากับ 3.5505% ตัวแบบนี้สามารถเพิ่มความ แม่นยำจากการพยากรณ์ด้วยตัวแบบเดี่ยว SVR และ ANN โดยมีความแม่นยำเพิ่มขึ้น 30.45% และ 23.58% ตามลำดับ เมื่อพิจารณาจากผลลัพธ์พบว่าตัวแบบผสมสามารถช่วยเพิ่มความแม่นยำให้กับ ตัวแบบเดี่ยวได้เป็นส่วนใหญ่ โดยเฉพาะตัวแบบเดี่ยวที่มีการนำผลการพยากรณ์ของตัวแบบจำนวน 2 ตัวแบบมารวมกันด้วยวิธี RANSAC-REG นั้นให้ผลการพยากรณ์ที่มีความแม่นยำเพิ่มขึ้นทุกตัวแบบ

ตารางที่ 22 ค่า MAPE ของแต่ละตัวแบบการเรียนรู้ของเครื่องสำหรับการตรวจสอบตัวแบบของน้ำมัน ดีเซล ในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557

ชนิดน้ำมัน	ตัวแบบ	MAPE (%) เฉลี่ย 5 ปี (20 ไตรมาส)
ดีเซล	SARIMAX-ANN	2.7680
	SARIMAX-SVR	2.7076
	ANN-SARIMAX	2.6240
	SVR-SARIMAX	2.9990
	SARIMAX-ANN-RANSAC-REG	2.4755
	SARIMAX-SVR-RANSAC-REG	2.5553
	SARIMAX-ANN-SVR-RANSAC-REG	2.2785

ตารางที่ 23 ค่า MAPE ของแต่ละตัวแบบการเรียนรู้ของเครื่องสำหรับการตรวจสอบตัวแบบของน้ำมันเบนซิน ในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557

ชนิดน้ำมัน	ตัวแบบ	MAPE (%) เฉลี่ย 5 ปี (20 ไตรมาส)
เบนซิน	SARIMAX-ANN	2.9528
	SARIMAX-SVR	3.2112
	ANN-SARIMAX	4.4433
	SVR-SARIMAX	4.2789
	SARIMAX-ANN-RANSAC-REG	1.9966
	SARIMAX-SVR-RANSAC-REG	2.1358
	SARIMAX-ANN-SVR-RANSAC-REG	2.2323

ตารางที่ 24 ค่า MAPE ของแต่ละตัวแบบการเรียนรู้ของเครื่องสำหรับการตรวจสอบตัวแบบของน้ำมันเครื่องบิน ในช่วง พ.ศ. พ.ศ. 2553 ถึงปี พ.ศ. 2557

ชนิดน้ำมัน	ตัวแบบ	MAPE (%) เฉลี่ย 5 ปี (20 ไตรมาส)
เครื่องบิน	SARIMAX-ANN	3.8487
	SARIMAX-SVR	3.5055
	ANN-SARIMAX	4.6279
	SVR-SARIMAX	4.9074
	SARIMAX-ANN-RANSAC-REG	4.0524
	SARIMAX-SVR-RANSAC-REG	4.3920
	SARIMAX-ANN-SVR-RANSAC-REG	3.9539

#### 4.14 การคัดเลือกตัวแบบการพยากรณ์

ในการคัดเลือกตัวแบบการพยากรณ์ที่เหมาะสม ผู้วิจัยใช้วิธีการคัดเลือกตัวแบบจากการเปรียบเทียบผลลัพธ์ของตัวแบบการพยากรณ์ โดยจะเลือกจากตัวแบบที่ให้ค่าความแม่นยำมากที่สุด เมื่อนำแต่ละตัวแบบไปประเมินผลความแม่นยำของตัวแบบกับชุดข้อมูลตรวจสอบจะได้ค่า MAPE ของการตรวจสอบ (Validation MAPE) เมื่อพิจารณาตารางที่ 25 26 และ 27 จะเห็นว่าตัวแบบ ADA ให้ความแม่นยำมากที่สุดกับการพยากรณ์ระยะกลางของน้ำมันสำเร็จรูปทั้ง 3 ชนิด แต่เมื่อพิจารณาผลลัพธ์ของตัวแบบในหัวข้อที่ 4.9 พบว่าตัวแบบเกิดปัญหาความเข้ากันเกินไปของข้อมูล ดังนั้นผู้วิจัยจึงไม่แนะนำตัวแบบ ADA สำหรับการพยากรณ์น้ำมันสำเร็จรูปทั้ง 3 ชนิด โดยจะพิจารณาเลือกใช้ตัวแบบที่มีความแม่นยำสูงในลำดับถัดไป สำหรับตัวแบบการพยากรณ์ของปริมาณการบริโภคน้ำมันดีเซล ผู้วิจัยเลือกใช้ตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG มีค่า Validation MAPE เท่ากับ 2.2324% จากนั้นจึงนำไปประเมินผลความแม่นยำของตัวแบบกับชุดข้อมูลทดสอบ ได้ค่า MAPE ของการทดสอบ (Testing MAPE) เท่ากับ 5.4835% สำหรับตัวแบบการพยากรณ์ของปริมาณการบริโภคน้ำมันเบนซิน ผู้วิจัยเลือกใช้ตัวแบบ SARIMAX-ANN-RANSAC-REG มีค่า Validation MAPE เท่ากับ 1.9966% และมีค่า Testing MAPE เท่ากับ 6.3873% สำหรับตัวแบบการพยากรณ์ของปริมาณการบริโภคน้ำมันเครื่องบิน ผู้วิจัยเลือกใช้ตัวแบบ SARIMAX-SVR มีค่า Validation MAPE เท่ากับ 3.5055% และมีค่า Testing MAPE เท่ากับ 7.2962% ค่า Validation MAPE และ Testing MAPE ของแต่ละตัวแบบที่ถูกคัดเลือกแสดงในตารางที่ 28

จะเห็นได้ว่าตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันดีเซล ตัวแบบ SARIMAX-ANN-RANSAC-REG ของน้ำมันเบนซิน และตัวแบบ SARIMAX-SVR ของน้ำมันเครื่องบิน ที่เลือกใช้เป็นตัวแบบพยากรณ์มีความแม่นยำสูงและเมื่อพิจารณาความแตกต่างระหว่างค่า Validation MAPE กับค่า Testing MAPE พบว่า Testing MAPE สูงกว่า Validation MAPE ประมาณ 2-3 เท่า และ Testing MAPE มีค่าไม่เกิน 10% ตัวแบบเหล่านี้จึงมีปัญหาค่าความเข้ากันเกินไปของข้อมูลไม่มากนัก จึงมีความเหมาะสมที่จะนำไปใช้พยากรณ์น้ำมันสำเร็จรูปทั้งแบบการพยากรณ์ระยะกลางและยังให้ผลการพยากรณ์ที่เหมาะสมกับการพยากรณ์ระยะสั้นด้วยเนื่องจากผลการพยากรณ์ระยะกลางครอบคลุมการพยากรณ์ระยะสั้นและมีค่าความแม่นยำสูง

ตารางที่ 25 ค่า Validation MAPE ของแต่ละตัวแบบการพยากรณ์ระยะกลางของน้ำมันดีเซล

กลุ่มตัวแบบ	ตัวแบบ	MAPE (%)
ตัวแบบอนุกรมเวลาเชิงสถิติ	Holt-Winters	4.4917
	SARIMA	5.5853
	<b>SARIMAX</b>	<b>3.4629</b>
ตัวแบบพยากรณ์เชิงสาเหตุ	MLR	6.2489
	<b>RANSAC-REG</b>	<b>4.8850</b>
ตัวแบบการเรียนรู้ของเครื่อง	KNN	2.6534
	SVR	3.3988
	ANN	3.3364
	<b>ADA</b>	<b>2.2348</b>
ตัวแบบผสม	SARIMAX-ANN	2.7680
	SARIMAX-SVR	2.7076
	ANN-SARIMAX	2.6240
	SVR-SARIMAX	2.9990
	SARIMAX-ANN-RANSAC-REG	2.4755
	SARIMAX-SVR-RANSAC-REG	2.5553
	<b>SARIMAX-ANN-SVR-RANSAC-REG</b>	<b>2.2785</b>

ตารางที่ 26 ค่า Validation MAPE ของแต่ละตัวแบบการพยากรณ์ระยะกลางของน้ำมันเบนซิน

กลุ่มตัวแบบ	ตัวแบบ	MAPE (%)
ตัวแบบอนุกรมเวลาเชิงสถิติ	Holt-Winters	4.9547
	SARIMA	7.1490
	<b>SARIMAX</b>	<b>3.5834</b>
ตัวแบบพยากรณ์เชิงสาเหตุ	MLR	3.2701
	<b>RANSAC-REG</b>	<b>2.8895</b>
ตัวแบบการเรียนรู้ของเครื่อง	KNN	3.0597
	SVR	2.9376
	ANN	2.0596
	<b>ADA</b>	<b>1.6357</b>
ตัวแบบผสม	SARIMAX-ANN	2.9528
	SARIMAX-SVR	3.2112
	ANN-SARIMAX	4.4433
	SVR-SARIMAX	4.2789
	<b>SARIMAX-ANN-RANSAC-REG</b>	<b>1.9966</b>
	SARIMAX-SVR-RANSAC-REG	2.1358
	SARIMAX-ANN-SVR-RANSAC-REG	2.2323

ตารางที่ 27 ค่า Validation MAPE ของแต่ละตัวแบบการพยากรณ์ระยะกลางของน้ำมันเครื่องบิน

กลุ่มตัวแบบ	ตัวแบบ	MAPE (%)
ตัวแบบอนุกรมเวลาเชิงสถิติ	Holt-Winters	5.4177
	SARIMA	5.5589
	<b>SARIMAX</b>	<b>4.7452</b>
ตัวแบบพยากรณ์เชิงสาเหตุ	MLR	4.9555
	<b>RANSAC-REG</b>	<b>4.7085</b>
ตัวแบบการเรียนรู้ของเครื่อง	KNN	3.5516
	SVR	5.0404
	ANN	4.5870
	<b>ADA</b>	<b>3.5401</b>
ตัวแบบผสม	SARIMAX-ANN	3.8487
	<b>SARIMAX-SVR</b>	<b>3.5055</b>
	ANN-SARIMAX	4.6279
	SVR-SARIMAX	4.9074
	SARIMAX-ANN-RANSAC-REG	4.0524
	SARIMAX-SVR-RANSAC-REG	4.3920
	SARIMAX-ANN-SVR-RANSAC-REG	3.9539

## จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 28 ค่า Validation MAPE และ Testing MAPE ของตัวแบบที่ถูกคัดเลือก

ชนิดน้ำมัน	ตัวแบบที่ถูกคัดเลือก	Validation MAPE (%)	Testing MAPE (%)
ดีเซล	SARIMAX-ANN-SVR-RANSAC-REG	2.2785	5.4835
เบนซิน	SARIMAX-ANN-RANSAC-REG	1.9966	6.3873
เครื่องบิน	SARIMAX-SVR	3.5055	7.2962

## บทที่ 5

### สรุปผลการดำเนินการและข้อเสนอแนะ

การศึกษากาการพยากรณ์ปริมาณน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยด้วยเทคนิคการเรียนรู้ของเครื่องมีวัตถุประสงค์เพื่อศึกษาวิธีการพยากรณ์และเลือกตัวแบบที่เหมาะสมสำหรับการพยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยโดยการเปรียบเทียบผลการพยากรณ์ของตัวแบบอนุกรมเวลาเชิงสถิติ ตัวแบบการพยากรณ์เชิงสาเหตุ ตัวแบบการเรียนรู้ของเครื่อง และตัวแบบผสม โดยน้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยที่ศึกษาในงานวิจัยนี้ได้แก่น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ผู้วิจัยได้ศึกษาตัวแบบการพยากรณ์โดยแบ่งออกเป็น 4 กลุ่ม ประกอบไปด้วยตัวแบบอนุกรมเวลาเชิงสถิติ 3 ตัวแบบ ตัวแบบการพยากรณ์เชิงสาเหตุ 2 ตัวแบบ ตัวแบบการเรียนรู้ของเครื่อง 4 ตัวแบบ และตัวแบบผสม 7 ตัวแบบ รวมทั้งหมด 16 ตัวแบบ ซึ่งแต่ละตัวแบบจะถูกนำไปสร้างตัวแบบการพยากรณ์ระยะกลางของปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ส่วนการพยากรณ์ระยะสั้นโดยใช้วิธีการเลื่อนตำแหน่งช่วงเวลาจะใช้ตัวแบบอนุกรมเวลาเชิงสถิติ 3 ตัวแบบ และตัวแบบการพยากรณ์เชิงสาเหตุ 2 ตัวแบบ จากนั้นจึงนำผลการพยากรณ์ที่ได้ไปประเมินประสิทธิภาพของตัวแบบด้วย MAPE โดยสามารถสรุปผลวิจัยได้ดังนี้

#### 5.1 สรุปผลการดำเนินการ

##### 1) ตัวแบบการพยากรณ์เชิงสาเหตุ

จากการสร้างตัวแบบการพยากรณ์เชิงสาเหตุสำหรับการพยากรณ์ระยะกลางด้วยวิธีการถดถอยเชิงเส้นพหุคูณโดยใช้เทคนิคการถดถอยแบบขั้นตอนในการคัดเลือกตัวแปรอิสระที่เกี่ยวข้องกับปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน ผลลัพธ์คือตัวแปรอิสระที่แตกต่างกันไปขึ้นอยู่กับชนิดน้ำมันสำเร็จรูป โดยตัวแปรอิสระสำหรับน้ำมันดีเซลได้แก่ ปริมาณการผลิตน้ำมันเบนซิน ราคาน้ำมันดีเซล ผลผลิตทั้งหมดรวม จำนวนรถใหม่ที่จะจดทะเบียน และจำนวนประชากร ตัวแปรอิสระสำหรับน้ำมันเบนซินได้แก่ ปริมาณการผลิตน้ำมันดีเซล ปริมาณการผลิตน้ำมันเบนซิน ปริมาณการผลิตก๊าซปิโตรเลียมเหลว ราคาน้ำมันดีเซล ผลผลิตทั้งหมดรวม จำนวนรถใหม่ที่จะจดทะเบียน และจำนวนประชากร และตัวแปรอิสระสำหรับน้ำมันเครื่องบินคือผลผลิตทั้งหมดรวม ดังแสดงในสมการที่ 4.1 4.2 และ 4.3 สำหรับผลลัพธ์ความสัมพันธ์ระหว่างตัวแปรอิสระที่ถูกคัดเลือกเหล่านี้กับปริมาณการบริโภคน้ำมันสำเร็จรูปทั้ง 3 ชนิดมีทั้งตัวแปรอิสระที่ผู้วิจัยเห็นด้วยและไม่เห็นด้วย ส่วนที่ผู้วิจัยเห็นด้วยกับตัวแปรอิสระที่ถูกคัดเลือกของน้ำมันดีเซลมีดังนี้ 1) ราคาน้ำมันดีเซลที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันดีเซลในทิศทางตรงข้าม อาจเป็นเพราะเมื่อราคาน้ำมันดีเซลสูงขึ้น ผู้บริโภคอาจเปลี่ยนไปบริโภคน้ำมันสำเร็จรูปชนิดอื่นทดแทน 2) ผลผลิตทั้งหมดรวมและจำนวนรถจดทะเบียนใหม่ซึ่งมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันดีเซลในทิศทางเดียวกัน อาจเป็น



เพราะเมื่อผลิตภัณฑ์มวลรวมและจำนวนรถใหม่ที่จดทะเบียนสูงขึ้น นั้นหมายถึงเศรษฐกิจดีขึ้น ผู้บริโภคมีกำลังซื้อเพิ่มขึ้น เกิดการจับจ่ายใช้สอยและการเดินทางทั้งรถส่วนตัวและระบบขนส่งมวลชนเพิ่มมากขึ้น ส่วนที่ผู้วิจัยไม่เห็นด้วยกับตัวแปรอิสระที่ถูกคัดเลือกของน้ำมันดีเซลมีดังนี้ 1) จำนวนประชากรที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันดีเซลในทิศทางตรงข้าม โดยผู้วิจัยเชื่อว่าเมื่อจำนวนประชากรเพิ่มมากขึ้นปริมาณการบริโภคน้ำมันดีเซลหรือน้ำมันสำเร็จรูปชนิดอื่นจะเพิ่มขึ้นด้วย เพราะมีจำนวนผู้บริโภคเพิ่มขึ้น 2) ปริมาณการผลิตน้ำมันเบนซิน ผู้วิจัยคิดว่าควรจะเป็นปริมาณการผลิตน้ำมันดีเซลที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันดีเซลมากกว่าปริมาณการผลิตน้ำมันเบนซิน เพราะหากมีการผลิตสูงขึ้นก็นำเข้าน้ำมันสำเร็จรูปลดลง ราคาของน้ำมันดีเซลอาจจะถูกลง ทำให้ผู้บริโภคมีกำลังซื้อน้ำมันดีเซลมากขึ้น ดังนั้นผู้วิจัยจึงเห็นว่าปริมาณการผลิตน้ำมันและปริมาณการบริโภคน้ำมันควรเป็นชนิดเดียวกันมากกว่า สำหรับตัวแปรอิสระที่ถูกคัดเลือกของน้ำมันเบนซินในส่วนที่ผู้วิจัยเห็นด้วยได้แก่ 1) ผลิตภัณฑ์มวลรวม จำนวนรถยนต์จดทะเบียนใหม่ และจำนวนประชากรมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันเบนซินในทิศทางเดียวกัน เหตุผลเป็นเช่นเดียวกันกับที่กล่าวไว้ในส่วนของน้ำมันดีเซล 2) กลุ่มตัวแปรอิสระปริมาณการผลิตน้ำมันสำเร็จรูปที่ตัวแบบได้เลือกมา ผู้วิจัยเห็นด้วยกับปริมาณการผลิตน้ำมันเบนซิน ซึ่งผู้วิจัยคิดว่าหากมีการผลิตน้ำมันเบนซินสูงขึ้นก็นำเข้าน้ำมันเบนซินลดลง ราคาของน้ำมันเบนซินอาจจะถูกลง ทำให้ผู้บริโภคมีกำลังซื้อน้ำมันเบนซินมากขึ้น สำหรับปริมาณการผลิตน้ำมันดีเซลและก๊าซปิโตรเลียมเหลว นั้นผู้วิจัยคิดว่าอาจส่งผลต่อปริมาณการบริโภคน้ำมันเบนซิน เนื่องจากในการกลั่นน้ำมันสำเร็จรูปจากน้ำมันดิบนั้นได้ผลิตภัณฑ์หลายชนิด เมื่อผลิตน้ำมันเบนซินเพิ่มขึ้น จะได้น้ำมันชนิดอื่นเพิ่มขึ้นด้วย ดังนั้นเมื่อนำชุดข้อมูลฝึกสอนมาสร้างตัวแบบด้วยเทคนิคการถดถอยแบบขั้นตอน ข้อมูลปริมาณการผลิตน้ำมันสำเร็จรูปชนิดอื่น ๆ จึงอาจมีอิทธิพลต่อปริมาณการบริโภคน้ำมันเบนซิน ตัวแปรอิสระที่ถูกคัดเลือกของน้ำมันเบนซินในส่วนที่ผู้วิจัยไม่เห็นด้วยได้แก่ 1) ราคาน้ำมันดีเซลที่มีความสัมพันธ์กับปริมาณการบริโภคน้ำมันเบนซินในทิศทางตรงข้าม ผู้วิจัยมีความเห็นว่าราคาน้ำมันเบนซินควรถูกเลือกมากกว่า เนื่องจากเมื่อราคาน้ำมันเบนซินเมื่อสูงขึ้น กำลังซื้อของผู้บริโภคอาจลดลง แต่ผู้วิจัยคิดว่าการที่ตัวแบบได้เลือกราคาน้ำมันดีเซล อาจเป็นเพราะราคาน้ำมันดีเซลและราคาน้ำมันเบนซินอาจมีความสัมพันธ์กันในทิศทางเดียวกัน โดยจากข้อมูลที่ใช้ในการสร้างตัวแบบอาจทำให้ราคาน้ำมันดีเซลมีความสัมพันธ์กับปริมาณการบริโภคน้ำมันเบนซินมากกว่าราคาน้ำมันเบนซิน และสำหรับตัวแปรอิสระที่ถูกคัดเลือกของน้ำมันเครื่องบินมีเพียงตัวแปรเดียวคือ ผลิตภัณฑ์มวลรวม ซึ่งผู้วิจัยเห็นด้วยเพราะการที่ผลิตภัณฑ์มวลรวมสูงขึ้น แสดงถึงผู้บริโภคมีการจับจ่ายใช้สอยมากขึ้น กำลังซื้อเพิ่มขึ้น การเดินทางด้วยเครื่องบินอาจเพิ่มขึ้น ส่งผลทำให้การบริโภคน้ำมันเครื่องบินอาจเพิ่มขึ้นด้วยเช่นกัน โดยตัวแปรอิสระของปริมาณการบริโภคน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินทั้งที่ผู้วิจัยเห็นด้วยและไม่เห็นด้วยจะถูกนำไปเป็นตัวแปรอิสระให้กับตัวแบบอื่น ๆ ต่อไป สำหรับในส่วนของการพยากรณ์ระยะสั้น

ผู้วิจัยขอสรุปว่าการที่ผู้วิจัยมีความเห็นด้วยและไม่เห็นด้วยกับตัวแปรอิสระของน้ำมันสำเร็จรูปแต่ละชนิดที่ถูกคัดเลือกด้วยเทคนิคการถดถอยแบบขั้นตอนนั้น มีความเห็นไปในทางเดียวกับการตัวแบบการพยากรณ์ระยะกลาง ซึ่งตัวแปรอิสระที่ถูกคัดเลือกนั้นแต่ละตัวแบบจะได้ผลลัพธ์ที่แตกต่างกันไปตามแต่ละช่วงเวลาที่มีการเลื่อนตำแหน่งช่วงเวลาตามวิธีการที่ได้กล่าวในหัวข้อที่ 3.2.1 และตัวแบบการถดถอยเชิงเส้นพหุคูณที่ได้ยังพบปัญหาความสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) โดยตัวแบบ MLR ที่ได้ส่วนใหญ่มีค่า MAPE อยู่ในช่วง 3.3% ถึง 6.5% ดังแสดงในตารางที่ 16 17 และ 18 ยกเว้นตัวแบบ MLR ระยะสั้นของน้ำมันเครื่องบินที่มีค่า MAPE ประมาณ 8.9% ส่วนตัวแบบ RANSAC-REG ที่ได้ส่วนใหญ่มีค่า MAPE อยู่ในช่วง 2.9% ถึง 6.5% ยกเว้นตัวแบบ RANSAC-REG ระยะสั้นของน้ำมันเครื่องบินที่มีค่า MAPE ประมาณ 8.7% ผู้วิจัยจึงเห็นว่าตัวแบบอื่น ๆ จึงมีความเหมาะสมที่จะนำไปพยากรณ์ปริมาณน้ำมันสำเร็จรูปมากกว่า และเมื่อพิจารณาการเปรียบเทียบความแม่นยำระหว่างกลุ่มตัวแบบพยากรณ์เชิงสาเหตุด้วยกันคือตัวแบบ MLR และตัวแบบ RANSAC-REG พบว่าตัวแบบ RANSAC-REG ให้ความแม่นยำในการพยากรณ์ดีกว่าตัวแบบ MLR ในการพยากรณ์ทั้งแบบระยะสั้นและแบบระยะยาว ซึ่งผู้วิจัยคิดว่าน่าจะเป็นเพราะตัวแบบ RANSAC-REG ที่มีความแกร่ง (Robust) ต่อค่าผิดปกติของข้อมูลป้อนเข้า (Outlier ของตัวแปรอิสระ) ในขณะที่ตัวแบบ MLR นั้นอ่อนไหวต่อค่าผิดปกติของข้อมูลป้อนเข้า นอกจากนี้ตัวแบบ RANSAC-REG และตัวแบบ MLR ทั้งการพยากรณ์แบบระยะสั้นและระยะยาวยังมีทิศทางของสัมประสิทธิ์ของตัวแปรอิสระที่เหมือนกันทั้งหมด แต่มีขนาดของสัมประสิทธิ์ที่แตกต่างกัน ซึ่งผู้วิจัยคิดว่าขนาดของสัมประสิทธิ์ที่มีค่าแตกต่างกันน่าจะเป็นเพราะตัวแบบ RANSAC-REG จะไม่นำข้อมูลป้อนเข้าที่มีค่าผิดปกติมาใช้ในการประมาณค่าสัมประสิทธิ์ของตัวแบบ และผู้วิจัยมีความเห็นด้วยและไม่เห็นกับทิศทางของสัมประสิทธิ์ของแต่ละตัวแปรอิสระที่ได้จากตัวแบบ RANSAC-REG เช่นเดียวกันกับตัวแบบ MLR เนื่องจากมีทิศทางของสัมประสิทธิ์ของตัวแปรอิสระเป็นทิศทางเดียวกัน

2) ตัวแบบการพยากรณ์ในกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติและการเปรียบเทียบผลลัพธ์การพยากรณ์ระยะและระยะกลาง

ผู้วิจัยได้ศึกษาตัวแบบ Holt-Winters ตัวแบบ SARIMA และตัวแบบ SARIAMAX ที่ใช้ตัวแปรอิสระจากการคัดเลือกด้วยวิธีการถดถอยแบบขั้นตอน ผลลัพธ์ของตัวแบบการพยากรณ์ระยะกลางที่ได้มีค่า MAPE อยู่ในช่วง 3.5% ถึง 7.1% ดังแสดงในตารางที่ 16 17 และ 18 โดยตัวแบบ SARIAMAX ของน้ำมันสำเร็จรูปทั้ง 3 ชนิดให้ความแม่นยำมากที่สุดในกลุ่มการพยากรณ์ระยะกลางด้วยตัวแบบอนุกรมเวลาเชิงสถิติ แต่เมื่อพิจารณาการพยากรณ์ระยะสั้นด้วยตัวแบบอนุกรมเวลาเชิงสถิติพบว่าตัวแบบ SARIAMAX ให้ผลลัพธ์การพยากรณ์ที่แม่นยำน้อยกว่าตัวแบบ Holt-Winters และตัวแบบ SARIMA ผลลัพธ์ของตัวแบบการพยากรณ์ระยะสั้นที่ได้มีค่า MAPE อยู่ในช่วง 2.6% ถึง 6.0%

ดังแสดงในตารางที่ 16 17 และ 18 เมื่อพิจารณาเปรียบเทียบการพยากรณ์ระยะสั้นกับการพยากรณ์ระยะกลางด้วยตัวแบบอนุกรมเวลาเชิงสถิติและตัวแบบ MLR พบว่าสำหรับการพยากรณ์ระยะสั้น ตัวแบบอนุกรมเวลาเชิงสถิติที่ไม่มีตัวแปรอิสระภายนอก (Holt-Winters และ SARIMA) จะมีความแม่นยำสูงกว่าตัวแบบที่มีตัวแปรอิสระภายนอก (MLR RANSAC-REG และ SARIMAX) นอกจากนี้เมื่อเปรียบเทียบกราฟการพยากรณ์ระยะสั้นและระยะกลางของตัวแบบอนุกรมเวลาเชิงสถิติของน้ำมันสำเร็จรูปทั้ง 3 ชนิดจากรูปที่ 24 ถึงรูปที่ 29 สำหรับตัวแบบ Holt-Winters และรูปที่ 39 ถึงรูปที่ 44 สำหรับตัวแบบ SARIMA พบว่าส่วนใหญ่ค่าพยากรณ์ที่ได้จะมีค่าต่ำกว่าค่าจริงโดยเฉพาะในช่วงกลางจนถึงช่วงท้ายของกราฟ แสดงว่าตัวแบบจะวิ่งตามแนวโน้มได้ไม่ดี แต่กราฟของการพยากรณ์ระยะสั้นของตัวแบบอนุกรมเวลาจะให้ค่าพยากรณ์ที่ตามแนวโน้มได้ดี การพยากรณ์ระยะสั้นจึงพบปัญหาค่าพยากรณ์มีค่าต่ำกว่าค่าจริงน้อยกว่าการพยากรณ์ระยะกลาง สำหรับการพยากรณ์ระยะกลางนั้นเมื่อพิจารณาจากค่า MAPE แล้วพบว่าตัวแบบที่มีตัวแปรอิสระภายนอก (MLR RANSAC-REG และ SARIMAX) จะให้ผลการพยากรณ์ที่แม่นยำกว่าตัวแบบอนุกรมเวลาเชิงสถิติที่ไม่มีตัวแปรอิสระภายนอก (Holt-Winters และ SARIMA) นอกจากนี้ตัวแบบ ตัวแบบอนุกรมเวลาเชิงสถิติที่ไม่มีตัวแปรอิสระภายนอก (Holt-Winters และ SARIMA) จะให้ผลการพยากรณ์ระยะสั้นที่นำมาเฉลี่ย 5 ปี แม่นยำว่าการพยากรณ์ระยะกลาง ส่วนตัวแบบที่มีตัวแปรอิสระภายนอก (MLR RANSAC-REG และ SARIMAX) นั้นมีทั้งที่ผลการพยากรณ์ระยะกลางแม่นยำกว่าระยะสั้น และผลการพยากรณ์ระยะสั้นแม่นยำกว่าระยะกลาง ผู้วิจัยคิดว่าน่าจะเป็นผลมาจากปัจจัยภายนอกที่อาจมีการเปลี่ยนแปลงบางช่วงที่ไม่ขึ้นกับรูปแบบอนุกรมเวลาหรืออาจมีความสัมพันธ์กันเอง

### 3) ตัวแบบการพยากรณ์ในกลุ่มตัวแบบการเรียนรู้ของเครื่อง

ผู้วิจัยได้ศึกษาตัวแบบ KNN ตัวแบบ SVR ตัวแบบ ANN และตัวแบบ ADA ผลลัพธ์ที่ได้จากการสร้างตัวแบบการพยากรณ์ระยะกลางพบว่า ตัวแบบ ADA ให้ความแม่นยำดีที่สุดในการพยากรณ์น้ำมันสำเร็จรูปทั้ง 3 ชนิด แต่เมื่อพิจารณาในแง่การนำตัวแบบไปใช้งานกับชุดข้อมูลอื่นที่ไม่เคยเห็นมาก่อน โดยที่ไม่เกิดปัญหาความเข้ากันเกินไปของข้อมูลกับตัวแบบที่สร้าง (Overfitting Problem) พบว่าตัวแบบ ADA และตัวแบบ KNN มีปัญหา Overfitting สูงมาก เมื่อเทียบกับตัวแบบ SVR และตัวแบบ ANN ซึ่งมีความแม่นยำน้อยกว่าตัวแบบ ADA ไม่มากนัก แต่มีปัญหา Overfitting น้อยกว่าตัวแบบ ADA และตัวแบบ KNN ค่อนข้างมาก สำหรับสาเหตุที่ตัวแบบ ADA และตัวแบบ KNN พบปัญหา Overfitting สูงในการใช้พยากรณ์ปริมาณการบริโภคน้ำมันสำเร็จรูปทั้ง 3 ชนิดนั้น ผู้วิจัยมีความคิดเห็นที่น่าจะเป็นเพราะ 1) ตัวแบบ KNN เป็นตัวแบบที่ใช้วิธีการพยากรณ์จากการคำนวณค่าเฉลี่ยของของตัวแปรตามของเพื่อนบ้านใกล้สุด  $k$  ตัวของตัวแปรอิสระ ซึ่งเป็นวิธีการที่มีรูปแบบซับซ้อนและมีจำนวนไฮเปอร์พารามิเตอร์ที่สามารถปรับแต่งตัวแบบได้น้อย มีไฮเปอร์พารามิเตอร์ที่

สำคัญคือ  $n\_neighbors(k)$  เนื่องจากตัวแบบ KNN มีรูปแบบที่ซับซ้อนและปรับแต่งไฮเปอร์พารามิเตอร์ของตัวแบบได้น้อยจึงมีโอกาสเกิดปัญหา Overfitting สูง 2) ตัวแบบ ADA เป็นตัวแบบที่ใช้เพิ่มความแม่นยำให้กับตัวแบบต้นไม้การถดถอย (Regression Tree) มีหลักการคือให้ Regression Tree เป็น Weak Learner (หมายถึงมีความแม่นยำที่ต่ำกว่าการเดาสุ่มเพียงเล็กน้อยเท่านั้น) เมื่อนำ Weak Learner หลาย ๆ ตัวแบบไปรวมกลุ่มด้วยวิธีการ Ensemble ในรูปแบบ Boosting ด้วยตัวแบบ Adaboost จะได้เป็น Strong Learner (ได้ความแม่นยำสูง) แต่มักจะพบปัญหา Overfitting กับข้อมูลที่มีค่าผิดพลาดน้อย (Low Noise Data) จากที่ได้กล่าวมาข้างต้น เมื่อพิจารณาจากค่าความแม่นยำของแต่ละตัวแบบการเรียนรู้ของเครื่องที่ผู้วิจัยได้ศึกษา จากตารางที่ 19 พบว่าค่าความแม่นยำของแต่ละตัวแบบไม่ได้แตกต่างกันมากนัก แต่ตัวแบบ KNN และตัวแบบ ADA พบปัญหา Overfitting ค่อนข้างสูง ผู้วิจัยจึงเลือกตัวแบบ SVR และตัวแบบ ANN ให้เป็นตัวแบบที่เหมาะสมในกลุ่มตัวแบบการเรียนรู้ของเครื่องของน้ำมันสำเร็จรูปทั้ง 3 ชนิดที่จะนำไปใช้ในการผสมตัวแบบต่อไป ซึ่งตัวแบบ SVR ผู้วิจัยเห็นว่าจุดเด่นคือ มีความสามารถในการจับรูปแบบข้อมูลที่เป็นฟังก์ชันไม่เชิงเส้นได้โดยการเลือก Kernel Function ให้เหมาะสมกับข้อมูล ส่วนตัวแบบ ANN ผู้วิจัยเห็นว่าจุดเด่นคือ มีความสามารถในการหาความสัมพันธ์ระหว่างตัวแปรอิสระด้วยกันที่ซับซ้อนได้ดีและยังมีความสามารถในการจับรูปแบบข้อมูลที่เป็นฟังก์ชันไม่เชิงเส้นได้เช่นกัน อีกทั้งตัวแบบ SVR และตัวแบบ ANN สามารถเลือกปรับไฮเปอร์พารามิเตอร์ได้หลากหลาย เมื่อนำไปใช้เป็นตัวแบบเดียวในการผสมตัวแบบ น่าจะช่วยเพิ่มความแม่นยำและลดปัญหา Overfitting ของตัวแบบการพยากรณ์น้ำมันสำเร็จรูปทั้ง 3 ชนิดได้

#### 4) ตัวแบบการพยากรณ์ในกลุ่มตัวแบบผสมและผลลัพธ์ของตัวแบบที่เหมาะสม

ผู้วิจัยเลือกศึกษาการผสมตัวแบบที่มีการพื้นฐานจากการผสมผสานส่วนประกอบของตัวแบบตามโครงสร้างการผสมแบบขนานที่สามารถช่วยลดความเสี่ยงจากการใช้แบบจำลองที่ไม่เหมาะสมได้ และตามโครงสร้างการผสมแบบอนุกรมที่สามารถช่วยปรับปรุงความแม่นยำของการพยากรณ์ให้ดีขึ้น เนื่องจากการตรวจจับรูปแบบข้อมูลและการสร้างตัวแบบที่ครอบคลุม และยังศึกษาการสร้างตัวแบบที่มีการผสมทั้งโครงสร้างการผสมแบบอนุกรมผสมกับแบบขนานเข้าด้วยกันเพื่อเป็นการนำจุดเด่นของโครงสร้างทั้งสองแบบมาใช้สร้างตัวแบบ โดยการเลือกใช้ตัวแบบเดี่ยวที่จะนำมาผสมตัวแบบนั้นมีจุดมุ่งหมายคือเพื่อให้ตัวแบบผสมมีความสามารถในการตรวจจับรูปแบบข้อมูลให้ได้ทั้งรูปแบบในฟังก์ชันเชิงเส้นและรูปแบบฟังก์ชันไม่เชิงเส้น และมีการสร้างตัวแบบที่ครอบคลุม ผู้วิจัยจึงเลือกตัวแบบการพยากรณ์จากกลุ่มตัวแบบอนุกรมเวลาเชิงสถิติเป็นตัวแบบ SARIMAX ที่ใช้ในการพยากรณ์ระยะกลาง ซึ่งให้ค่าความแม่นยำที่ค่อนข้างสูงเมื่อเทียบกับตัวอื่น และยังสามารถใช้ทดแทนการพยากรณ์ระยะสั้นได้เนื่องจากเมื่อพยากรณ์ไปข้างหน้าจำนวน 20 ไตรมาสยังให้ค่า MAPE ที่ค่อนข้าง

ต่ำคือไม่เกิน 5% โดยจะใช้ตัวแบบ SARIMAX ไปพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้น ส่วนตัวแบบในกลุ่มการเรียนรู้ของเครื่องที่เหมาะสมสำหรับนำไปพยากรณ์ข้อมูลในส่วนของฟังก์ชันไม่เชิงเส้น ผู้วิจัยเลือกใช้ตัวแบบ SVR หรือ ANN แบบใช้ตัวแปรอิสระที่ถูกคัดเลือกจากการถดถอยแบบขั้นตอน เพื่อให้สามารถผสมกับตัวแบบ SARIMAX ที่ใช้ตัวแปรอิสระภายนอกจากการคัดเลือกจากการถดถอยแบบขั้นตอนได้แบบใช้ตัวแปรอิสระตัวเดียวกันในการสร้างตัวแบบผสม ผลลัพธ์ของตัวแบบผสมที่เหมาะสมในการพยากรณ์น้ำมันดีเซลทั้งแบบระยะสั้นและระยะกลางคือตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ที่มีโครงสร้างการผสมแบบอนุกรมผสมกับแบบขนาน ตัวแบบ SARIMAX-ANN-SVR-RANSAC-REG ของน้ำมันดีเซล มีค่า Validation MAPE เท่ากับ 2.2785% และมีค่า Generalization Error MAPE เท่ากับ 5.4835% ซึ่งมีความเหมาะสมสำหรับการพยากรณ์ปริมาณการบริโภคน้ำมันดีเซล ส่วนผลลัพธ์ตัวแบบผสมที่เหมาะสมในการพยากรณ์น้ำมันเบนซินคือ ตัวแบบ SARIMAX-ANN-RANSAC-REG ที่มีโครงสร้างการผสมแบบขนาน มีค่า Validation MAPE เท่ากับ 1.9966% และมีค่า Generalization Error MAPE เท่ากับ 6.3873% ซึ่งมีความเหมาะสมสำหรับการพยากรณ์ปริมาณการบริโภคน้ำมันเบนซิน ส่วนผลลัพธ์ตัวแบบผสมที่เหมาะสมในการพยากรณ์น้ำมันเครื่องบินคือ ตัวแบบ SARIMAX-SVR ที่มีโครงสร้างการผสมแบบอนุกรม มีค่า Validation MAPE เท่ากับ 3.5055% และมีค่า Generalization Error MAPE เท่ากับ 7.2962% ซึ่งมีความเหมาะสมสำหรับการพยากรณ์ปริมาณการบริโภคน้ำมันเครื่องบิน ตัวแบบเหล่านี้มีความแม่นยำสูงและมีปัญหาความเข้ากันเกินไปของข้อมูลไม่มาก และจากการเปรียบเทียบผลลัพธ์ของตัวแบบการพยากรณ์ทั้งหมดของน้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบิน พบว่าการพยากรณ์ด้วยวิธีการผสมตัวแบบที่มีพื้นฐานจากการผสมผสานส่วนประกอบของตัวแบบสามารถเพิ่มความแม่นยำของตัวแบบการพยากรณ์ได้ดี อีกทั้งตัวแบบผสมที่ศึกษาในงานวิจัยนี้มีความสามารถในการจับรูปแบบข้อมูลทั้งที่เป็นฟังก์ชันเชิงเส้นและไม่เป็นฟังก์ชันเชิงเส้นได้จึงช่วยเพิ่มความแม่นยำในการพยากรณ์และยังสามารถช่วยลดความเสี่ยงในการเลือกใช้ตัวแบบการพยากรณ์ที่ไม่เหมาะสมโดยการใช้การรวมผลการพยากรณ์ด้วยวิธีการ RANSAC Regression ในส่วนของความซับซ้อนในการสร้างตัวแบบนั้นเป็นตัวแบบที่มีความซับซ้อนไม่มาก สามารถเข้าใจกระบวนการสร้างตัวแบบได้ไม่ยากและใช้เวลาในการสร้างตัวแบบไม่นาน ผู้วิจัยจึงเห็นว่าตัวแบบผสมที่ได้มีความเหมาะสมที่จะนำไปใช้พยากรณ์น้ำมันสำเร็จรูปกลุ่มหลักของประเทศไทยทั้งการพยากรณ์แบบระยะกลางและแบบระยะสั้นด้วยเนื่องจากผลการพยากรณ์ระยะกลางครอบคลุมการพยากรณ์ระยะสั้นอยู่แล้ว

## 5.2 ข้อเสนอแนะ

1) วิธีการที่จะช่วยเพิ่มความแม่นยำให้กับตัวแบบการพยากรณ์นั้นมีหลากหลายวิธี สำหรับตัวแบบการเรียนรู้ของเครื่องนั้นความแม่นยำของตัวแบบขึ้นอยู่กับค่าไฮเปอร์พารามิเตอร์ที่เลือกใช้

วิธีการค้นหาไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบ SVR และตัวแบบ ANN มีหลากหลายวิธี ผู้วิจัยเลือกใช้ GridSearchCV ซึ่งก็สามารถค้นหาไฮเปอร์พารามิเตอร์ที่เหมาะสมได้ตามค่ากำหนดที่ผู้วิจัยเลือกใช้ แต่ค่ากำหนดที่เลือกใช้อาจจะยังไม่ใช่ค่าที่ทำให้ได้ตัวแบบที่แม่นยำที่สุดก็ได้ ทางแก้คือเพิ่มจำนวนค่ากำหนดของไฮเปอร์พารามิเตอร์ให้มากขึ้น แต่ก็ยังทำให้ใช้เวลาในการประมวลผลสำหรับสร้างตัวแบบนานยิ่งขึ้นไปอีก ผู้วิจัยเห็นว่ายังมีวิธีการค้นหาไฮเปอร์พารามิเตอร์ที่เหมาะสมวิธีอื่นที่อาจมีประสิทธิภาพมากกว่า อย่างการใช้ฮิวริสติกค้นหาค่าไฮเปอร์พารามิเตอร์ เช่น Genetic Algorithm, Particle Swarm Optimization, Bayesian Optimization หรือใช้วิธีการอื่น ๆ การนำฮิวริสติกมาใช้ในการค้นหาค่าไฮเปอร์พารามิเตอร์ของตัวแบบนั้นเป็นวิธีการผสมตัวแบบอย่างหนึ่งที่สามารถช่วยเพิ่มความแม่นยำให้กับตัวแบบการพยากรณ์ได้

2) ตัวแปรอิสระที่ใช้ในตัวแบบการพยากรณ์ปริมาณการบริโภคน้ำมันเครื่องบินมีตัวแปรอิสระที่ถูกคัดเลือกจากวิธีการถดถอยแบบขั้นตอนเพียงตัวแปรเดียวคือผลิตภัณฑ์มวลรวม การหาตัวแปรอิสระอื่นที่ส่งผลต่อปริมาณการบริโภคน้ำมันเครื่องบินมาเพิ่มในตัวแบบอาจช่วยเพิ่มความแม่นยำของตัวแบบได้ นอกจากนี้การใช้วิธีการคัดเลือกตัวแปรอิสระด้วยเทคนิคการถดถอยแบบขั้นตอนนั้นอาจไม่ได้ให้คำตอบที่ดีที่สุด เช่น กรณีน้ำมันเครื่องบิน ถ้าใส่ตัวแปรอิสระทุกตัว ตัวแบบที่ได้อาจจะมี MAPE ที่ต่ำกว่าการใช้ตัวแปรอิสระที่ได้จากเทคนิคการถดถอยแบบขั้นตอนที่เลือกตัวแปรอิสระมาเพียงตัวแปรเดียว

3) การใช้เทคนิคการถดถอยแบบขั้นตอนคัดเลือกตัวแปรอิสระที่ส่งผลต่อปริมาณการบริโภคน้ำมันสำเร็จรูปทั้ง 3 ชนิด แล้วได้ผลลัพธ์เป็นตัวแปรอิสระที่ผู้วิจัยไม่เห็นด้วยดังที่ได้อธิบายไว้ในหัวข้อที่ 4.3 อาจมีสาเหตุจากตัวแปรอิสระมีความสัมพันธ์กันเองทำให้เกิดปัญหา Multicollinearity จึงควรใช้ค่าสหสัมพันธ์ (Correlation) เพื่อตัดตัวแปรอิสระที่มีความซ้ำซ้อนกัน (Redundant) ออก

4) ในการนำผลการพยากรณ์ในงานวิจัยไปใช้งานในหน่วยงานที่เกี่ยวข้อง เช่น สำนักงานนโยบายและแผนพลังงานมีการประมาณการความต้องการใช้น้ำมันดีเซล น้ำมันเบนซิน และน้ำมันเครื่องบินระยะสั้นเป็นราย 6 เดือน แต่ในงานของผู้วิจัยเป็นการพยากรณ์แบบรายไตรมาส การพยากรณ์ระยะสั้นเป็นราย 6 เดือนสามารถใช้วิธีการรวมผลการพยากรณ์รายไตรมาส โดยนำผลการพยากรณ์ไตรมาสที่ 1 และ 2 ของปีรวมกัน จะได้เป็นผลการพยากรณ์ราย 6 เดือนแรกของปี และนำผลการพยากรณ์ไตรมาสที่ 3 และ 4 ของปีรวมกัน จะได้เป็นผลการพยากรณ์ราย 6 เดือนหลังของปี หรือหากต้องการผลการพยากรณ์แบบรายปี สามารถใช้วิธีรวมผลการพยากรณ์ไตรมาสที่ 1 2 3 และ 4 ของปี จะได้เป็นผลการพยากรณ์รายปี เป็นต้น

บรรณานุกรม



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

- Al-Dosary, N. M. N., Al-Hamed, S. A., & Aboukarima, A. M. (2019). K-Nearest Neighbors Method for Prediction of Fuel Consumption in Tractor-Chisel Plow Systems. *Engenharia Agrícola*, 39(6), 729-736.
- Al-Fattah, S. M. (2020). A new artificial intelligence GANNATS model predicts gasoline demand of Saudi Arabia. *Journal of Petroleum Science and Engineering*, 194.
- Azadeh, A., Khakestani, M., & Saberi, M. (2009). A flexible fuzzy regression algorithm for forecasting oil consumption estimation. *Energy Policy*, 37(12), 5567-5579.
- Breiman, L. (1996). Stacked regressions. *Machine Learning*, 24(1), 49-64.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A. (1984). *Classification and Regression Trees*. New York, NY: Chapman and Hall/CRC.
- Fischler, M.A., & Bolles, R.C. (1981). Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6), 381-395.
- Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and Computation*, 121 (2), 256-285.
- Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn Keras and TensorFlow* (2nd ed.). Sebastopol, CA: O'Reilly Media, Inc.
- Ghahremani, M., Williams, K., Corke, F., Tiddeman, B., Liu, Y., Wang, X., & Doonan, J. H. (2021). Direct and accurate feature extraction from 3D point clouds of plants using RANSAC. *Computers and Electronics in Agriculture*, 187, 106240.
- Hajirahimi, Z., & Khasheji, M. (2019). Hybrid structures in time series modeling and forecasting: A review. *Engineering Applications of Artificial Intelligence*, 86, 83-106.
- Hanke, J. E., & Wichern, D. W. (2009). *Business forecasting* (9th ed.). Upper Saddle River, N J: Pearson Education.
- Hyndman, R. J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice* (2nd ed.). Melbourne, Australia: OTexts. OTexts.com/fpp2. Accessed on 2020, Dec 14.
- Karakurt, I. (2021). Modelling and forecasting the oil consumptions of the BRICS-T Countries. *Energy*, 220, 119720.



- Kristjanpoller, W., Prado, F., & Minutolo, M. C. (2020). Forecasting based on an ensemble Autoregressive Moving Average -Adaptive neuro - Fuzzy inference system Neural network – Genetic Algorithm Framework. *Energy*, 197, 117159.
- Lee, S., Kc, B., & Choeh, J. Y. (2020). Comparing performance of ensemble methods in predicting movie box office revenue. *Heliyon*, 6(6), e04260.
- Li, Y., & Gans, N. R., (2017). Predictive RANSAC: Effective model fitting and tracking approach under heavy noise and outliers. *Computer Vision and Image Understanding*, 161, 99-113.
- Liddle, B., & Huntington, H. (2020). ‘On the Road Again’: A 118 country panel analysis of gasoline and diesel demand. *Transportation Research Part A: Policy and Practice*, 142, 151-167.
- Lu, H., Azimi, M., & Iseley, T. (2019). Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine. *Energy Reports*, 5, 666-677.
- Maehashi, K., & Shintani, M. (2020). Macroeconomic forecasting using factor models and machine learning: an application to Japan. *Journal of the Japanese and International Economies*, 58.
- Melikoglu, M. (2014). Demand forecast for road transportation fuels including gasoline, diesel, LPG, bioethanol and biodiesel for Turkey between 2013 and 2023. *Renewable Energy*, 64, 164-171.
- Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, 323, 203-213.
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), 210-229.
- Schapire, R.E. (1990). The strength of weak learnability. *Machine Learning*, 5 (2), 197–227.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms* (1st ed.). New York, NY: Cambridge University Press.
- Sheather, S. (2009). *A modern approach to regression with R*. New York, NY: Springer.

Xiao, J., Li, Y., Xie, L., Liu, D., & Huang, J. (2018). A hybrid model based on selective ensemble for energy consumption forecasting in China. *Energy*, 159, 534-546.

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL: Chapman and Hall/CRC Press.

เอกจิตต์ จีงเจริญ. (2560). *การพยากรณ์ทางธุรกิจ* (ฉบับพิมพ์ครั้งที่ 1 ed.). กรุงเทพมหานคร: สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์.

นันทชัย กานตานันทะ. (2555). การพยากรณ์ด้วยวิธีการพยากรณ์เชิงสาเหตุ. *วารสารวิศวกรรมศาสตร์*, 4(1), 33-48.

สำนักงานนโยบายและแผนพลังงาน. (2563). ข้อมูลพลังงาน. สืบค้นจาก <http://www.eppo.go.th/index.php/th/energy-information#>. สืบค้นเมื่อ 10 กันยายน 2563.

สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ. (2563). ภาวะเศรษฐกิจในประเทศไทยไตรมาส. สืบค้นจาก [https://www.nesdc.go.th/main.php?filename=qgdp\\_page](https://www.nesdc.go.th/main.php?filename=qgdp_page). สืบค้นเมื่อ 10 กันยายน 2563.

### ภาคผนวก

ตารางผนวกที่ 1 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ KNN ของน้ำมันดีเซลแบบใช้ตัวแปรอิสระที่คัดเลือก

No	metric	n_neighbors	Mean_CV_MSE	Mean_Train_MSE
1	euclidean	4	152,850	29,502
2	minkowski	4	152,850	29,502
3	euclidean	3	154,232	24,899
4	minkowski	3	154,232	24,899
5	euclidean	2	160,707	21,128
6	minkowski	2	160,707	21,128
7	manhattan	4	171,695	29,596
8	euclidean	5	176,329	32,528
9	minkowski	5	176,329	32,528
10	manhattan	3	177,862	24,071

ตารางผนวกที่ 2 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ KNN ของน้ำมันเบนซินแบบใช้ตัวแปรอิสระที่คัดเลือก

No	metric	n_neighbors	Mean_CV_MSE	Mean_Train_MSE
1	euclidean	10	15,100	5,637
2	minkowski	10	15,100	5,637
3	euclidean	11	15,166	7,047
4	minkowski	11	15,166	7,047
5	euclidean	9	15,185	4,222
6	minkowski	9	15,185	4,222
7	manhattan	9	15,324	3,878
8	manhattan	10	15,334	5,069
9	euclidean	12	15,358	8,535
10	minkowski	12	15,358	8,535

ตารางผนวกที่ 3 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ KNN ของน้ำมันเครื่องบินแบบใช้ตัวแปรอิสระที่คัดเลือก

No	metric	n_neighbors	Mean_CV_MSE	Mean_Train_MSE
1	euclidean	2	12,342	899
2	manhattan	2	12,342	899
3	minkowski	2	12,342	899
4	euclidean	4	17,186	1,219
5	manhattan	4	17,186	1,219
6	minkowski	4	17,186	1,219
7	euclidean	3	18,227	1,078
8	manhattan	3	18,227	1,078
9	minkowski	3	18,227	1,078
10	euclidean	6	21,470	1,496

ตารางผนวกที่ 4 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ SVR ของน้ำมันดีเซลแบบใช้ตัวแปรอิสระที่คัดเลือก

No	C	degree	gamma	kernel	Mean_CV_MSE	Mean_Train_MSE
1	3000	1	0.1	rbf	79,185	62,628
2	3000	2	0.1	rbf	79,185	62,628
3	3000	3	0.1	rbf	79,185	62,628
4	3000	4	0.1	rbf	79,185	62,628
5	3000	5	0.1	rbf	79,185	62,628
6	3000	6	0.1	rbf	79,185	62,628
7	10000	1	0.1	sigmoid	84,037	51,085
8	10000	2	0.1	sigmoid	84,037	51,085
9	10000	3	0.1	sigmoid	84,037	51,085
10	10000	4	0.1	sigmoid	84,037	51,085

ตารางผนวกที่ 5 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ SVR ของน้ำมันเบนซินแบบใช้ตัวแปรอิสระที่คัดเลือก

No	C	degree	gamma	kernel	Mean_CV_MSE	Mean_Train_MSE
1	1000	1	0.1	sigmoid	7,007	6,778
2	1000	2	0.1	sigmoid	7,007	6,778
3	1000	3	0.1	sigmoid	7,007	6,778
4	1000	4	0.1	sigmoid	7,007	6,778
5	1000	5	0.1	sigmoid	7,007	6,778
6	1000	6	0.1	sigmoid	7,007	6,778
7	10000	1	0.01	sigmoid	7,236	6,813
8	10000	2	0.01	sigmoid	7,236	6,813
9	10000	3	0.01	sigmoid	7,236	6,813
10	10000	4	0.01	sigmoid	7,236	6,813

ตารางผนวกที่ 6 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ SVR ของน้ำมันเครื่องบินแบบใช้ตัวแปรอิสระที่คัดเลือก

No	C	degree	gamma	kernel	Mean_CV_MSE	Mean_Train_MSE
1	10000	1	0.1	rbf	4,175	1,422
2	10000	2	0.1	rbf	4,175	1,422
3	10000	3	0.1	rbf	4,175	1,422
4	10000	4	0.1	rbf	4,175	1,422
5	10000	5	0.1	rbf	4,175	1,422
6	10000	6	0.1	rbf	4,175	1,422
7	10000	1	1	poly	4,256	1,417
8	10000	1	1	linear	4,256	1,417
9	10000	1	0.1	linear	4,256	1,417
10	10000	1	0.01	linear	4,256	1,417

ตารางผนวกที่ 7 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ ANN ของน้ำมันดีเซลแบบใช้ตัวแปรอิสระที่คัดเลือก

No	activation	Batch_size	dropout	epochs	NumLayers	optimizer	seed	units	Mean_CV MSE	Mean_Train MSE
1	selu	48	0.2	100	2	adam	23	30	0.005279	0.005792
2	selu	48	0.2	500	2	sgd	23	30	0.005807	0.005179
3	selu	24	0	1000	1	adam	456	3	0.005916	0.004489
4	selu	88	0	1000	1	adam	456	3	0.006031	0.004697
5	elu	24	0.2	1000	1	adam	456	3	0.006121	0.005455
6	selu	88	0.2	500	2	sgd	23	30	0.006179	0.005115
7	selu	24	0.2	500	2	sgd	23	30	0.006210	0.005056
8	selu	48	0	100	2	adam	456	30	0.006374	0.004731
9	elu	88	0.2	1000	1	adam	456	3	0.006441	0.006054
10	selu	48	0	1000	1	adam	456	3	0.006469	0.004639

ตารางผนวกที่ 8 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ ANN ของน้ำมันเบนซินแบบใช้ตัวแปรอิสระที่คัดเลือก

No	activation	Batch_size	dropout	epochs	NumLayers	optimizer	seed	units	Mean_CV MSE	Mean_Train MSE
1	elu	24	0.2	1000	2	adam	1	5	0.002397	0.002863
2	elu	48	0.2	1000	2	adam	1	5	0.002893	0.003102
3	elu	24	0.2	500	2	adam	1	5	0.002993	0.003859
4	elu	24	0.4	1000	2	adam	1	5	0.003378	0.004385
5	elu	48	0.4	1000	2	adam	1	5	0.003443	0.004717
6	relu	48	0	50	2	adam	456	30	0.003450	0.003110
7	relu	48	0	100	2	adam	456	30	0.003462	0.001994
8	relu	24	0	50	2	adam	456	30	0.003552	0.002657
9	relu	88	0	100	2	adam	456	30	0.003597	0.002092
10	selu	24	0.4	100	2	adam	456	30	0.003726	0.006393



ตารางผนวกที่ 9 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ ANN ของน้ำมันเครื่องบินแบบใช้ตัวแปรอิสระที่คัดเลือก

No	activation	Batch_size	dropout	epochs	NumLayers	optimizer	seed	units	Mean_CV MSE	Mean_Train MSE
1	elu	88	0.4	500	1	adam	456	3	0.005039	0.002182
2	selu	88	0.2	50	2	adam	456	50	0.005142	0.002366
3	elu	88	0.4	500	1	adam	456	5	0.005144	0.002145
4	elu	48	0.2	500	1	adam	456	3	0.005148	0.002071
5	relu	24	0.2	1000	1	sgd	456	3	0.005152	0.003204
6	selu	48	0.2	50	2	adam	456	50	0.005164	0.002016
7	relu	48	0	25	1	adam	456	5	0.005168	0.001967
8	relu	24	0.4	100	1	adam	456	3	0.005183	0.003268
9	relu	48	0.4	100	1	adam	456	3	0.005203	0.004897
10	elu	48	0.2	500	1	adam	456	5	0.005213	0.002039

ตารางผนวกที่ 10 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ ADA ของน้ำมันดีเซลแบบใช้ตัวแปรอิสระที่คัดเลือก

No	learning_rate	n_estimators	Random state	Mean_CV_MSE	Mean_Train_MSE
1	0.01	50	1	203,892	17,101
2	0.01	50	456	205,040	15,997
3	0.1	10	7890	207,996	22,541
4	0.1	10	111213	208,154	21,014
5	0.0001	100	7890	211,472	18,243
6	0.001	100	7890	212,198	17,991
7	0.01	10	111213	213,225	18,205
8	0.001	50	7890	213,294	18,166
9	0.01	100	1	214,648	15,425
10	0.0001	50	7890	215,124	17,940

ตารางผนวกที่ 11 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ ADA ของน้ำมันเบนซินแบบใช้ตัวแปรอิสระที่คัดเลือก

No	learning_rate	n_estimators	Random state	Mean_CV_MSE	Mean_Train_MSE
1	0.1	10	23	13,403	1,267
2	0.0001	10	456	14,533	1,635
3	0.0001	10	1	15,835	1,213
4	1	10	1	16,350	1,179
5	0.01	10	23	16,601	1,356
6	0.01	100	1	17,127	1,113
7	0.001	10	456	17,218	1,487
8	0.1	10	111213	17,233	1,270
9	0.001	50	23	17,241	1,230
10	0.0001	50	23	17,287	1,311

ตารางผนวกที่ 12 ค่าเฉลี่ย MSE ที่ต่ำสุด 10 อันดับแรกของการปรับไฮเปอร์พารามิเตอร์ของตัวแบบ ADA ของน้ำมันเครื่องบินแบบใช้ตัวแปรอิสระที่คัดเลือก

No	learning_rate	n_estimators	Random state	Mean_CV_MSE	Mean_Train_MSE
1	0.01	100	23	11,987	800
2	0.01	50	23	12,021	835
3	0.001	100	23	12,038	829
4	0.001	50	23	12,055	865
5	0.0001	50	23	12,105	862
6	0.001	50	456	12,146	841
7	0.001	1000	1	12,156	805
8	0.0001	50	456	12,156	843
9	0.001	1000	111213	12,162	798
10	0.001	500	23	11,987	800

## ประวัติผู้เขียน

ชื่อ-สกุล	พิศาล สามัง
วัน เดือน ปี เกิด	29 กรกฎาคม 2532
สถานที่เกิด	ร้อยเอ็ด
วุฒิการศึกษา	จุฬาลงกรณ์มหาวิทยาลัย
ที่อยู่ปัจจุบัน	1150/36 หมู่ที่ 3 แขวงคลองสองต้นนุ่น เขตลาดกระบัง กรุงเทพมหานคร



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY