

การพยากรณ์ราคาเม็ดพลาสติกพอลิโพรพีลีนในประเทศไทยโดยใช้วิธีการเรียนรู้ของเครื่อง



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมอุตสาหการ ภาควิชาวิศวกรรมอุตสาหการ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2565

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

FORECASTING OF POLYPROPYLENE GRANULE PRICE IN THAILAND USING MACHINE
LEARNING



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 2022

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพยากรณ์ราคาเม็ดพลาสติกพอลิโพรพิลีนในประเทศไทย ไทยโดยใช้วิธีการเรียนรู้ของเครื่อง
โดย	นายภัทรพล ภัทรอัมพรชัย
สาขาวิชา	วิศวกรรมอุตสาหการ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.นันทชัย กานตานันทะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

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

ภัทรพล ภัทรอัมพรชัย : การพยากรณ์ราคาเม็ดพลาสติกพอลิโพรพิลีนในประเทศไทย
โดยใช้วิธีการเรียนรู้ของเครื่อง. (FORECASTING OF POLYPROPYLENE GRANULE
PRICE IN THAILAND USING MACHINE LEARNING) อ.ที่ปรึกษาหลัก : ผศ. ดร.นันท
ชัย กานตานันทะ

ในปัจจุบันธุรกิจอุตสาหกรรมพลาสติกมีความสำคัญมากกับประเทศไทยเนื่องจากสินค้า
อุตสาหกรรมส่วนใหญ่ใช้พลาสติก โดยจากสถิติปริมาณการใช้งานพลาสติกพบว่าพลาสติกโพลีโพร
พิลีน (PP) มีปริมาณการใช้งานมากที่สุดเมื่อเทียบกับพลาสติกชนิดอื่น ทำให้เป็นพลาสติกชนิดที่มี
ความสำคัญที่สุดต่ออุตสาหกรรมพลาสติกในไทย งานวิจัยฉบับนี้จัดทำเพื่อเป็นการนำเสนอและ
เปรียบเทียบรูปแบบการพยากรณ์สำหรับราคาเม็ดพลาสติก PP ในประเทศไทย จุดประสงค์ของ
งานวิจัยนี้เพื่อหาตัวแบบการพยากรณ์ราคาเม็ดพลาสติก PP ที่ให้ผลลัพธ์ที่แม่นยำที่สุด โดยใช้ค่า
ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของตัว
แบบพยากรณ์ ซึ่งตัวแบบพยากรณ์อนุกรมเวลาที่ใช้ในการทำวิจัยครั้งนี้ได้แก่ ตัวแบบพยากรณ์ปรับ
เรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์ ตัวแบบพยากรณ์ SARIMA ตัวแบบพยากรณ์เชิง
สาเหตุที่ใช้คือ ตัวแบบพยากรณ์ถดถอยเชิงเส้นพหุคูณ ตัวแบบพยากรณ์เทคนิคการเรียนรู้ด้วย
เครื่องที่ใช้คือ ตัวแบบพยากรณ์ Support Vector Regression ตัวแบบพยากรณ์ Decision Tree
ตัวแบบพยากรณ์ Random Forest ตัวแบบพยากรณ์ XGBoost ตัวแบบพยากรณ์ Artificial
Neural Networks และตัวแบบพยากรณ์ผสมที่เกิดจากการนำตัวแบบพยากรณ์เดี่ยวที่มีความ
แม่นยำสูงมาจับคู่กัน โดยในงานวิจัยนี้ได้ใช้ข้อมูลราคาเม็ดพลาสติก PP ตั้งแต่ เดือนมกราคม พ.ศ.
2554 ถึง เดือนธันวาคม พ.ศ. 2565 จากผลการศึกษาพบว่าตัวแบบพยากรณ์ผสม SARIMA –
ANN มีความแม่นยำสูงที่สุดและมีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยในช่วงชุดข้อมูล
ตรวจสอบอยู่ที่ 5.54% และเมื่อนำมาประเมินประสิทธิภาพกับชุดข้อมูลทดสอบจะได้ค่าร้อยละ
ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเท่ากับ 6.92%

สาขาวิชา วิศวกรรมอุตสาหกรรม
ปีการศึกษา 2565

ลายมือชื่อนิสิต
ลายมือชื่อ อ.ที่ปรึกษาหลัก

6270207521 : MAJOR INDUSTRIAL ENGINEERING

KEYWORD: Polypropylene Granule Price, Hybrid Forecasting Model, Machine Learning Forecasting Model

Pattarapol Pattaraumpornchai : FORECASTING OF POLYPROPYLENE GRANULE PRICE IN THAILAND USING MACHINE LEARNING. Advisor: Asst. Prof. NANTACHAI KANTANANTHA, Ph.D.

Plastic is vital to Thailand's economy due to its widespread use in various merchandise industries. According to the statistics of plastic consumption in Thailand, the Polypropylene (PP) has the highest consumption quantity compared to other plastic types which makes it becomes the most important plastic to the plastic industry in Thailand. This research is conducted to present and compare the forecasting models of PP granule price in Thailand. The objective of this research is to determine the most accurate forecasting model of PP granule price by using mean absolute percentage error (MAPE) as a criterion to compare the efficiency of the forecasting models. The forecasting models in this research are Holt - Winters exponential smoothing model, SARIMA model, multiple regression model, Support Vector Regression Model, Decision Tree Model, Random Forest Model, XGBoost Model, Artificial Neural Networks Model and Hybrid Forecasting Model. This research uses the data of PP granule price from January 2011 to December 2022. The results show that SARIMA – ANN Model have the highest accuracy with the value of MAPE equal to 5.54% at CV Data Set and when evaluating with the Testing Data Set, the MAPE is equal to 6.92%.

Field of Study: Industrial Engineering

Student's Signature

Academic Year: 2022

Advisor's Signature

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบพระคุณผู้ช่วยศาสตราจารย์ ดร. นันทชัย กานตานันทะ อาจารย์ที่ปรึกษาวิทยานิพนธ์เป็นอย่างสูงที่กรุณาสละเวลาคอยให้คำแนะนำปรึกษา ให้ความช่วยเหลือ ชี้แนะข้อบกพร่องข้อปรับปรุงเพื่อทำการแก้ไขข้อบกพร่องต่าง ๆ รวมถึงได้ให้คำแนะนำอันเป็นประโยชน์รวมถึงช่วยตรวจสอบความสมบูรณ์ของงานวิทยานิพนธ์ฉบับนี้

ข้าพเจ้าขอขอบพระคุณรองศาสตราจารย์ ดร. ดาริชา สุธีวงศ์ ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. นระเกณท์ พุ่มชูศรี กรรมการสอบ และรองศาสตราจารย์ ดร. จันทร์ศิริ สิงห์เถื่อน กรรมการภายนอกมหาวิทยาลัย ที่ได้สละเวลามาเป็นคณะกรรมการสอบวิทยานิพนธ์ของข้าพเจ้าและให้คำแนะนำ แนวทางแก้ไขข้อบกพร่องในวิทยานิพนธ์ฉบับนี้ เพื่อให้มีความสมบูรณ์มากยิ่งขึ้น

ข้าพเจ้าขอกราบขอบพระคุณคณะอาจารย์ภาควิชาวิศวกรรมอุตสาหการ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยทุกท่าน ที่ได้ถ่ายทอดวิชาวิชาความรู้และคำแนะนำ ซึ่งสามารถช่วยให้ข้าพเจ้าทำวิทยานิพนธ์ฉบับนี้สำเร็จไปได้ด้วยดี

ข้าพเจ้าขอขอบพระคุณครอบครัวและคนรักที่ให้การสนับสนุน ให้กำลังใจในการทำวิทยานิพนธ์ของข้าพเจ้า และเพื่อน ๆ ที่คอยให้คำปรึกษาแนะนำ ให้กำลังใจ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภัทรพล ภัทรอัมพรชัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ค
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญ.....	ฉ
สารบัญตาราง.....	ฎ
สารบัญรูปภาพ.....	ฏ
บทที่ 1 บทนำ	1
1.1 ที่มาและความสำคัญ.....	1
1.1.1 สถานการณ์อุตสาหกรรมการส่งออกของประเทศไทย	1
1.1.2 สถานการณ์อุตสาหกรรมพลาสติกของประเทศไทย	2
1.1.3 การประยุกต์ใช้งานของพลาสติกแต่ละชนิด	3
1.2 ปัญหาที่พบ	4
1.3 วัตถุประสงค์ของงานวิจัย	6
1.4 ขอบเขตของงานวิจัย.....	6
1.5 ผลลัพธ์ที่คาดว่าจะได้รับ.....	7
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	7
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	8
2.1 ทฤษฎีที่เกี่ยวข้อง	8
2.1.1 การพยากรณ์.....	8
2.1.2 ช่วงระยะเวลาการพยากรณ์.....	8
2.1.3 ขั้นตอนในการพยากรณ์.....	8

2.1.4 ประเภทของการพยากรณ์เชิงสถิติ	10
2.1.4.1 การพยากรณ์เชิงคุณภาพ (Qualitative Forecasting Techniques)	10
2.1.4.2 การพยากรณ์เชิงปริมาณ (Quantitative Forecasting Techniques)	10
2.1.5 วิธีการพยากรณ์ข้อมูลเชิงสาเหตุ (Casual Method).....	10
2.1.5.1 วิธีการวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis).....	10
2.1.5.2 วิธีการวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Regression Analysis)...	11
2.1.6 วิธีการพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data Method).....	11
2.1.6.1 ข้อมูลอนุกรมเวลา (Time Series Data).....	11
2.1.6.2 รูปแบบของอนุกรมเวลา.....	13
2.1.6.3 วิธีปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์ (Holt - Winters Exponential Smoothing Method).....	15
2.1.6.4 วิธีบ็อกซ์-เจนกินส์ (Box-Jenkins Method : ARIMA Model)	16
2.1.7 การวัดความแม่นยำของตัวแบบพยากรณ์	17
2.1.8 การเรียนรู้ของเครื่อง (Machine Learning).....	18
2.1.9 Supervised Learning	20
2.1.9.1 Linear Regression.....	20
2.1.9.2 Support Vector Regression (SVR).....	24
2.1.9.3 Ensemble Learning.....	26
2.1.9.4 ต้นไม้ตัดสินใจ (Decision Tree).....	27
2.1.9.5 ป่าสุ่ม (Random Forests).....	29
2.1.9.6 XGBoost (eXtreme Gradient Boosting)	29
2.1.9.7 โครงข่ายประสาทเทียม (Artificial Neural Network).....	30
2.1.9.8 Hyperparameter and Model Validation	32
2.2 งานวิจัยที่เกี่ยวข้อง.....	35

2.2.1 งานวิจัยการพยากรณ์ราคาเม็ดพลาสติกชนิดอื่น	35
2.2.2 งานวิจัยเกี่ยวกับน้ำมันปาล์ม.....	38
2.2.3 งานวิจัยเกี่ยวกับน้ำมันดิบ.....	39
2.2.4 งานวิจัยเกี่ยวกับยาง	42
บทที่ 3 ขั้นตอนการดำเนินงาน.....	44
3.1 รวบรวมข้อมูลและกำหนดขอบเขตการวิจัย	44
3.2 วิเคราะห์ข้อมูล.....	46
3.3 สร้างตัวแบบการพยากรณ์จากชุดข้อมูลฝึกสอน	49
3.3.1 วิธีการพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data Method).....	49
3.3.1.1 วิธีปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์ (Holt – Winters Exponential Smoothing Method).....	50
3.3.1.2 วิธี Seasonal Autoregressive Integrated Moving Average.....	50
3.3.2 วิธีการพยากรณ์ข้อมูลเชิงสาเหตุ (Casual Method).....	51
3.3.3 วิธีพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง	52
3.3.3.1 การปรับไฮเปอร์พารามิเตอร์ของตัวแบบการเรียนรู้ของเครื่อง.....	52
3.3.3.2 วิธีพยากรณ์ด้วยตัวแบบ Support Vector Regression (SVR).....	53
3.3.3.3 วิธีพยากรณ์ด้วยตัวแบบต้นไม้การตัดสินใจ (Decision Tree).....	54
3.3.3.4 วิธีพยากรณ์ด้วยตัวแบบป่าสุ่ม (Random Forest).....	55
3.3.3.5 วิธีพยากรณ์ด้วยตัวแบบ XGBoost.....	56
3.3.3.6 วิธีพยากรณ์ด้วยตัวแบบโครงข่ายประสาทเทียม (Artificial Neural Network: ANN).....	58
3.3.4 วิธีพยากรณ์ด้วยตัวแบบพยากรณ์ผสม (Hybrid Forecasting Model).....	59
3.4 การเปรียบเทียบความแม่นยำของตัวแบบการพยากรณ์ โดยเทียบกับชุดข้อมูล Cross Validation Data Set.....	63
3.5 การวัดความแม่นยำกับข้อมูล Testing Data Set	63

บทที่ 4 ผลการศึกษา.....	64
4.1 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ Holt-Winters	64
4.2 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SARIMA	65
4.3 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์วิเคราะห์การถดถอยเชิงเส้นพหุคูณ.....	68
4.4 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ Support Vector Regression.....	71
4.5 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ Decision Tree	72
4.6 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ Random Forest.....	74
4.9 เปรียบเทียบผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์เดี่ยว.....	78
4.10 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ผสม.....	78
4.10.1 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SARIMA - SVR.....	79
4.10.2 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SARIMA – XGBoost.....	81
4.10.3 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SARIMA – ANN	82
4.10.4 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SVR - SARIMA.....	84
4.10.5 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ XGBoost - SARIMA	85
4.10.6 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ ANN - SARIMA.....	87
4.10.7 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ ANN - XGBoost	89
4.10.8 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ ANN - SVR.....	91
4.10.9 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ XGBoost - ANN.....	93
4.10.10 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ XGBoost - SVR.....	95
4.10.11 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SVR - ANN	97
4.10.12 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SVR - XGBoost.....	99
4.11 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์.....	101
4.11.1 วิเคราะห์ผลลัพธ์การพยากรณ์ของกลุ่มตัวแบบพยากรณ์หลัก SARIMA	103
4.11.2 วิเคราะห์ผลลัพธ์การพยากรณ์ของกลุ่มตัวแบบพยากรณ์รอง SARIMA	103

4.11.3 วิเคราะห์ผลลัพธ์การพยากรณ์ของกลุ่มตัวแบบพยากรณ์หลัก และตัวแบบพยากรณ์ รองจากตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่อง	104
4.12 ผลลัพธ์การพยากรณ์ในชุดข้อมูลทดสอบ (Testing Set).....	106
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	107
5.1 สรุปผลการวิจัย	107
5.2 สรุปผลการวิจัยตัวแบบพยากรณ์อนุกรมเวลา.....	107
5.3 สรุปผลการวิจัยตัวแบบพยากรณ์เชิงสาเหตุ.....	108
5.4 สรุปผลการวิจัยตัวแบบพยากรณ์เทคนิคการเรียนรู้ด้วยเครื่อง	108
5.5 สรุปผลการวิจัยตัวแบบพยากรณ์แบบผสม	108
5.6 ข้อเสนอแนะเพิ่มเติม.....	109
บรรณานุกรม.....	110
ประวัติผู้เขียน.....	114

สารบัญตาราง

หน้า

ตารางที่ 1 สถิติสัดส่วนโครงสร้างการส่งของไทยตั้งแต่ พ.ศ. 2556 ถึง พ.ศ. 2563	1
ตารางที่ 2 ผลลัพธ์ความแม่นยำจากการพยากรณ์จากงานวิจัยของปฎิมาพร (2556)	37
ตารางที่ 3 ปัจจัยที่ถูกนำมาพิจารณาในแต่ละงานวิจัยจากบทที่ 2	45
ตารางที่ 4 ข้อมูลราคาและปัจจัยที่เกี่ยวข้องกับราคาเม็ดพลาสติก PP	46
ตารางที่ 5 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ SVR	53
ตารางที่ 6 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ Decision Tree ...	54
ตารางที่ 7 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ Random Forest	55
ตารางที่ 8 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ XGBoost	57
ตารางที่ 9 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ ANN	58
ตารางที่ 10 สรุปการแบ่งกลุ่มในการสร้างตัวแบบพยากรณ์ผสม	60
ตารางที่ 11 ค่าพารามิเตอร์ของตัวแบบพยากรณ์ Holt-Winters และค่า MAPE	64
ตารางที่ 12 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ SVR	72
ตารางที่ 13 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ Decision Tree	73
ตารางที่ 14 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ Random Forest	74
ตารางที่ 15 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ XGBoost	76
ตารางที่ 16 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ ANN	77
ตารางที่ 17 เปรียบเทียบผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์เดี่ยว	78
ตารางที่ 18 สรุปการแบ่งกลุ่มในการสร้างตัวแบบพยากรณ์ผสม	78
ตารางที่ 19 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SARIMA - SVR และค่า MAPE	80

ตารางที่ 20 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SARIMA – XGBoost และค่า MAPE	81
ตารางที่ 21 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SARIMA – ANN และค่า MAPE	83
ตารางที่ 22 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SVR - SARIMA และค่า MAPE	84
ตารางที่ 23 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ XGBoost - SARIMA และค่า MAPE	86
ตารางที่ 24 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ ANN - SARIMA และค่า MAPE	88
ตารางที่ 25 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ ANN - XGBoost และค่า MAPE	90
ตารางที่ 26 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ ANN - SVR และค่า MAPE ...	92
ตารางที่ 27 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ XGBoost - ANN และค่า MAPE	94
ตารางที่ 28 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ XGBoost - SVR และค่า MAPE	96
ตารางที่ 29 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SVR - ANN และค่า MAPE ...	98
ตารางที่ 30 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SVR - XGBoost และค่า MAPE	100
ตารางที่ 31 สรุปค่า MAPE ของแต่ละตัวแบบพยากรณ์ของราคาเม็ดพลาสติก PP ในประเทศไทย	102
ตารางที่ 32 เปรียบเทียบค่า MAPE ระหว่างตัวแบบพยากรณ์เดี่ยวและตัวแบบพยากรณ์ผสม	103
ตารางที่ 33 เปรียบเทียบผลลัพธ์การพยากรณ์ระหว่างกลุ่มตัวแบบพยากรณ์หลัก SARIMA และตัวแบบพยากรณ์เดี่ยว	103
ตารางที่ 34 เปรียบเทียบผลลัพธ์การพยากรณ์ระหว่างกลุ่มตัวแบบพยากรณ์รอง SARIMA และตัวแบบพยากรณ์เดี่ยว	104

ตารางที่ 35 เปรียบเทียบผลลัพธ์การพยากรณ์ของกลุ่มตัวแบบพยากรณ์หลัก และตัวแบบพยากรณ์ รองจากตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่อง	104
ตารางที่ 36 ผลลัพธ์ค่าความแม่นยำของตัวแบบพยากรณ์ SARIMA – ANN.....	106



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญรูปภาพ

หน้า

รูปที่ 1 ดัชนีผลผลิตของอุตสาหกรรมเม็ดพลาสติกตั้งแต่เดือนมกราคม พ.ศ. 2554 ถึงเดือนเมษายน พ.ศ. 2564 (ที่มา : http://www.oie.go.th).....	2
รูปที่ 2 สถิติการใช้งานเม็ดพลาสติกแต่ละชนิดตั้งแต่ พ.ศ. 2555 ถึง พ.ศ. 2562.....	3
รูปที่ 3 สถิติราคาเม็ดพลาสติก PP ในประเทศตั้งแต่เดือนมกราคม พ.ศ. 2548 ถึง เดือนมิถุนายน พ.ศ. 2564	5
รูปที่ 4 ข้อมูลที่มีลักษณะของแนวโน้มและวัฏจักร	12
รูปที่ 5 ข้อมูลที่มีลักษณะเป็นฤดูกาล.....	13
รูปที่ 6 ข้อมูลอนุกรมเวลารูปแบบการบวก.....	14
รูปที่ 7 ข้อมูลอนุกรมเวลารูปแบบการคูณ.....	14
รูปที่ 8 ลักษณะของโครงข่ายประสาทเทียม (Hurwitz, 2018).....	19
รูปที่ 9 ภาพรวมของปัญญาประดิษฐ์ (Hurwitz, 2018).....	19
รูปที่ 10 รูปแบบชุดข้อมูลของ Training ของ Supervised Learning ที่มี Label (Géron, 2019). 20	
รูปที่ 11 หลักการทำงานของ Gradient descent (Géron, 2019).....	22
รูปที่ 12 การทำงานของ Gradient Descent ที่ Learning Rate ต่ำเกินไป (Géron, 2019).....	22
รูปที่ 13 การทำงานของ Gradient Descent ที่ Learning Rate สูงเกินไป (Géron, 2019)	23
รูปที่ 14 Cost Function ที่มี Local Minimum และ Global Minimum (Géron, 2019).....	23
รูปที่ 15 หลักการทำงานของ Support Vector Machine (ชิตพงษ์, 2563).....	24
รูปที่ 16 การเปรียบเทียบผลลัพธ์ของตัวแบบ SVR ที่ค่า ϵ ต่างกัน (Géron, 2019).....	25
รูปที่ 17 หลักการทำงานของ Bagging (Géron, 2019)	26
รูปที่ 18 หลักการทำงานของ Boosting.....	27
รูปที่ 19 หลักการทำงานของ Decision Tree (Géron, 2019)	28
รูปที่ 20 การเปรียบเทียบผลลัพธ์ของตัวแบบ Decision Tree ที่ค่า Max Depth ต่างกัน.....	28

รูปที่ 21 หลักการทำงานของ Random Forest	29
รูปที่ 22 หลักการทำงานของ XGBoost.....	30
รูปที่ 23 ส่วนประกอบและโครงสร้างของตัวแบบพยากรณ์โครงข่ายประสาทเทียม	31
รูปที่ 24 ตัวอย่างรูปแบบของ k-fold Cross-validation ที่ $k = 5$	33
รูปที่ 25 การเปรียบเทียบระหว่างตัวแบบพยากรณ์ที่มีค่า Bias สูง และตัวแบบที่มีค่า Variance สูง	34
รูปที่ 26 กราฟแสดงจุดสมดุลระหว่าง Bias และ Variance	34
รูปที่ 27 ราคาเม็ดพลาสติก PP ในประเทศตั้งแต่เดือนมกราคม พ.ศ. 2554.....	47
รูปที่ 28 ราคาเม็ดพลาสติก PP รายเดือนแยกตามปี ตั้งแต่ พ.ศ. 2554 ถึง พ.ศ. 2560.....	47
รูปที่ 29 โครงสร้างองค์ประกอบข้อมูลราคาเม็ดพลาสติก PP	48
รูปที่ 30 กราฟฟังก์ชันอัตสหสัมพันธ์ (ACF) ของราคาเม็ดพลาสติก PP.....	49
รูปที่ 31 ขั้นตอนของวิธีพยากรณ์ด้วยตัวแบบพยากรณ์ผสม	62
รูปที่ 32 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ตรวจสอบของตัวแบบพยากรณ์ Holt-Winters	65
รูปที่ 33 ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบ SARIMA ของราคาเม็ดพลาสติก PP.....	66
รูปที่ 34 ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบ SARIMA ของราคาเม็ดพลาสติก PP.....	67
รูปที่ 35 การวิเคราะห์ตรวจสอบค่าความคลาดเคลื่อนของตัวแบบพยากรณ์ SARIMA (Diagnostic Checking).....	67
รูปที่ 36 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ตรวจสอบของตัวแบบพยากรณ์ SARIMA	68
รูปที่ 37 ค่าสัมประสิทธิ์และ VIF ของแต่ละปัจจัยจากโปรแกรม Minitab	69
รูปที่ 38 ผลลัพธ์การทดสอบสมมติฐานการถดถอยของราคาเม็ดพลาสติก PP.....	70
รูปที่ 39 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ตรวจสอบของตัวแบบพยากรณ์วิเคราะห์การถดถอยเชิงเส้นพหุคูณ.....	71
รูปที่ 40 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ตรวจสอบของตัวแบบพยากรณ์ SVR.....	72

รูปที่ 54 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ตรวจสอบของตัวแบบพยากรณ์ XGBoost - SVR.....	97
รูปที่ 55 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ตรวจสอบของตัวแบบพยากรณ์ SVR - ANN	99
รูปที่ 56 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ตรวจสอบของตัวแบบพยากรณ์ SVR - XGBoost.....	101
รูปที่ 57 เปรียบเทียบความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SVR-NN	105
รูปที่ 58 เปรียบเทียบความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SVR-XG	105
รูปที่ 59 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ทดสอบของตัวแบบพยากรณ์ SARIMA - ANN	106



บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

ปัจจุบันอุตสาหกรรมพลาสติกมีความสำคัญมากเนื่องจากเป็นวัตถุดิบที่ใช้กันทั่วโลก และมีการนำไปผลิตเป็นสินค้าหลายชนิด ซึ่งวัตถุดิบเริ่มต้นของพลาสติกคือเม็ดพลาสติก โดยราคาเม็ดพลาสติกไม่คงที่ มีการเปลี่ยนแปลงตามปัจจัยที่ไม่แน่นอนหากสามารถทราบรูปแบบราคาเม็ดพลาสติกได้จะสามารถทำให้ผู้ผลิตสินค้าลดภาระค่าใช้จ่ายจากการสั่งซื้อเม็ดพลาสติกได้

1.1.1 สถานการณ์อุตสาหกรรมการส่งออกของประเทศไทย

จากสถิติสัดส่วนโครงสร้างการส่งออกของไทยในปี พ.ศ. 2563 พบว่าแต่ละหมวดมีสัดส่วนในการส่งออกเรียงลำดับจากมากไปน้อยดังนี้ หมวดสินค้าอุตสาหกรรม 80.23% หมวดสินค้าเกษตรกรรม 9.14% หมวดสินค้าอุตสาหกรรมเกษตร 7.8% และหมวดสินค้าแร่และเชื้อเพลิง 2.83% ซึ่งหมวดที่มีสัดส่วนมากที่สุดคือหมวดสินค้าอุตสาหกรรมและจากตารางที่ 1 พบว่าสัดส่วนการส่งออกของหมวดสินค้าอุตสาหกรรมนั้นมากที่สุดตั้งแต่ พ.ศ. 2556 จนถึง พ.ศ. 2563 โดยตัวอย่างสินค้าในหมวดนี้ได้แก่ เครื่องคอมพิวเตอร์ รถยนต์ เม็ดพลาสติก แผงวงจรไฟฟ้า เป็นต้น ซึ่งสินค้าเหล่านี้ล้วนมีการใช้เม็ดพลาสติกเป็นส่วนประกอบหรือเป็นวัตถุดิบทั้งสิ้น ทำให้เม็ดพลาสติกมีความสำคัญมากต่อการส่งออกของประเทศไทย

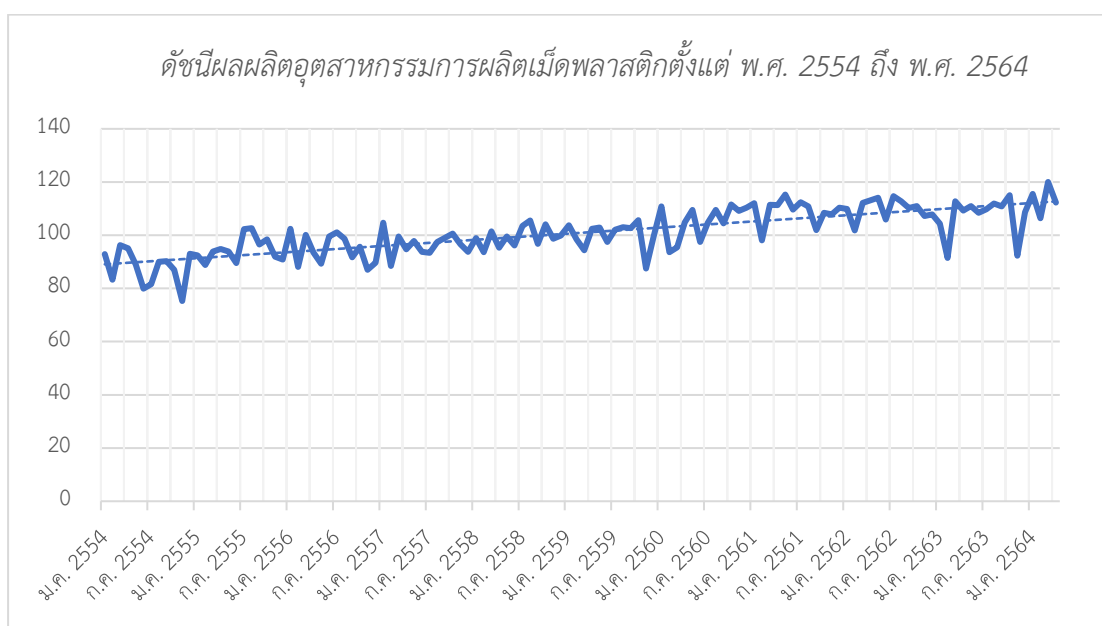
ตารางที่ 1 สถิติสัดส่วนโครงสร้างการส่งของไทยตั้งแต่ พ.ศ. 2556 ถึง พ.ศ. 2563

สินค้า	สัดส่วน (%)							
	2556	2557	2558	2559	2560	2561	2562	2563
รวมทุกสินค้า	100	100	100	100	100	100	100	100
1. สินค้าเกษตรกรรม (กลีกรวม, ปศุสัตว์, ประมง)	9.95	9.84	9.41	9.01	9.71	9.18	8.85	9.14
2. สินค้าอุตสาหกรรมเกษตร	7.55	7.48	7.76	7.76	7.21	7.21	7.62	7.80
3. สินค้าอุตสาหกรรม	76.04	77.21	78.68	79.95	79.43	79.21	79.9	80.23
4. สินค้าแร่และเชื้อเพลิง	6.46	5.48	4.15	3.10	3.66	4.39	3.63	2.83

ที่มา : <http://tradereport.moc.go.th>

1.1.2 สถานการณ์อุตสาหกรรมพลาสติกของประเทศไทย

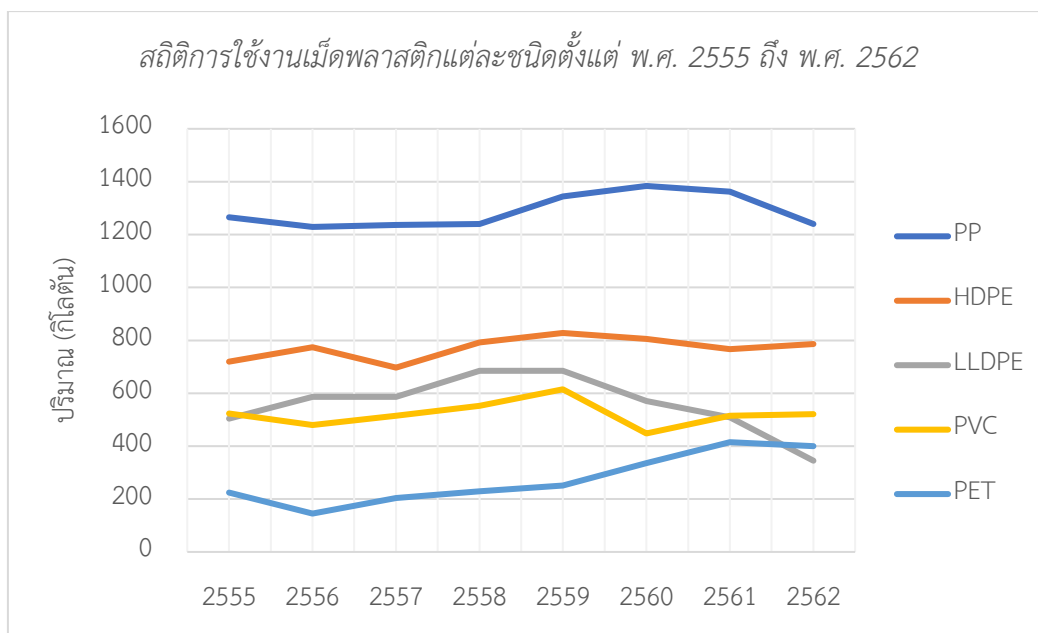
ดัชนีผลผลิตของอุตสาหกรรมเม็ดพลาสติกตั้งแต่เดือนมกราคม พ.ศ. 2554 จนถึงเดือนเมษายน พ.ศ. 2564 มีแนวโน้มเพิ่มสูงขึ้นดังแสดงในรูปที่ 1 ซึ่งสามารถบ่งชี้ได้ว่าทางภาคอุตสาหกรรมมีความต้องการใช้เม็ดพลาสติกเป็นวัตถุดิบในการผลิตเพิ่มมากขึ้น โดยได้ทำการรวบรวมข้อมูลดังกล่าวจากสำนักงานเศรษฐกิจอุตสาหกรรม กระทรวงอุตสาหกรรม



รูปที่ 1 ดัชนีผลผลิตของอุตสาหกรรมเม็ดพลาสติกตั้งแต่เดือนมกราคม พ.ศ. 2554 ถึงเดือนเมษายน พ.ศ. 2564 (ที่มา : <http://www.oie.go.th>)

CHULALONGKORN UNIVERSITY

ในรูปที่ 1-2 แสดงถึงสถิติปริมาณการใช้งานเม็ดพลาสติก 5 อันดับแรกในประเทศไทยตั้งแต่ พ.ศ. 2555 ถึง พ.ศ. 2562 ที่รวบรวมจากสถาบันพลาสติก ได้แก่ PP (Polypropylene), HDPE (High-Density Polyethylene), LLDPE (Linear Low-density Polyethylene), PVC (Polyvinyl Chloride) และ PET (Polyethylene Terephthalate) โดยแกนในแนวตั้งคือปริมาณการใช้งานในหน่วยตัน ส่วนแกนในแนวนอนคือปี โดยสามารถสรุปจากรูปที่ 2 ได้ว่า PP มีปริมาณการใช้งานต่อปีสูงที่สุด



รูปที่ 2 สถิติการใช้งานเม็ดพลาสติกแต่ละชนิดตั้งแต่ พ.ศ. 2555 ถึง พ.ศ. 2562
(ที่มา : สถาบันพลาสติก)

1.1.3 การประยุกต์ใช้งานของพลาสติกแต่ละชนิด

จากรูปที่ 1-2 เมื่อนำพลาสติกทั้ง 5 ชนิดมาพิจารณาถึงประโยชน์และการนำไปใช้สามารถสรุปได้ดังนี้

1.) Polypropylene (PP) มีคุณสมบัติแข็ง เหนียว ทนต่อการหักงอ ทนความร้อน (จุดหลอมเหลวอยู่ที่ 165 องศาเซลเซียส) สามารถทนอุณหภูมิในการฆ่าเชื้อ (สามารถผ่านกระบวนการสเตอริไลซ์ได้ในอุณหภูมิ 100 องศาเซลเซียส) ทนต่อสารเคมี ด้วยคุณสมบัติดังกล่าวพลาสติก PP จึงสามารถนำมาผลิตเป็นพลาสติกสำหรับใช้งานในทางการแพทย์ ผลิตเป็นภาชนะบรรจุอาหาร เครื่องมือ กระจา เป่า ปกแพ้มเอกสาร ตลับเครื่องสำอาง ถังร้อน ฟิล์มใส กระจาอบข้าว ถังบรรจุปุ๋ย พลาสติกหุ้มซองบุหรี่ เชือก แห อวน และยังสามารถนำมาผลิตเป็นกล่องแบตเตอรี่และชิ้นส่วนรถยนต์

2.) Polyethylene Terephthalate (PET) มีคุณสมบัติเหนียวแต่ยืดหยุ่น ทำให้ทนทานต่อแรงกระแทกและไม่แตกหักง่าย ด้วยคุณสมบัติดังกล่าวทำให้พลาสติก PET นิยมนำมาผลิตเป็นขวดน้ำอัดลม น้ำดื่ม น้ำผลไม้ เครื่องดื่มแอลกอฮอล์ และสามารถนำไปผลิตเป็นใยสังเคราะห์ เสื้อผ้า กระจา เป่า

3.) High Density Polyethylene (HDPE) มีคุณสมบัติแข็งแรง ยืดหยุ่น มีสีขาวขุ่น โปร่งแสง สามารถนำมาผลิตเป็นบรรจุภัณฑ์พลาสติก เช่น ขวด ถัง ถาด ถู ที่ต้องการความแข็งแรง และสามารถนำมาแปรรูปได้เป็นขวดน้ำมันเครื่อง ท่อ ลัง

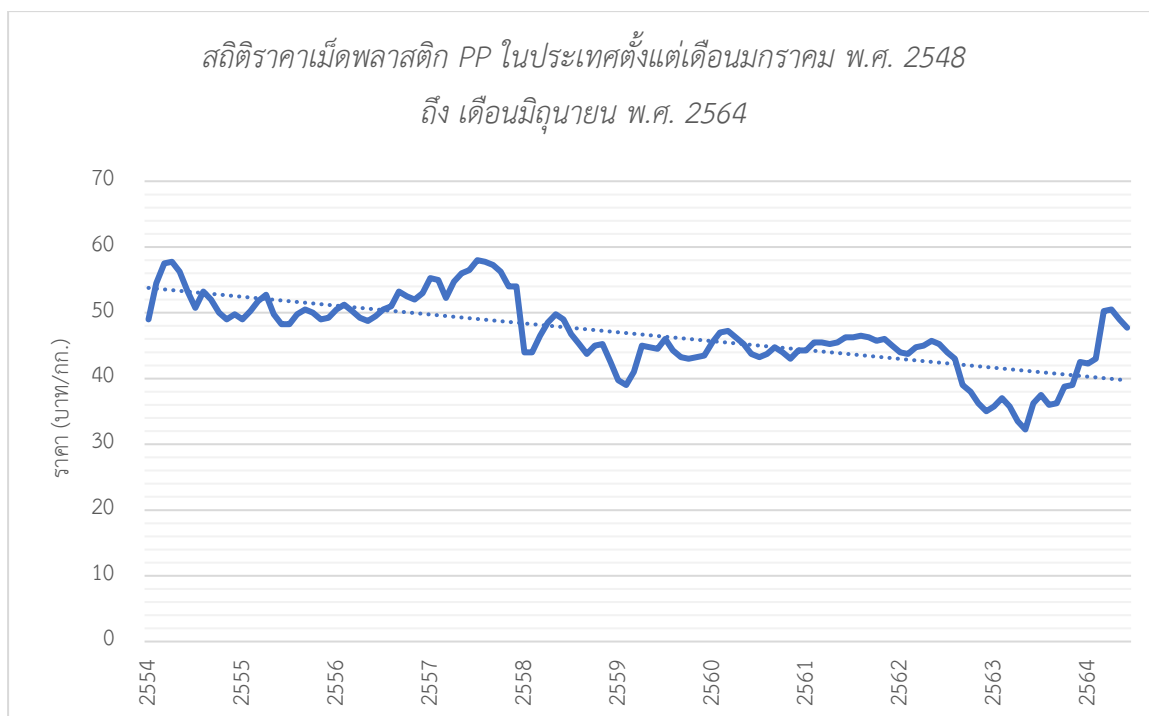
4.) Linear Low Density Polyethylene (LLDPE) มีคุณสมบัติอยู่ระหว่าง LDPE และ HDPE แต่จะนิ่ม และเหนียวกว่า LDPE และ HDPE (ในส่วนของ LDPE มีคุณสมบัติ ได้แก่ ความหนาแน่นต่ำ มีความทนทานต่อสารเคมีที่เป็นกรดและด่าง สามารถนำไปแปรรูปได้ง่ายเป็นฉนวนได้ และไม่มีสารพิษที่เป็นอันตรายต่อร่างกายมนุษย์ รวมถึงความสามารถในการรับแรงกระแทก และมีความยืดหยุ่นได้ดี) โดยสามารถนำ LLDPE มาผลิตเป็น ถูบรรจุสิ่งของที่มีน้ำหนักสูง ท่อน้ำ เคลือบสายไฟ และของเด็กเล่น นอกจากนี้ยังนิยมนำ LLDPE มาผลิตสินค้าใช้งานที่อุณหภูมิต่ำ เช่นบรรจุอาหารแช่เย็น หรือ แช่แข็ง

5.) Polyvinylchloride (PVC) มีคุณสมบัติดังนี้ มีความยืดหยุ่นสูง แข็งแต่เปราะ เมื่อตัดไฟจะดับได้ด้วยตัวเอง ป้องกันไขมันได้ และสลายตัวได้ง่ายเมื่อได้รับความร้อน นิยมนำมาผลิตเป็นท่อน้ำ ท่อพีวีซี ข้อต่อ ฉนวนหุ้มสายไฟ สายเคเบิล แผ่นพลาสติก ฟิล์ม หนังสติ๊ก รองเท้า บัตรเครดิต อุปกรณ์รถยนต์ ขวดพลาสติก ของเล่น

จากข้อมูลคุณสมบัติของพลาสติกแต่ละชนิดสามารถสรุปได้ว่า Polypropylene (PP) มีการใช้กับผลิตภัณฑ์หลายชนิดตั้งแต่บรรจุภัณฑ์ ของใช้ทั่วไป ชิ้นส่วนรถยนต์ จนถึงอุปกรณ์ทางการแพทย์ ต่างกับเม็ดพลาสติกชนิดอื่นที่การใช้งานค่อนข้างเจาะจงกับผลิตภัณฑ์บางชนิด ดังนั้น PP จึงเป็นพลาสติกชนิดที่มีความสำคัญมากกับอุตสาหกรรมการผลิตและการส่งออกในประเทศไทยเพราะหลายผลิตภัณฑ์จากหลายอุตสาหกรรมล้วนใช้เม็ดพลาสติก PP และมีปริมาณการใช้งานต่อปีสูงที่สุดเมื่อเทียบกับพลาสติกชนิดอื่น จากสาเหตุดังกล่าวทำให้งานวิจัยฉบับนี้เลือกใช้เม็ดพลาสติก PP ในการนำมาพยากรณ์ราคา

1.2 ปัญหาที่พบ

จากดัชนีการผลิตเม็ดพลาสติกที่เพิ่มสูงขึ้นเป็นการบ่งบอกว่าความต้องการพลาสติกเพิ่มมากขึ้นในทุกปีทำให้ผู้ประกอบการต้องมีการสั่งซื้อเม็ดพลาสติกเพิ่มมากขึ้นเพื่อรองรับกับความต้องการ และราคาเม็ดพลาสติก PP มีความผันผวนไม่คงที่ดังรูปที่ 3 การที่จะสั่งซื้อมาทำการผลิตหากสามารถสั่งซื้อถูกจังหวะจะสามารถทำให้ลดค่าใช้จ่ายจากส่วนต่างราคาได้ เนื่องจากปริมาณการสั่งซื้อต่อครั้งจะอยู่ที่ประมาณ 10 – 100 ตัน ดังนั้นปัญหาส่วนต่างราคาจะก่อให้เกิดค่าเสียโอกาสคิดเป็นมูลค่าได้ตั้งแต่หลักหมื่นถึงหลักล้านบาทขึ้นกับจังหวะและปริมาณที่สั่งซื้อ



รูปที่ 3 สถิติราคาเม็ดพลาสติก PP ในประเทศตั้งแต่เดือนมกราคม พ.ศ. 2548
ถึง เดือนมิถุนายน พ.ศ. 2564

ในปัจจุบันยังไม่มี การสร้างตัวแบบการพยากรณ์ราคาเม็ดพลาสติกโพลิโพรพิลีนในประเทศไทยโดยใช้วิธีการเรียนรู้ของเครื่อง ซึ่งตัวอย่างงานวิจัยเกี่ยวกับการพยากรณ์ราคาเม็ดพลาสติกในประเทศไทยจะใช้วิธีการถดถอยเชิงพหุคูณ (จากงานวิจัยของวรวรรณิภา และคณะ) และวิธี ARIMA (จากงานวิจัยของ Adisak and Vichai) โดยรายละเอียดจะอยู่ในบทที่ 2 ในส่วนของงานวิจัยที่เกี่ยวข้อง เนื่องจากการเรียนรู้ด้วยเครื่องสามารถนำมาใช้สร้างตัวแบบการพยากรณ์ที่ให้ความแม่นยำมากขึ้นได้ อีกทั้งเพื่อให้ผู้ที่สนใจหรือบริษัทที่ต้องสั่งซื้อเม็ดพลาสติกสามารถนำไปประยุกต์ใช้ในการวางแผนการสั่งซื้อเม็ดพลาสติก PP ได้ และสุดท้ายเพื่อเป็นการนำทฤษฎีการเรียนรู้ด้วยเครื่องที่กำลังเป็นที่นิยมในปัจจุบันมาสร้างตัวแบบพยากรณ์โดยทำการรวมหลายตัวแบบของการเรียนรู้ด้วยเครื่องเข้าด้วยกันมาสร้างตัวแบบการพยากรณ์แบบผสม (Hybrid Forecasting Model) ทำให้ผู้ที่สนใจสามารถศึกษาและนำไปสร้างตัวแบบการพยากรณ์ในการพยากรณ์อย่างอื่นได้ เพราะทฤษฎีการเรียนรู้ด้วยเครื่องแม้ว่าจะเป็นที่นิยมในประเทศไทยแต่ยังมีปัญหาในเรื่องของจำนวนงานวิจัยที่ใช้ศึกษามีจำนวนน้อยถึงแม้ว่าจะมีหน่วยงานเอกชนวิจัยเฉพาะทาง เช่น McKinsey ที่สามารถพยากรณ์ราคาเม็ดพลาสติก PP ได้แม่นยำ แต่ค่าใช้จ่ายในการดำเนินการที่สูงมากอาจทำให้ไม่คุ้มทุนในการลงทุน โดยใน

วิทยานิพนธ์ฉบับนี้มุ่งหมายที่จะทำการสร้างตัวแบบพยากรณ์ที่มีความคลาดเคลื่อน MAPE ต่ำกว่า 10%

1.3 วัตถุประสงค์ของงานวิจัย

1. สร้างและเปรียบเทียบตัวแบบการพยากรณ์ราคาเม็ดพลาสติก PP ของประเทศไทยที่เหมาะสมเพื่อพยากรณ์ราคาเม็ดพลาสติก PP ในประเทศไทยเป็นรายเดือน

2. สร้างตัวแบบการพยากรณ์แบบผสม (Hybrid Forecasting Model) จากตัวแบบการเรียนรู้ด้วยเครื่องแล้วนำมาเปรียบเทียบกับตัวแบบชนิดอื่น โดยใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error, MAPE) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

1.4 ขอบเขตของงานวิจัย

1. งานวิจัยฉบับนี้ได้รวบรวมข้อมูลราคาเม็ดพลาสติก PP มาจากสถาบันพลาสติก โดยเป็นข้อมูลราคาเฉลี่ยจากผู้ผลิตแต่ละบริษัทตั้งแต่ เดือนมกราคม พ.ศ. 2554 ถึง เดือนมิถุนายน พ.ศ. 2564 รวมเป็นระยะเวลา 12 ปี โดยจะแบ่งเป็น 3 ช่วงได้แก่

1.1 ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2554 ถึงเดือนธันวาคม พ.ศ. 2560 จำนวน 84 เดือน โดยใช้เป็นชุดข้อมูลที่ใช้ในการสร้างตัวแบบการพยากรณ์ (Training Data Set)

1.2 ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563 จำนวน 30 เดือน โดยใช้เป็นชุดข้อมูลที่ใช้ในการตรวจสอบความแม่นยำและเปรียบเทียบความแม่นยำของแต่ละตัวแบบการพยากรณ์เพื่อเลือกตัวแบบการพยากรณ์ที่แม่นยำที่สุด (Cross Validation Data Set)

1.3 ข้อมูลตั้งแต่เดือนกรกฎาคม พ.ศ. 2563 ถึงเดือนธันวาคม พ.ศ. 2565 จำนวน 30 ข้อมูล โดยใช้เป็นชุดข้อมูลที่ใช้ทดสอบตัวแบบการพยากรณ์ที่แม่นยำที่สุดที่ได้จากชุดข้อมูล Cross Validation Data Set (Testing Data Set)

2. ตัวแบบที่ใช้พยากรณ์ราคาเม็ดพลาสติก PP ได้แก่

2.1 ตัวแบบพยากรณ์ปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์

2.2 ตัวแบบพยากรณ์ Seasonal Autoregressive Integrated Moving Average (SARIMA)

2.3 ตัวแบบพยากรณ์การวิเคราะห์การถดถอยเชิงพหุคูณเชิงเส้น (Multiple Linear Regression Analysis)

2.4 ตัวแบบพยากรณ์ Support Vector Regression

- 2.5 ตัวแบบพยากรณ์ Decision Tree
- 2.6 ตัวแบบพยากรณ์ Random Forest
- 2.7 ตัวแบบพยากรณ์ Extreme Gradient Boosting (XGBoost)
- 2.8 ตัวแบบพยากรณ์ Artificial Neural Network (ANN)
- 2.9 ตัวแบบพยากรณ์แบบผสม (Hybrid Forecasting Model)

1.5 ผลลัพธ์ที่คาดว่าจะได้รับ

1. ทราบปัจจัยที่มีผลกระทบต่อราคาเม็ดพลาสติก PP
2. ได้ผลลัพธ์ตัวแบบการพยากรณ์ราคาเม็ดพลาสติก PP ที่เหมาะสม

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. ผลลัพธ์จากผลการพยากรณ์ราคาเม็ดพลาสติก PP ช่วยให้ผู้ประกอบการหรือเจ้าของธุรกิจที่สนใจ สามารถนำไปใช้วางแผนในการสั่งซื้อและจัดเก็บปริมาณเม็ดพลาสติก PP ได้
2. สามารถบ่งบอกถึงปัจจัยที่มีผลกระทบต่อราคาเม็ดพลาสติกให้ผู้ประกอบการและเจ้าของธุรกิจสามารถเข้าใจกลไกราคาเม็ดพลาสติก PP มากขึ้น
3. สร้างตัวแบบพยากรณ์ผสมโดยใช้ทฤษฎีการเรียนรู้ด้วยเครื่อง ในการพยากรณ์ข้อมูลซึ่งสามารถให้ผู้สนใจมาทำการศึกษาและนำไปประยุกต์ใช้กับการพยากรณ์อย่างอื่นได้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การพยากรณ์

การพยากรณ์ หมายถึง การทำนายเหตุการณ์ในอนาคต อาจนำข้อมูลในอดีตมาพยากรณ์อนาคตโดยใช้หลักการทางคณิตศาสตร์ ใช้ดุลยพินิจของผู้พยากรณ์หรือนำหลายวิธีมารวมกันเพื่อให้การพยากรณ์มีความแม่นยำมากที่สุด โดยการพยากรณ์มีประโยชน์อย่างมากต่อธุรกิจที่มีการแข่งขันสูงในปัจจุบันเพราะสามารถนำผลจากการพยากรณ์มาเป็นข้อมูลในการวางแผนกิจกรรมต่างๆ ให้เกิดประสิทธิภาพและประสิทธิผลสูงที่สุด เช่น การพยากรณ์ความต้องการเพื่อนำมาใช้ในการวางแผนการผลิต การพยากรณ์ราคาวัตถุดิบตั้งต้นที่ใช้ในการผลิตเพื่อวางแผนการสั่งซื้อ โดยตัวอย่างกรณีศึกษาของบริษัท Disney ที่นำการพยากรณ์มาใช้ที่ Disney World ในการประมาณผู้เข้าชมในรายวัน รายสัปดาห์ และ รายปี รวมทั้งอีก 5 ปีข้างหน้า เพื่อนำไปใช้ประโยชน์ต่อการตัดสินใจดำเนินกิจกรรมต่าง ๆ โดยที่ใช้วิธีการตัดสินใจทั้งในเชิงคุณภาพและเชิงปริมาณ ซึ่งทาง Disney ได้ทำการเก็บข้อมูลด้วยการใช้แบบสอบถามจากนักท่องเที่ยวประมาณ 1 ล้านคนต่อปี โดยพฤติกรรมของนักท่องเที่ยวสามารถนำมาสร้างตัวแบบพยากรณ์ที่มีความแม่นยำ (กฤษดา, 2557)

2.1.2 ช่วงระยะเวลาการพยากรณ์

ระยะเวลาของการพยากรณ์สามารถแบ่งได้ 3 ประเภท (ซัชชญา, 2560)

1. การพยากรณ์ในระยะสั้น เป็นการพยากรณ์เหตุการณ์ที่ไม่เกิน 1 ปีโดยทั่วไปมักจะอยู่ในช่วงไม่เกิน 3 เดือน เช่น การพยากรณ์การวางแผนจัดซื้อ การจัดทำตารางการทำงาน การมอบหมายงาน การพยากรณ์ยอดขาย และการพยากรณ์ระดับการผลิต

2. การพยากรณ์ระยะกลาง เป็นการพยากรณ์เหตุการณ์ที่อยู่ในช่วง 1 ปี ถึง 3 ปี จะใช้มากในการพยากรณ์การวางแผนการขาย การวางแผนการผลิต การวางแผนด้านงบประมาณเงินสด และการวิเคราะห์การวางแผนการดำเนินงานต่าง ๆ

3. การพยากรณ์ระยะยาว เป็นการพยากรณ์เหตุการณ์ที่มากกว่า 3 ปีขึ้นไป มักใช้สำหรับการวางแผนผลิตภัณฑ์ใหม่ ค่าใช้จ่ายในการลงทุน การขยายทำเลที่ตั้ง และการวิจัยพัฒนา

2.1.3 ขั้นตอนในการพยากรณ์

1. กำหนดปัญหาและรวบรวมข้อมูล (Define problem and Data Collection)

การกำหนดปัญหาและการรวบรวมข้อมูลต้องอยู่ในขั้นตอนเดียวกันเนื่องจาก ปัญหาที่ต้องการศึกษาจะเป็นตัวกำหนดข้อมูลที่ต้องรวบรวม ในกรณีของการพยากรณ์เชิงปริมาณต้องมีข้อมูล

ที่ถูกต้องและสอดคล้องเหมาะสมกับปัญหาที่กำหนดขึ้นมา หากไม่มีข้อมูลที่เกี่ยวข้อง ควรจะกำหนดหัวข้อปัญหาใหม่หรือเปลี่ยนไปใช้การพยากรณ์เชิงคุณภาพแทน การรวบรวมข้อมูลต้องคำนึงถึงระยะเวลาที่เก็บข้อมูลย้อนหลัง ความถูกต้องของข้อมูล ความน่าเชื่อถือของข้อมูลที่ได้รับมา ความสม่ำเสมอในการเก็บข้อมูล (ช่วงระยะเวลาการเก็บข้อมูลต้องเท่ากันทุกตัว และ ใช้วิธีการเก็บข้อมูลแบบเดียวกัน)

2. ตรวจสอบข้อมูลและแก้ไขข้อมูล (Data Manipulation and Cleaning)

กระบวนการตรวจสอบและการแก้ไขหรือลบรายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูลหรือ ฐานข้อมูล ซึ่งเป็นหลักสำคัญของฐานข้อมูล เพราะหมายถึงความไม่สมบูรณ์ ความไม่ถูกต้อง ความไม่สัมพันธ์กับข้อมูลอื่น เป็นต้น จึงต้องมีการแทนที่ การปรับปรุง หรือการลบข้อมูลที่ไม่ถูกต้อง เพื่อให้ข้อมูลมีคุณภาพ

3. สร้างตัวแบบการพยากรณ์และประเมินประสิทธิภาพ

ขั้นตอนนี้จะนำข้อมูลที่รวบรวมมาสร้างตัวแบบการพยากรณ์ที่เหมาะสมโดยพิจารณาจากรูปแบบข้อมูล องค์ประกอบของข้อมูลเพื่อใช้เลือกวิธีการพยากรณ์ที่เหมาะสม โดยตัวแบบที่ไม่ซับซ้อนจะเข้าใจง่ายแล้วได้รับการอนุมัติจากผู้ตัดสินใจ (เช่น ผู้บริหาร) ง่ายกว่าตัวแบบพยากรณ์ที่ซับซ้อนซึ่งเข้าใจยากสำหรับผู้ตัดสินใจ นอกจากนี้ยังต้องคำนึงถึงประสิทธิภาพของตัวแบบด้วย กล่าวคือ ตัวแบบที่จะนำไปใช้ต้องมีความคลาดเคลื่อนที่ต่ำที่สุดเมื่อเปรียบเทียบกับตัวแบบอื่น

นอกจากนั้นในขั้นตอนนี้จะทำการแบ่งข้อมูลออกเป็น 3 ส่วนได้แก่ Training Data Set ใช้สำหรับการสร้างตัวแบบการพยากรณ์ Cross Validation Data Set ใช้สำหรับการตรวจสอบความแม่นยำของตัวแบบการพยากรณ์และ Testing Data Set ใช้สำหรับการวัดประสิทธิภาพของตัวแบบการพยากรณ์ที่ดีที่สุดที่ได้จากการคัดเลือกโดยใช้ Cross Validation Data Set

4. การนำตัวแบบการพยากรณ์ไปใช้พยากรณ์จริง (Model Implementation)

เมื่อข้อมูลที่รวบรวมมาผ่านการตรวจสอบและแก้ไข และได้ทำการเลือกตัวแบบพยากรณ์ที่ดีที่สุดมาแล้ว ในขั้นตอนนี้จะนำตัวแบบพยากรณ์ที่ดีที่สุดที่ถูกเลือกมาทำการพยากรณ์

5. ประเมินผลตัวแบบการพยากรณ์ (Forecast Evaluation)

นำผลลัพธ์การพยากรณ์ที่ได้จากตัวแบบพยากรณ์ที่ดีที่สุดมาทำการเปรียบเทียบกับข้อมูลจริง ณ ช่วงเวลาเดียวกัน (Testing Data Set) เพื่อหาความคลาดเคลื่อนของตัวแบบการพยากรณ์ (Hanke and Wichern, 2009)

2.1.4 ประเภทของการพยากรณ์เชิงสถิติ

2.1.4.1 การพยากรณ์เชิงคุณภาพ (Qualitative Forecasting Techniques)

การพยากรณ์เชิงคุณภาพเป็นวิธีการพยากรณ์ที่นำความรู้ในอดีตมาตัดสิน หรือใช้ประสบการณ์ ความคิดเห็น มาตัดสินในเรื่องนั้น ๆ โดยการพยากรณ์เชิงคุณภาพนี้ไม่ได้ใช้วิธีการทางวิทยาศาสตร์และสถิติแต่อย่างใด (นิฉา, 2560)

2.1.4.2 การพยากรณ์เชิงปริมาณ (Quantitative Forecasting Techniques)

การพยากรณ์เชิงปริมาณเป็นการนำข้อมูลตั้งแต่อดีตจนถึงปัจจุบันมาวิเคราะห์แล้วสร้างตัวแบบพยากรณ์ โดยสามารถแบ่งออกได้เป็น 2 วิธีได้แก่ วิธีการพยากรณ์ข้อมูลเชิงสาเหตุ (Casual Method) และ วิธีการพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data Method) (ศิริเทพ, 2560)

2.1.5 วิธีการพยากรณ์ข้อมูลเชิงสาเหตุ (Casual Method)

การพยากรณ์วิธีนี้เป็น การนำความสัมพันธ์ระหว่างปัจจัยหรือตัวแปรมาวิเคราะห์หาความสัมพันธ์กับตัวแปรที่สนใจหรือผลลัพธ์ที่ต้องการ โดยจะนำความสัมพันธ์ที่ได้มาสร้างตัวแบบการพยากรณ์เพื่อพยากรณ์ผลลัพธ์ที่ต้องการ และปัจจัยหรือตัวแปรที่นำมาศึกษาต้องมีอย่างน้อย 2 ปัจจัย หรือ 2 ตัวแปร การพยากรณ์ในรูปแบบนี้สามารถแบ่งได้เป็น 2 แบบดังนี้

2.1.5.1 วิธีการวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis)

วิธีการวิเคราะห์การถดถอยอย่างง่ายจะเป็นการนำความสัมพันธ์ระหว่างตัวแปรตาม (Dependent Variable) หรือ ผลลัพธ์ กับตัวแปรอิสระ (Independent Variable) หรือ ปัจจัยเพียง 1 ตัวมาสร้างสมการพยากรณ์ผลลัพธ์ โดยมีรูปแบบดังสมการ 2.1 (ศิริเทพ, 2560)

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad (2.1)$$

โดยที่	Y	หมายถึง ผลลัพธ์
	β_0, β_1	หมายถึง สัมประสิทธิ์ถดถอย
	X_1	หมายถึง ตัวแปรอิสระ หรือ ปัจจัย
	ε	หมายถึง ความคลาดเคลื่อน

2.1.5.2 วิธีการวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Regression Analysis)

วิธีการวิเคราะห์การถดถอยเชิงพหุคูณ เป็นวิธีการพยากรณ์โดยใช้ตัวแปรอิสระหรือปัจจัยที่เกี่ยวข้องตั้งแต่ 2 ตัวขึ้นไป มาสร้างสมการเส้นตรงเพื่อพยากรณ์ผลลัพธ์โดยมีรูปแบบดังสมการ 2.2 (ศิริเทพ, 2560)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \varepsilon \quad (2.2)$$

โดยที่	Y	หมายถึง ผลลัพธ์
	$\beta_0, \beta_1, \beta_2, \dots, \beta_K$	หมายถึง สัมประสิทธิ์ถดถอย
	$X_1, X_2, X_3, \dots, X_K$	หมายถึง ตัวแปรอิสระ หรือ ปัจจัย
	K	หมายถึง จำนวนของปัจจัยหรือตัวแปรอิสระ
	ε	หมายถึง ความคลาดเคลื่อน

โดยในการคัดเลือกปัจจัยต้องมีการทำการทดสอบทางสมมติฐานทดสอบ p-value เพื่อพิสูจน์ว่าปัจจัยดังกล่าวมีผลต่อตัวแปร Y และต้องทำการทดสอบโดยการหาค่า Variance Inflation Factor (VIF) เพื่อพิสูจน์ว่าตัวแปรอิสระแต่ละตัวไม่ได้ส่งผลต่อกัน

2.1.6 วิธีการพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data Method)

การพยากรณ์ข้อมูลอนุกรมเวลาจะเป็นการนำข้อมูลในอดีตมาพยากรณ์ข้อมูลในอนาคต โดยตั้งอยู่บนสมมติฐานว่ารูปแบบของข้อมูลตั้งแต่อดีตจนถึงอนาคตเป็นรูปแบบเดียวกัน

2.1.6.1 ข้อมูลอนุกรมเวลา (Time Series Data)

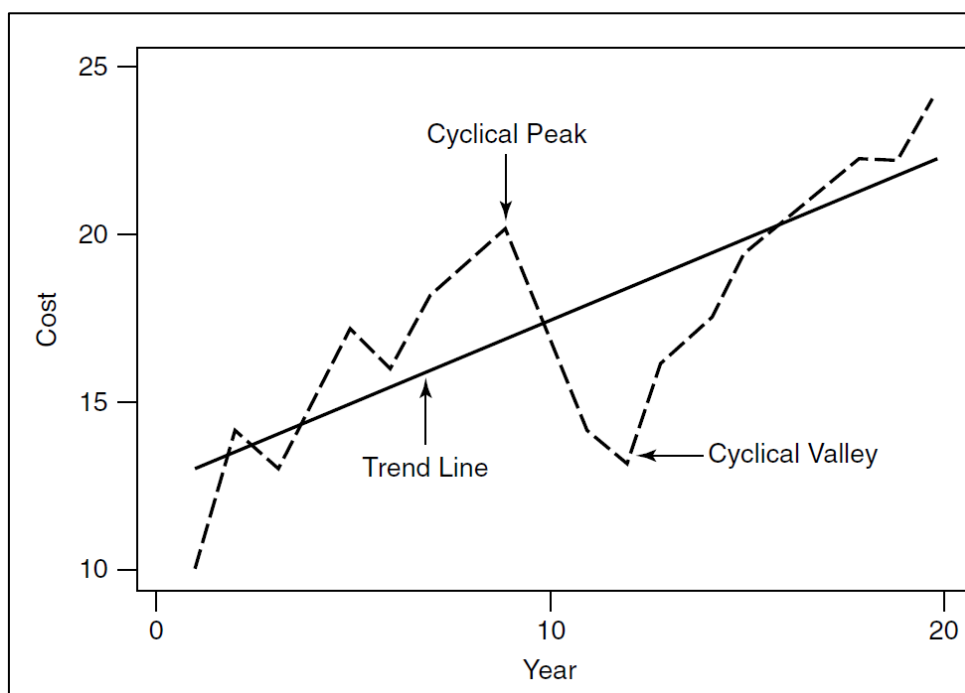
ข้อมูลอนุกรมเวลา คือ ข้อมูลที่ได้จากการเก็บรวบรวมมาอย่างต่อเนื่องตั้งแต่อดีตจนถึงปัจจุบันโดยมีช่วงระยะเวลาการเก็บแต่ละข้อมูลห่างเท่ากันอย่างสม่ำเสมอ โดยข้อมูลอนุกรมเวลาประกอบไปด้วย 4 องค์ประกอบหลัก ได้แก่ (Hanke and Wichern, 2009)

1. แนวโน้ม (Trend) เป็นข้อมูลที่มีลักษณะการเปลี่ยนแปลงค่าตามระยะเวลาไปในทิศทางเพิ่มขึ้นหรือลดลง โดยอาจจะเป็นได้ทั้งแนวโน้มแบบเส้นตรง (Linear) หรือเส้นโค้ง (Quadratic)

2. ฤดูกาล (Seasonal) เป็นข้อมูลที่มีลักษณะการเปลี่ยนแปลงค่าเหมือนกันในแต่ละช่วงเวลาซึ่งสามารถเป็นได้ตั้งแต่ รายวัน รายสัปดาห์ รายเดือน รายไตรมาส โดยระยะเวลาในแต่ละช่วงเวลาต้องไม่เกิน 1 ปี

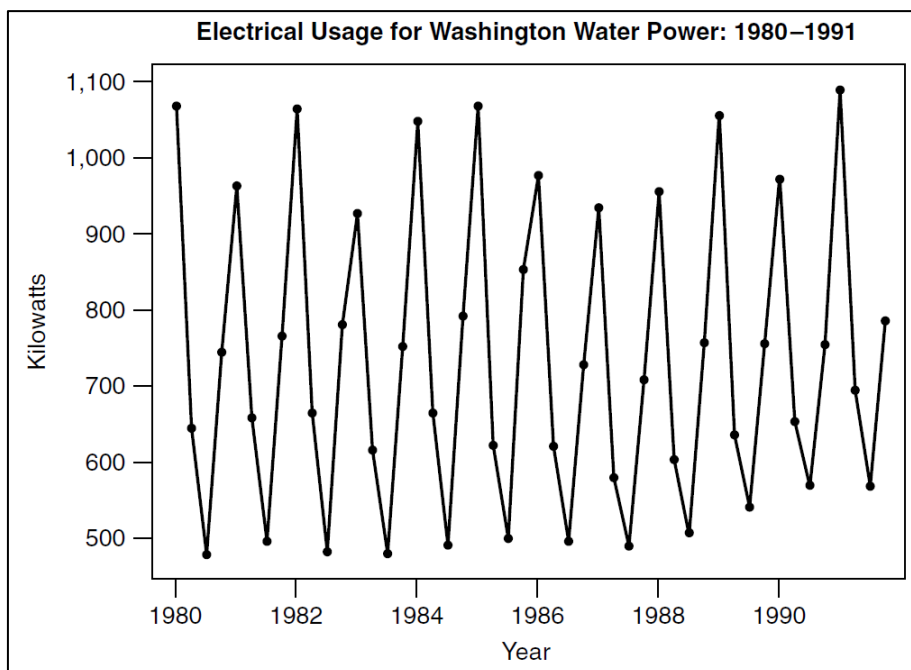
3. วัฏจักร (Cyclical) เป็นข้อมูลที่มีลักษณะการเปลี่ยนแปลงค่าแบบเพิ่มขึ้นและลดลงเป็นช่วงคล้ายรูปแบบฤดูกาล แต่ในรูปแบบวัฏจักรระยะเวลาในแต่ละช่วงจะมากกว่า 1 ปี และขนาดของการเพิ่มขึ้นและลดลงในแต่ละช่วงอาจไม่เท่ากัน

4. ผิดปกติ (Irregular) เป็นข้อมูลที่มีลักษณะการเปลี่ยนแปลงค่าที่ไม่มีแบบแผน ไม่มีสาเหตุในการเกิด โดยรูปแบบชนิดนี้เป็นส่วนที่เหลือจาก แนวโน้ม ฤดูกาล และ วัฏจักรของข้อมูล โดยในรูปแบบที่ 1 ถึง 3 แสดงดังตัวอย่างในรูปแบบที่ 4 และ 5



รูปที่ 4 ข้อมูลที่มีลักษณะของแนวโน้มและวัฏจักร

(Hanke and Wichern, 2009)

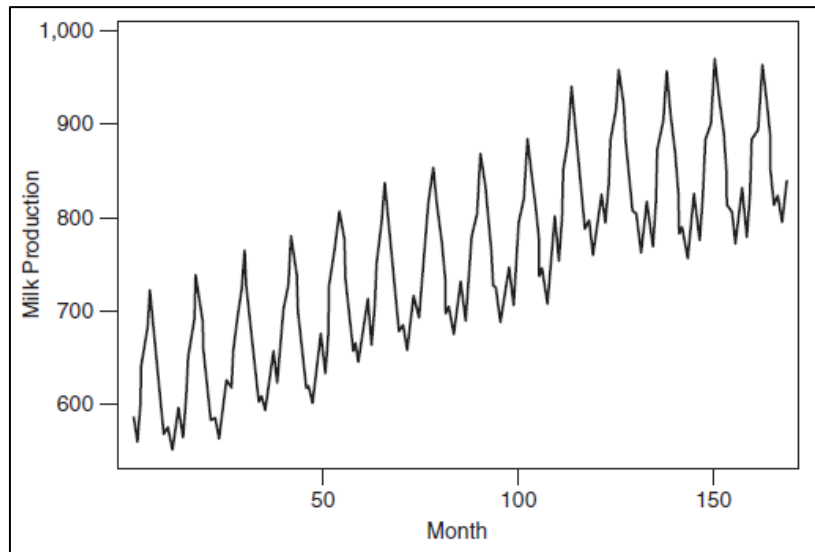


รูปที่ 5 ข้อมูลที่มีลักษณะเป็นฤดูกาล
(Hanke and Wichern, 2009)

2.1.6.2 รูปแบบของอนุกรมเวลา

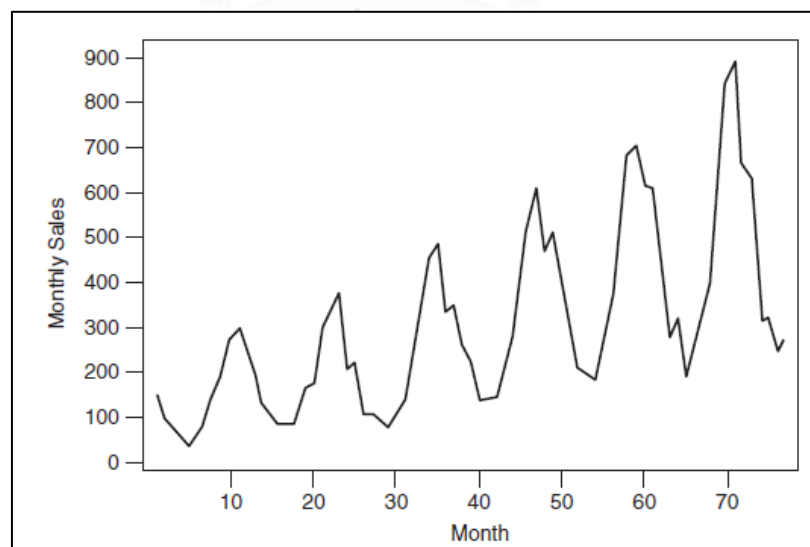
รูปแบบของอนุกรมเวลาสามารถแบ่งเป็น 2 รูปแบบดังนี้

1.) รูปแบบการบวก (Additive Model) เป็นรูปแบบที่องค์ประกอบทั้ง 4 อย่างของข้อมูลเป็นอิสระต่อกันคือเมื่อองค์ประกอบตัวใดมีการเปลี่ยนแปลงเกิดขึ้นไม่ว่าจะเป็นในทิศทางเพิ่มขึ้นหรือลดลง จะไม่ส่งผลกับองค์ประกอบส่วนที่เหลือของข้อมูล โดยเมื่อนำข้อมูลที่มีรูปแบบการบวกมาทำการพล็อตลงบนกราฟโดยแกนในแนวนอนคือระยะเวลา และแกนในแนวตั้งเป็นข้อมูลจะได้กราฟดังรูป 6 ที่มีลักษณะความแปรปรวนคงที่ (ศิริเทพ, 2560)



รูปที่ 6 ข้อมูลอนุกรมเวลารูปแบบการบวก
(Hanke and Wichern, 2009)

2.) รูปแบบการคูณ (Multiplicative Model) เป็นรูปแบบที่องค์ประกอบทั้ง 4 อย่างของข้อมูลไม่เป็นอิสระต่อกันคือเมื่อองค์ประกอบตัวใดมีการเปลี่ยนแปลงเกิดขึ้นไม่ว่าจะเป็นในทิศทางเพิ่มขึ้นหรือลดลง จะส่งผลกับองค์ประกอบส่วนที่เหลือของข้อมูล โดยเมื่อนำข้อมูลที่มีรูปแบบการคูณมาทำการพล็อตลงบนกราฟโดยแกนในแนวนอนคือระยะเวลา และแกนในแนวตั้งเป็นข้อมูล จะได้กราฟดังรูป 7 ที่มีลักษณะความแปรปรวนไม่คงที่ (ศิริเทพ, 2560)



รูปที่ 7 ข้อมูลอนุกรมเวลารูปแบบการคูณ
(Hanke and Wichern, 2009)

2.1.6.3 วิธีปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์ (Holt - Winters Exponential Smoothing Method)

วิธีปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์ เป็นวิธีการพยากรณ์ที่ใช้กับข้อมูลที่มีแนวโน้ม (Trend) และเป็นข้อมูลเชิงฤดูกาล (Seasonal) หลักการคือจะเป็นการถ่วงน้ำหนักข้อมูลด้วย Trend Level และ Seasonal แล้วนำมาพยากรณ์ ซึ่งเป็นวิธีการที่พัฒนามาจาก Holt's Exponential Smoothing ที่เป็นวิธีการพยากรณ์ข้อมูลที่มีลักษณะเป็นแนวโน้มเพียงอย่างเดียว

เนื่องจากความแปรปรวนจากส่วนประกอบของข้อมูลที่เป็นฤดูกาล (Seasonal) มีอยู่ 2 แบบ ทำให้การพยากรณ์ด้วยวิธีนี้มีอยู่ด้วยกัน 2 กรณีได้แก่ Additive Method ใช้กับข้อมูลที่มีความแปรปรวนคงที่ และ Multiplicative Method ใช้กับข้อมูลที่มีความแปรปรวนเพิ่มขึ้น หรือลดลง ดังนั้นวิธีปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์ สามารถแบ่งออกเป็น 2 กรณีดังนี้ (O'Connell and Koehler, 2005)

1.) Additive Holt – Winters Exponential Smoothing Method

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

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

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

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

2.) Multiplicative Holt – Winters Exponential Smoothing Method

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

$$L_t = \alpha(Y_t/S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (2.9)$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \quad (2.10)$$

$$S_t = \gamma(Y_t/L_t) + (1 - \gamma)S_{t-s} \quad (2.11)$$

โดยที่	Y_t	หมายถึง ค่าข้อมูลจริง ณ เวลา t
	\hat{Y}_{t+p}	หมายถึง ค่าพยากรณ์ ณ เวลา t+p
	L_t	หมายถึง ค่าประมาณระดับของข้อมูล
	T_t	หมายถึง ค่าประมาณแนวโน้มของข้อมูล
	S_t	หมายถึง ค่าประมาณฤดูกาลของข้อมูล
	α	หมายถึง ค่าคงที่ปรับระดับของข้อมูล (มีค่าอยู่ระหว่าง 0 ถึง 1)

β	หมายถึง ค่าคงที่ปรับแนวโน้มข้อมูล (มีค่าอยู่ระหว่าง 0 ถึง 1)
γ	หมายถึง ค่าคงที่ปรับฤดูกาลของข้อมูล (มีค่าอยู่ระหว่าง 0 ถึง 1)
p	หมายถึง ช่วงระยะเวลาที่จะทำการพยากรณ์
s	หมายถึง ระยะเวลาของฤดูกาล

2.1.6.4 วิธีบ็อกซ์-เจนกินส์ (Box-Jenkins Method : ARIMA Model)

วิธีบ็อกซ์-เจนกินส์ เป็นการนำความสัมพันธ์ข้อมูลในอดีตมาใช้สร้างตัวแบบการพยากรณ์โดยจะพิจารณาความสัมพันธ์ระหว่างตัวแปรที่สนใจ Y ณ เวลา t กับตัวแปรที่สนใจ Y ในอดีต ณ เวลา $t-1$ แล้วนำมาพยากรณ์ตัวแปรที่สนใจ Y ในอนาคต ณ เวลา $t+1$ ซึ่ง ARIMA เกิดจากการรวมกันระหว่าง Autoregressive Model (AR) และ Moving Average Model (MA) เป็นตัวแบบ Autoregressive Integrated Moving Average (ARIMA)

หากตัวข้อมูลมีรูปแบบฤดูกาลนั้นเมื่อนำมาใช้กับตัวแบบบ็อกซ์-เจนกินส์ จะได้เป็นตัวแบบ Seasonal Autoregressive Integrated Moving Average Model (SARIMA) ซึ่งอยู่ในรูปแบบดังสมการที่ 2.12

$$SARIMA(p, d, q)(P, D, Q)_m \quad (2.12)$$

โดยที่ p หมายถึง ระดับในส่วนของ Autoregressive
 q หมายถึง ระดับในส่วนของ Moving Average
 d หมายถึง จำนวนครั้งในการทำ Differencing
 P หมายถึง ระดับในส่วนของ Seasonal Autoregressive
 Q หมายถึง ระดับในส่วนของ Seasonal Moving Average
 D หมายถึง จำนวนครั้งในการทำ Differencing ของฤดูกาล
 m หมายถึง จำนวนของฤดูกาลในแต่ละปี เช่น รายเดือน $m = 12$

สำหรับขั้นตอนการพยากรณ์ด้วยวิธีบ็อกซ์-เจนกินส์มีด้วยกัน 4 ขั้นตอน ดังนี้

1. กำหนดตัวแบบ ARIMA ที่เหมาะสม (Identify Model)

ในขั้นตอนนี้ข้อมูลที่จะนำเข้าตัวแบบการพยากรณ์ต้องต้องเป็นข้อมูลที่มีลักษณะคงที่ (Stationary) หากข้อมูลมีลักษณะที่ค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) ไม่คงที่ หรือมีลักษณะอื่น ๆ ที่ไม่ใช่ลักษณะคงที่จะต้องทำการหาผลต่างมากกว่าหรือเท่ากับ 1 ครั้งซึ่งคือ

ค่าพารามิเตอร์ของ d หรือ D เมื่อได้ข้อมูลที่มีลักษณะคงที่แล้วจึงสามารถนำมาสร้างตัวแบบ ARIMA โดยพิจารณาจากกราฟอัตสหสัมพันธ์ (Autocorrelation Function: ACF) ซึ่งจะบ่งบอกถึงค่าพารามิเตอร์ของ q หรือ Q และ กราฟอัตสหสัมพันธ์บางส่วน (Partial Autocorrelation Function: PACF) ที่จะบ่งบอกค่าพารามิเตอร์ p หรือ P ของตัวแบบการพยากรณ์

2. ประมาณค่าพารามิเตอร์ (Estimate Parameter)

ในขั้นตอนนี้จะทำการประมาณค่าพารามิเตอร์ไปจนกว่าจะได้ตัวแบบการพยากรณ์ที่มีค่าความคลาดเคลื่อนต่ำที่สุด

3. ตรวจสอบความเหมาะสมของตัวแบบ (Model Adequate Inspection)

ขั้นตอนนี้เป็นขั้นตอนที่มีความสำคัญมากที่สุด โดยจะเป็นการตรวจสอบว่าค่าความคลาดเคลื่อนมีความสัมพันธ์กันเองหรือไม่ โดยตัวแบบการพยากรณ์ที่คืนค่าความคลาดเคลื่อนต้องไม่มีความสัมพันธ์เหลืออยู่หรืออยู่ในรูปแบบสุ่ม (Random)

4. พยากรณ์ (Forecasting)

เมื่อผ่านขั้นตอนที่ 3 จะได้ผลลัพธ์เป็นตัวแบบการพยากรณ์ที่เหมาะสม ในขั้นตอนนี้จะเป็นการนำตัวแบบการพยากรณ์นั้นมาทำการพยากรณ์ค่าในอนาคตได้ ซึ่งสามารถพยากรณ์ได้ทั้งแบบจุด (Point Forecast) และแบบช่วง (Interval Forecast) (Hanke and Wichern, 2009)

2.1.7 การวัดความแม่นยำของตัวแบบพยากรณ์

ในการคัดเลือกตัวแบบการพยากรณ์ในงานวิจัยฉบับนี้ใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์โดยเขียนได้ดังสมการที่ 2.13 (Hanke and Wichern, 2009)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \times 100\% \quad (2.13)$$

โดยที่ Y_t หมายถึง ข้อมูลจริง ณ เวลา t

\hat{Y}_t หมายถึง ข้อมูลจากการพยากรณ์ ณ เวลา t

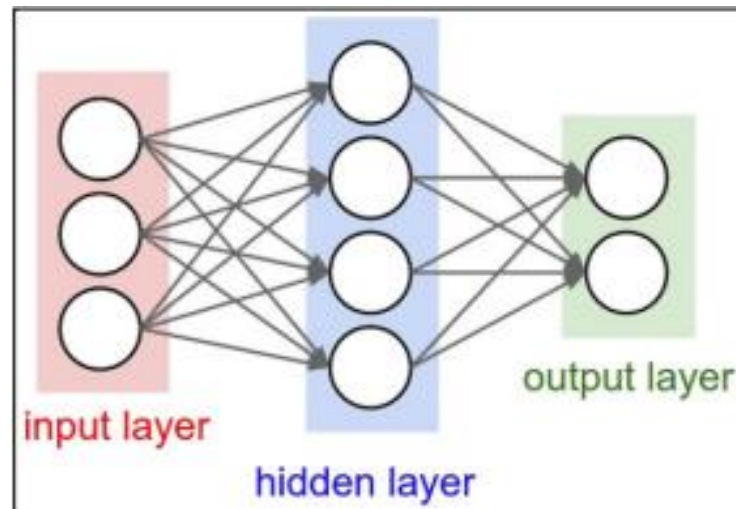
n หมายถึง จำนวนข้อมูล

2.1.8 การเรียนรู้ของเครื่อง (Machine Learning)

ปัญญาประดิษฐ์ (Artificial Intelligence) เกิดจากการสร้างโดยมนุษย์เพื่อที่จะสามารถช่วยแบ่งเบาภาระการทำงานของมนุษย์และจะต้องได้รับคำสั่งจากมนุษย์หรือการกระตุ้นจากสิ่งแวดล้อม แล้วทำงานตามโปรแกรมที่มนุษย์ได้ออกแบบไว้ โดยจะไม่สามารถทำงานนอกเหนือจากคำสั่งมนุษย์ได้ นอกจากนี้จะต้องมีการทำ Turing Test เพื่อเป็นการทดสอบว่าปัญญาประดิษฐ์ที่ถูกสร้างขึ้นมานั้นมีสติปัญญาจริงหรือไม่

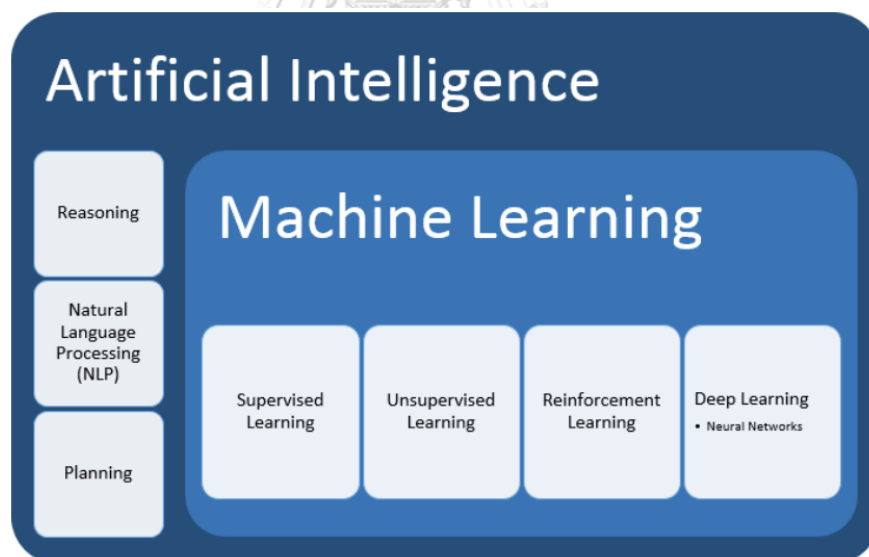
การเรียนรู้ด้วยเครื่องเป็นหนึ่งในรูปแบบของปัญญาประดิษฐ์ ที่มีความสามารถในการเรียนรู้ด้วยตัวเองเพื่อแก้ปัญหาที่ซับซ้อนขึ้นหรือช่วยในการตัดสินใจได้ดีขึ้น จากการนำชุดข้อมูล (Datasets) เข้าไปแล้วนำไปวิเคราะห์และเรียนรู้ผ่านอัลกอริทึม (Algorithms) โดยในปัจจุบันมีอัลกอริทึมอยู่หลายแบบซึ่งแต่ละแบบจะมีความเหมาะสมกับชุดข้อมูลที่แตกต่างกัน หลังจากนั้นจะได้ผลลัพธ์เป็นรูปแบบการเรียนรู้ด้วยเครื่องที่สามารถนำไปใช้ทำงานในด้านต่าง ๆ ได้แก่ การตรวจจับรูปภาพ (Image Recognition) การตรวจจับคำพูด (Speech Recognition) การแนะนำสินค้า (Product Recommendation) และการพยากรณ์ (Forecasting)

การเรียนรู้เชิงลึก (Deep Learning) และโครงข่ายประสาทเทียม (Neural Network) จัดว่าเป็นการเรียนรู้ด้วยเครื่องประเภทหนึ่งที่มีลักษณะการทำงานที่ใกล้เคียงกับการทำงานของสมองในมนุษย์คือมีชั้น (Layer) และในแต่ละชั้น จะมี Node เชื่อมต่อกันเป็นโครงข่าย ซึ่งชั้นของโครงข่ายประสาทเทียมจะแบ่งเป็น 3 ประเภทได้แก่ ชั้นข้อมูลเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นข้อมูลออก (Output Layer) ดังที่แสดงในรูปที่ 8 โดยถ้าเป็นโครงข่ายประสาทเทียมจะมีจำนวนชั้นซ่อนที่น้อยกว่าการเรียนรู้เชิงลึก (Deep Learning) หรือสามารถกล่าวได้ว่าการเรียนรู้เชิงลึกคือโครงข่ายประสาทเทียมที่มีความซับซ้อนของตัวโครงสร้างมากขึ้นเพื่อใช้ในการแก้ปัญหาที่ซับซ้อนมาก (Hurwitz, 2018)



รูปที่ 8 ลักษณะของโครงข่ายประสาทเทียม (Hurwitz, 2018)

ดังนั้นสามารถสรุปภาพรวมและความแตกต่างระหว่างปัญญาประดิษฐ์ การเรียนรู้ด้วยเครื่อง และการเรียนรู้เชิงลึกได้ดังรูปที่ 9

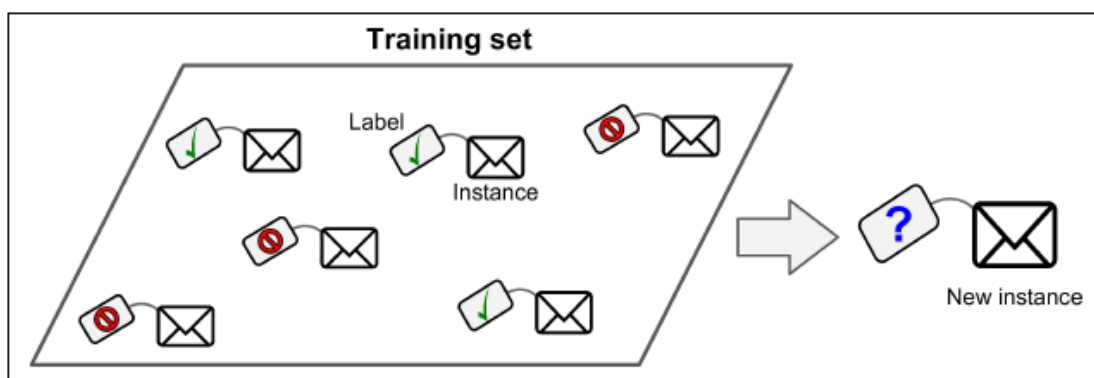


รูปที่ 9 ภาพรวมของปัญญาประดิษฐ์ (Hurwitz, 2018)

การเรียนรู้ด้วยเครื่องสามารถแบ่งได้เป็น 4 ประเภทได้แก่ 1. Supervised Learning 2. Unsupervised Learning 3. Semisupervised Learning และ 4. Reinforcement Learning

2.1.9 Supervised Learning

ชุดข้อมูล Training ของ Supervised Learning จะมีระบุลักษณะหรือรูปแบบของผลลัพธ์ที่ต้องการ (Label) ดังแสดงในรูปที่ 10



รูปที่ 10 รูปแบบชุดข้อมูลของ Training ของ Supervised Learning ที่มี Label (Géron, 2019)

การนำ Supervised Learning ไปใช้ในการพยากรณ์จะมีอยู่ 2 แบบได้แก่ 1. การจัดหมวดหมู่ (Classification) ผลลัพธ์ที่ได้จะเป็นค่าที่ไม่ต่อเนื่อง เช่น ใช่หรือไม่ใช่ ตัวอย่างการนำไปใช้งานเช่น การคัดกรองสแปมในอีเมล (Spam Filter) โดยมีหลักการทำงานคือนำตัวอย่างอีเมลที่เป็นสแปมมาทำการเรียนรู้เพื่อทำการสร้างอัลกอริทึมที่จะใช้คัดกรองอีเมลฉบับอื่นว่าเป็นสแปมหรือไม่ 2. การวิเคราะห์การถดถอย (Regression) ใช้ในการทำนายค่าต่อเนื่องเป็นตัวเลข เช่น ทำนายราคารถยนต์ โดยจะนำข้อมูลปัจจัยที่เกี่ยวข้องและผลลัพธ์ที่มีการระบุ Label ไว้มาทำการสร้างสมการพยากรณ์เพื่อใช้ทำนายค่าผลลัพธ์ (Géron, 2019)

ในงานวิจัยฉบับนี้จะเป็นการนำ Supervised Learning ในส่วนของ Regression มาใช้งาน ซึ่งมีตัวอย่างอัลกอริทึมดังต่อไปนี้

2.1.9.1 Linear Regression

หลักการทำงานของตัวแบบนี้ในการพยากรณ์คือการถ่วงน้ำหนักของแต่ละปัจจัยและรวมเข้ากับ Bias Term (Intercept Term) ดังสมการที่ 2.14

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n \quad (2.14)$$

โดยที่ \hat{y} หมายถึง ผลลัพธ์การพยากรณ์
 x_i หมายถึง ค่าของปัจจัยที่ i
 n หมายถึง จำนวนข้อมูลปัจจัย
 θ_j หมายถึง model parameter

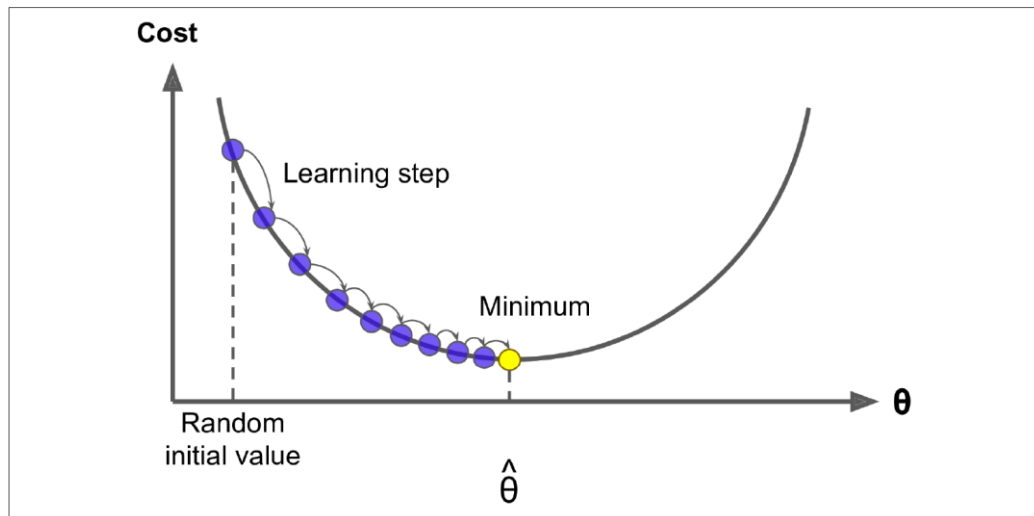
จากสมการที่ 2.14 สามารถเขียนให้อยู่ในรูปของเวกเตอร์ได้ดังสมการที่ 2.15

$$\hat{y} = h_{\theta}(x) = \theta \cdot x \quad (2.15)$$

โดยที่ \hat{y} หมายถึง ผลลัพธ์การพยากรณ์
 x หมายถึง ค่าของปัจจัยที่ i
 θ หมายถึง เวกเตอร์พารามิเตอร์ของตัวแบบ
 h_{θ} หมายถึง ฟังก์ชันสมมติฐาน
 $\theta \cdot x$ หมายถึง ผลคูณเวกเตอร์เชิงสเกลาร์ที่มีผลลัพธ์อยู่ในรูป $\theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n$

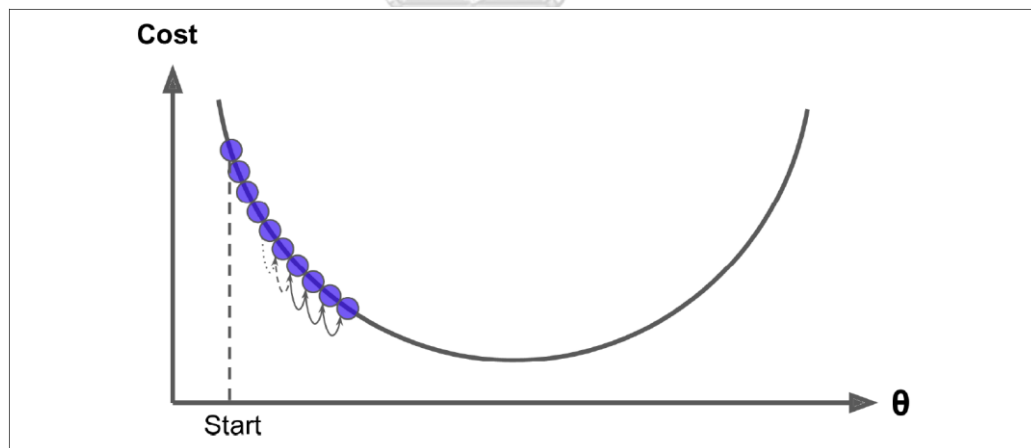
สำหรับตัวแบบ Linear Regression มี 2 วิธีที่จะทำให้ได้ค่าพารามิเตอร์ของตัวแบบที่สามารถลดค่าความคลาดเคลื่อนให้ได้ค่าที่ต่ำที่สุดโดยวิธีแรกคือ Normal Equation ซึ่งเป็นการคำนวณหาพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดจากชุดข้อมูล Training Data Set และวิธีต่อมาคือ การใช้อัลกอริทึมการเรียนรู้ เช่น Gradient Descent ซึ่งเป็นกระบวนการทำซ้ำในการปรับค่าพารามิเตอร์ของตัวแบบให้ได้ผลลัพธ์ Cost Function ที่มีค่าต่ำที่สุด (Géron, 2019)

Gradient Descent เป็นอัลกอริทึมการเรียนรู้ที่ใช้ในการหาค่าตอบที่ดีที่สุด (Optimal Solution) โดยหลักการจะเป็นดังรูปที่ 11 คือเริ่มจากการสุ่มค่า θ เริ่มต้นแล้วทำการขยับค่าซึ่งกำหนดจากอัตราการเรียนรู้ (Learning Rate) ให้ได้ค่า θ ใหม่จากนั้นวัดค่า Cost Function ที่ได้ หากมีค่าลดลงให้ทำต่อไปจนกระทั่งถึงจุดที่ได้ค่า Cost Function ที่ต่ำที่สุดให้หยุดและค่า θ ที่ได้จะเป็นค่า θ ที่ทำให้ได้ Cost Function ที่ต่ำที่สุด (Géron, 2019)

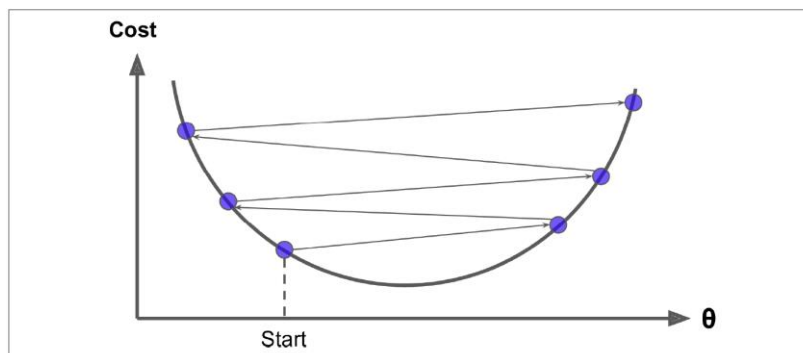


รูปที่ 11 หลักการทำงานของ Gradient descent (Géron, 2019)

หากค่า Learning Rate มีค่าต่ำจะส่งผลให้ระยะเวลาในการประมวลผลเพื่อให้ได้ค่า Cost Function ที่ต่ำที่สุดช้าไปด้วยดังรูปที่ 12 แต่ถ้ค่า Learning Rate มีค่ามากจะทำให้เกิดการลู่ออกจากค่าต่ำสุดดังรูปที่ 13

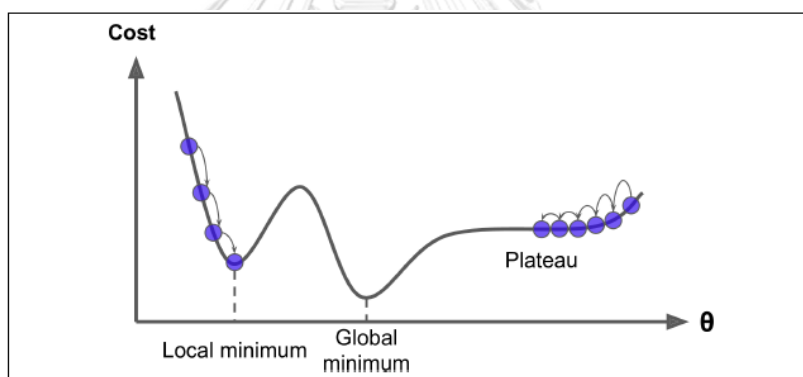


รูปที่ 12 การทำงานของ Gradient Descent ที่ Learning Rate ต่ำเกินไป (Géron, 2019)



รูปที่ 13 การทำงานของ Gradient Descent ที่ Learning Rate สูงเกินไป (Géron, 2019)

ในการทำ Gradient Descent บางครั้ง Cost Function ไม่ได้อยู่ในลักษณะที่เว้าแค่จุดเดียว (Convex Function) แต่อาจมีจุดที่เว้าได้หลายที่ทำให้เกิดคำตอบ 2 แบบนั่นคือ Local Minimum และ Global Minimum ดังรูปที่ 14

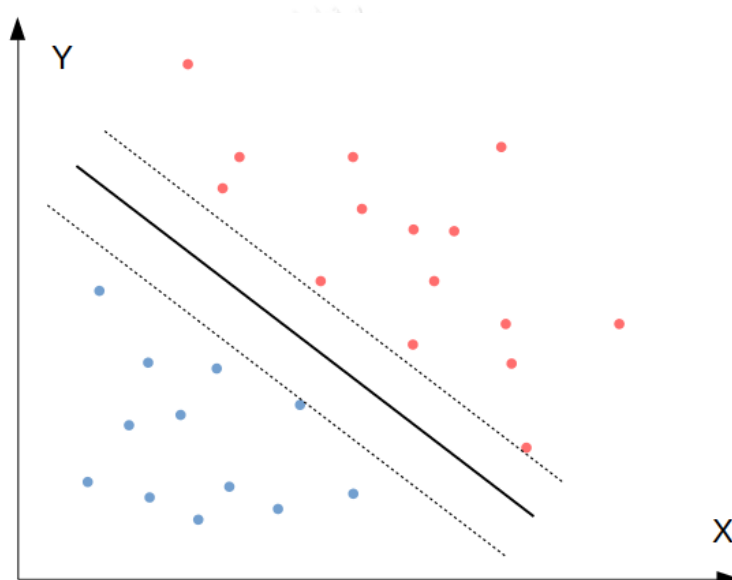


รูปที่ 14 Cost Function ที่มี Local Minimum และ Global Minimum (Géron, 2019)

ในการตรวจสอบว่าตัวแบบ Underfitting (High Bias) หรือ Overfitting (High Variance) นั้นสามารถทำได้จากการพิจารณาค่าความคลาดเคลื่อนจากชุดข้อมูล Training Data Set และชุดข้อมูล Cross Validation Set โดยถ้าหากค่าความคลาดเคลื่อนของชุดข้อมูล Training Data Set มีค่าต่ำแต่ค่าความคลาดเคลื่อนในชุดข้อมูล Cross Validation Set มีค่าสูงจะถือว่าเป็น Overfitting แต่ในทางตรงกันข้าม หากว่า Training Data Set มีค่าความคลาดเคลื่อนสูงแต่ค่าความคลาดเคลื่อนในชุดข้อมูล Cross Validation Set มีค่าใกล้เคียงกับ Training Data Set จะถือว่าเป็น Underfitting (Géron, 2019)

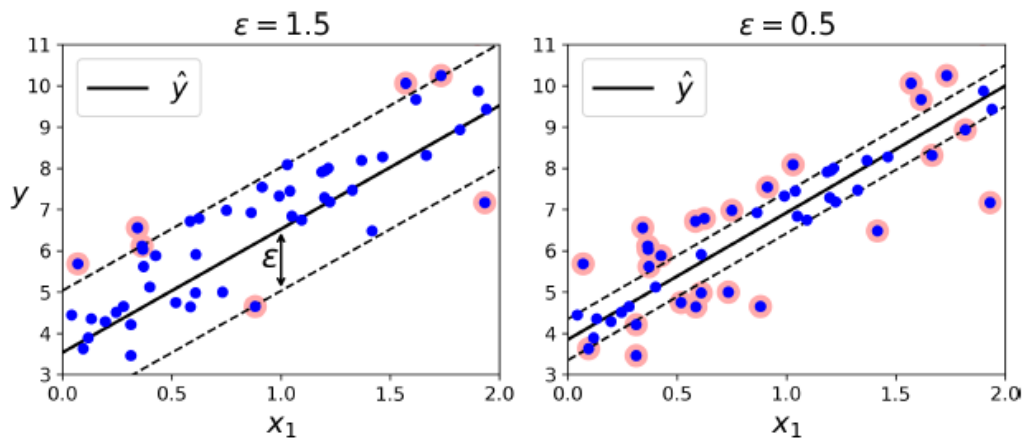
2.1.9.2 Support Vector Regression (SVR)

อัลกอริทึมของ Support Vector Machine (SVM) สามารถใช้งานได้หลากหลายหลายเอนกประสงค์ตั้งแต่ใช้แก้ปัญหาการแบ่งหมวดหมู่ทั้งรูปแบบเส้นตรงและที่ไม่ใช่เส้นตรง (Linear and Nonlinear Classification) จนถึงใช้แก้ปัญหาเชิงถดถอยทั้งรูปแบบเส้นตรงและที่ไม่ใช่เส้นตรง (Linear and Nonlinear Regression) โดยหลักการของ SVM คือจะทำการแบ่งข้อมูลออกด้วย Decision Boundary และยังกำหนดระยะห่างให้ได้มากที่สุดระหว่างเส้นแบ่งข้อมูลกับตัวข้อมูลที่อยู่ใกล้กับเส้นที่สุดด้วยเส้นประดังรูปที่ 15 (ชิตพงษ์, 2563)



รูปที่ 15 หลักการทำงานของ Support Vector Machine (ชิตพงษ์, 2563)

ในขณะที่ Support Vector Regression (SVR) จะทำตรงกันข้ามกล่าวคือแทนที่จะพยายามทำให้ระยะห่างของเส้นประมีค่ามากที่สุดเพื่อให้ข้อมูลอยู่ห่างจาก Decision Boundary และไม่เข้ามาอยู่ภายในเส้นประ แต่ใน SVR จะพยายามทำให้ข้อมูลทุกตัวอยู่ใกล้ Decision Boundary มากที่สุดหรืออยู่ในระหว่างเส้นประมากที่สุดโดยระยะระหว่างเส้นประกำหนดด้วยตัวแปร ϵ โดยในรูป 16 จะเป็นตัวอย่าง 2 ตัวแบบ SVR ที่ค่า $\epsilon = 1.5$ และ 0.5 (Géron, 2019)



รูปที่ 16 การเปรียบเทียบผลลัพธ์ของตัวแบบ SVR ที่ค่า ϵ ต่างกัน (Géron, 2019)

ในส่วนของการปรับค่าตัวแบบ Support Vector Regression (SVR) จะประกอบไปด้วย 3 ส่วน ได้แก่

- Regularization (C) ใช้สำหรับการควบคุมความซับซ้อนของตัวแบบพยากรณ์และความผิดพลาดของตัวแบบพยากรณ์ในข้อมูลฝึกฝน เมื่อค่า C มีค่าเพิ่มขึ้นจะทำให้ตัวแบบพยากรณ์มีแนวโน้มที่จะสร้าง Decision Boundary ที่ซับซ้อนและมีการวิเคราะห์เชิงถดถอยในชุดข้อมูลฝึกฝนได้ถูกต้องมากขึ้น แต่สามารถทำให้ตัวแบบพยากรณ์ Overfit กับชุดข้อมูลฝึกฝน ซึ่งอาจทำให้มีประสิทธิภาพต่ำในการพยากรณ์กับชุดข้อมูลที่ไม่เคยเห็น (Unseen Data) ในขณะที่หากค่า C มีค่าลดลง จะทำให้ตัวแบบพยากรณ์มีแนวโน้มที่จะสร้าง Decision Boundary ที่ไม่ซับซ้อน แต่สามารถทำให้ตัวแบบพยากรณ์ Underfit กับชุดข้อมูลฝึกฝน ซึ่งอาจทำให้ไม่สามารถวิเคราะห์เชิงถดถอยในชุดข้อมูลฝึกฝนได้ถูกต้อง ซึ่งอาจทำให้มีประสิทธิภาพต่ำในการพยากรณ์กับชุดข้อมูลที่ไม่เคยเห็น
- Gamma คือ การกำหนดระยะห่างของข้อมูลแต่ละตัว ซึ่งผลต่อการคำนวณเส้นแบ่งชั้น (Separation Line) หากมีค่าต่ำแสดงว่าตัวอย่างชุดข้อมูลฝึกฝนที่อยู่ห่างจากเส้นแบ่งชั้นที่เป็นไปได้จะถูกพิจารณาในการคำนวณเส้นแบ่งชั้น จะทำให้ตัวแบบพยากรณ์มีแนวโน้ม Underfit กับชุดข้อมูลฝึกฝน ซึ่งอาจทำให้ไม่สามารถจำแนกข้อมูลในชุดทดสอบได้อย่างแม่นยำ ในขณะที่หากมีค่าสูงแสดงว่าตัวอย่างข้อมูลชุดฝึกฝนที่อยู่ใกล้กับเส้นแบ่งชั้นที่เป็นไปได้ จะถูกพิจารณาในการคำนวณเส้นแบ่งชั้นจะทำให้ตัวแบบพยากรณ์มีแนวโน้ม Overfit กับชุด

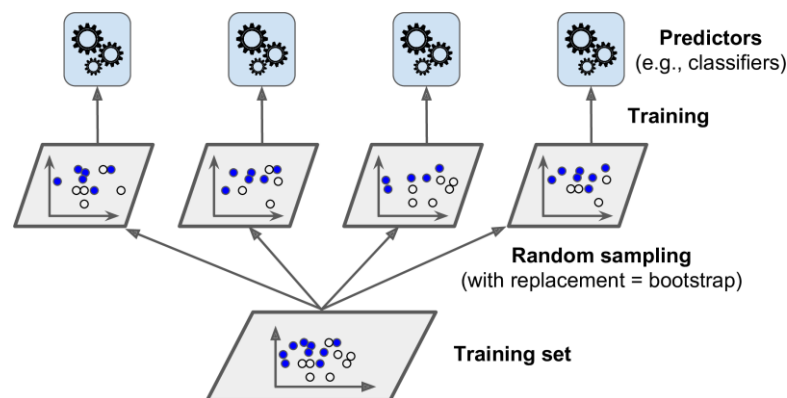
ข้อมูลฝึกฝน ซึ่งทำให้ไม่สามารถพยากรณ์วิเคราะห์เชิงถดถอยกับชุดข้อมูลทดสอบได้อย่างแม่นยำ

- Epsilon คือ การกำหนดค่าความคลาดเคลื่อนที่ยอมให้เกิดขึ้นในการทำนายของตัวแบบพยากรณ์ และค่าเป้าหมายของชุดข้อมูลทดสอบ หากค่า Epsilon ต่ำจะทำให้ตัวแบบพยากรณ์มีการทำนายที่เข้าใกล้ค่าเป้าหมายมากขึ้น ในทางกลับกันหากค่า Epsilon สูงจะทำให้ตัวแบบพยากรณ์มีความยืดหยุ่นในการทำนายและยอมให้เกิดความคลาดเคลื่อนที่มากขึ้น ดังนั้นควรเลือกและตั้งค่า Epsilon ที่เหมาะสมเพื่อควบคุมระดับความคลาดเคลื่อนที่ในการทำนายของตัวแบบพยากรณ์ให้เป็นไปตามที่ต้องการ

2.1.9.3 Ensemble Learning

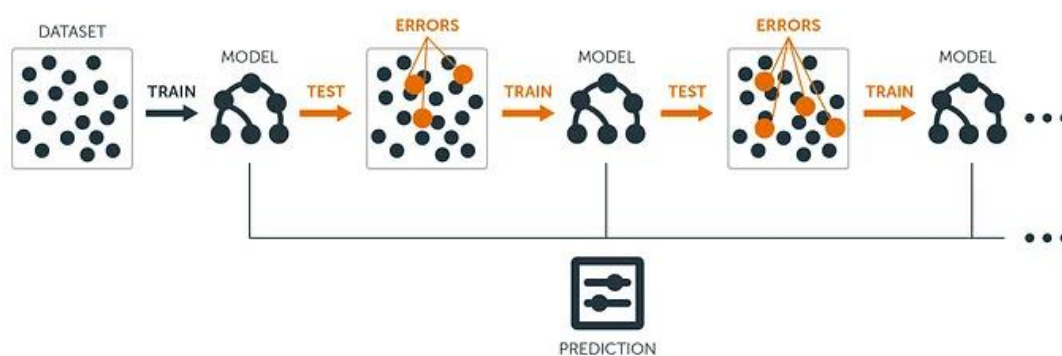
Ensemble Learning เป็นรูปแบบในการเรียนรู้ของเครื่องที่ใช้การรวมกันของตัวแบบพยากรณ์มากกว่า 1 ตัว เพื่อเป็นการเสริมจุดแข็งและปรับแก้จุดด้อยของแต่ละตัวแบบพยากรณ์ด้วยกัน เพื่อทำการพยากรณ์ที่แม่นยำมากขึ้นเมื่อเปรียบเทียบกับตัวแบบพยากรณ์เดี่ยว สำหรับ Ensemble learning สามารถแบ่งวิธีการเป็น 2 วิธีหลัก ได้แก่ Bagging และ Boosting

Bagging (Bootstrap Aggregating) คือการนำตัวแบบพยากรณ์เดี่ยวหลายตัวแบบแล้วให้แต่ละตัวแบบพยากรณ์ฝึกสอนกับ Subset ของชุดข้อมูลฝึกสอนที่แตกต่างกัน จากนั้นนำผลพยากรณ์ของแต่ละตัวแบบพยากรณ์เดี่ยวมาทำการหาค่าเฉลี่ยด้วยน้ำหนักที่เท่ากันในแต่ละตัวแบบพยากรณ์เดี่ยว เพื่อหาผลพยากรณ์สุดท้าย ตัวอย่างตัวแบบพยากรณ์ในรูปแบบนี้ เช่น Random Forest ที่เกิดจากตัวแบบ Decision Tree เดี่ยวหลายตัวมารวมกันแบบ Bagging



รูปที่ 17 หลักการทำงานของ Bagging (Géron, 2019)

Boosting คือการนำตัวแบบพยากรณ์มาทำการฝึกสอนเป็นลำดับโดยที่ตัวแบบพยากรณ์ตัวก่อนหน้าจะถูกสร้างและฝึกสอนกับชุดข้อมูลฝึกสอนและพยากรณ์หาความคลาดเคลื่อน จากนั้นตัวแบบพยากรณ์ถัดไปจะทำการปรับปรุงลดค่าความคลาดเคลื่อน ซึ่งเป็นการช่วยปรับปรุงความแม่นยำของตัวแบบพยากรณ์ และให้ผลลัพธ์ที่มีความสอดคล้องกับข้อมูลมากขึ้น ตัวอย่างตัวแบบพยากรณ์ในรูปแบบนี้ เช่น Gradient Boosting Machines (GBM), AdaBoost และ XGBoost (Géron, 2019)

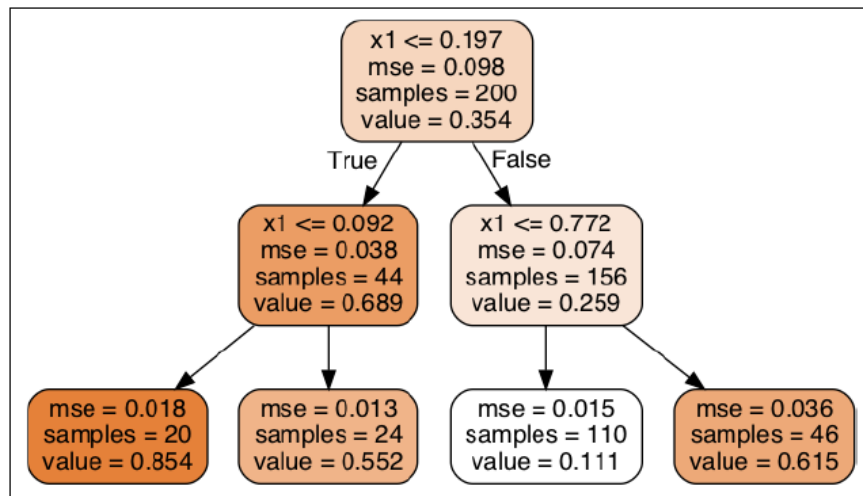


รูปที่ 18 หลักการทำงานของ Boosting

(Maria, 2017)

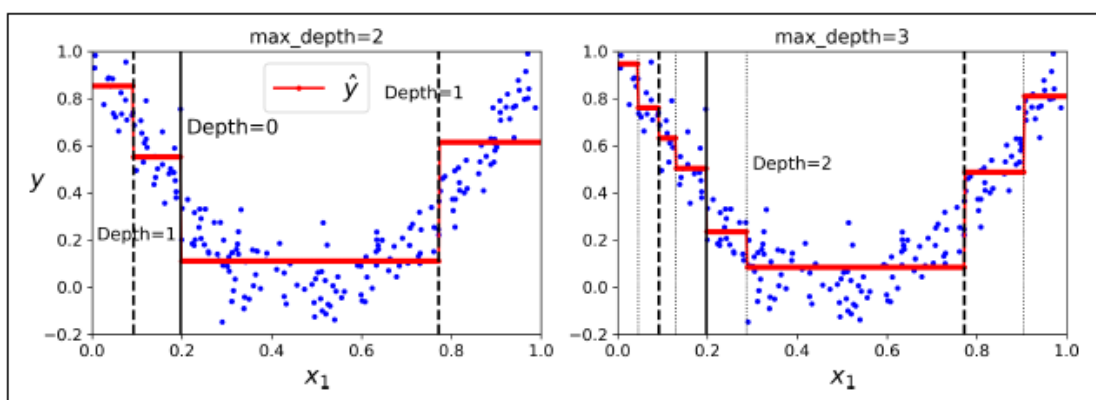
2.1.9.4 ต้นไม้ตัดสินใจ (Decision Tree)

Decision Tree จัดเป็นอัลกอริทึมที่ใช้งานได้หลากหลายเช่นเดียวกับ SVM เพราะสามารถใช้แก้ปัญหาการจำแนกหมวดหมู่ (Classification Problem) และปัญหาเชิงถดถอย (Regression Problem) และเป็นพื้นฐานสำคัญของอัลกอริทึมป่าสุ่ม (Random Forest) โดยตัวอย่างของต้นไม้ตัดสินใจเชิงถดถอยเป็นดังรูปที่ 19 ซึ่งจะมีลักษณะเหมือนกับต้นไม้โดยที่ชั้นบนสุดคือราก (Root Node) และปลายสุดคือใบไม้ (Leaf Node) โดยสามารถกำหนดระดับของต้นไม้ (Depth) ได้



รูปที่ 19 หลักการทำงานของ Decision Tree (Géron, 2019)

รูปแบบของต้นไม้ตัดสินใจเชิงถดถอยจะคล้ายกับต้นไม้ตัดสินใจสำหรับการแบ่งหมวดหมู่ แต่จะต่างกันในแต่ละ Node จะถูกใช้ในการทำนายค่าของข้อมูล ตัวอย่างการทำงานของต้นไม้ตัดสินใจเช่น หากกำหนดค่าของ x_1 เท่ากับ 0.6 เมื่อไล่ลำดับตั้งแต่ตำแหน่งบนสุดของต้นไม้ (ราก) จนกระทั่งถึงผลลัพธ์ (ใบไม้) พบว่าจะได้ผลลัพธ์เท่ากับ 0.1106 โดยค่าดังกล่าวได้มาจากการเฉลี่ยชุดข้อมูล Training จำนวน 110 ข้อมูลที่อยู่ใน Leaf Node นั้นและได้ค่า MSE เท่ากับ 0.00151 ผลลัพธ์การพยากรณ์ของตัวแบบทั้งหมดสามารถแสดงได้ในรูปที่ 20 โดยจะพบว่าหากกำหนดค่า Depth มากเกินไปจะทำให้ผลการพยากรณ์ใกล้เคียงกับชุดข้อมูล Training มากขึ้นและเกิดการ Overfitting ขึ้นได้ (Géron, 2019)

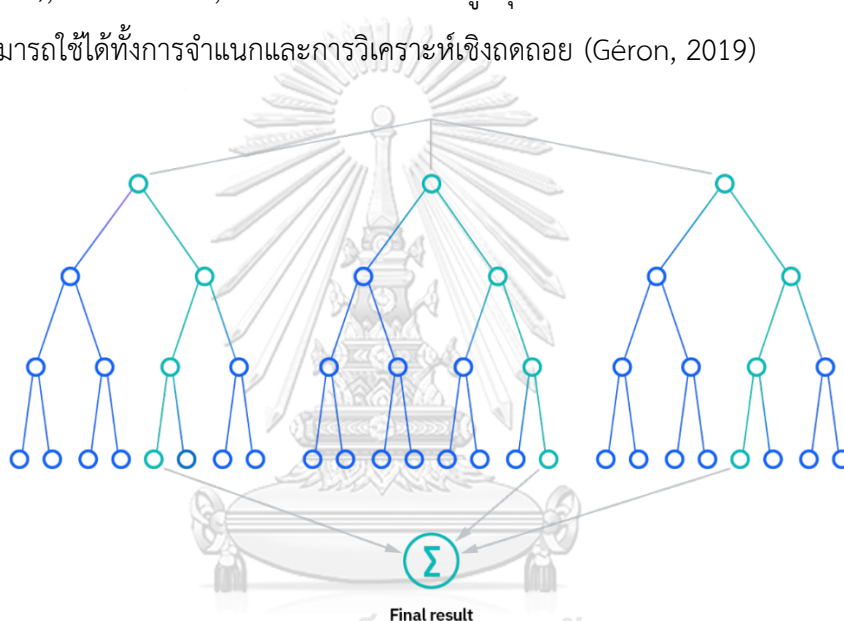


รูปที่ 20 การเปรียบเทียบผลลัพธ์ของตัวแบบ Decision Tree ที่ค่า Max Depth ต่างกัน (Géron, 2019)

2.1.9.5 ป่าสุ่ม (Random Forests)

Random Forest จัดเป็นรูปแบบหนึ่งของ Ensemble Learning ที่เป็นการเทรนตัวแบบซ้ำหลายครั้งบนข้อมูลชุดเดียวกันแต่เปลี่ยนสัดส่วนไปในทุกครั้งแล้วเลือกผลลัพธ์ที่ดีที่สุดออกมา โดย Random Forest จะเปรียบเสมือนป่าที่ประกอบไปด้วย Decision Tree หลายต้นที่มีการแบ่งสัดส่วนข้อมูลที่ไม่เหมือนกันด้วยวิธี Bagging (การสุ่มเลือกข้อมูลที่สามารถสุ่มซ้ำได้) หรือ Pasting (การสุ่มเลือกข้อมูลที่ทำซ้ำ) มารวมกันแล้วเลือกผลลัพธ์ที่ดีที่สุดออกมา

พารามิเตอร์หลักของอัลกอริทึม Random Forest ประกอบด้วย Node Size (จำนวนโหนด), จำนวนต้นไม้, และจำนวนฟีเจอร์ที่ถูกสุ่มตัวอย่าง ซึ่งตัวแบบพยากรณ์ Random Forest สามารถใช้ได้ทั้งการจำแนกและการวิเคราะห์เชิงถดถอย (Géron, 2019)



รูปที่ 21 หลักการทำงานของ Random Forest

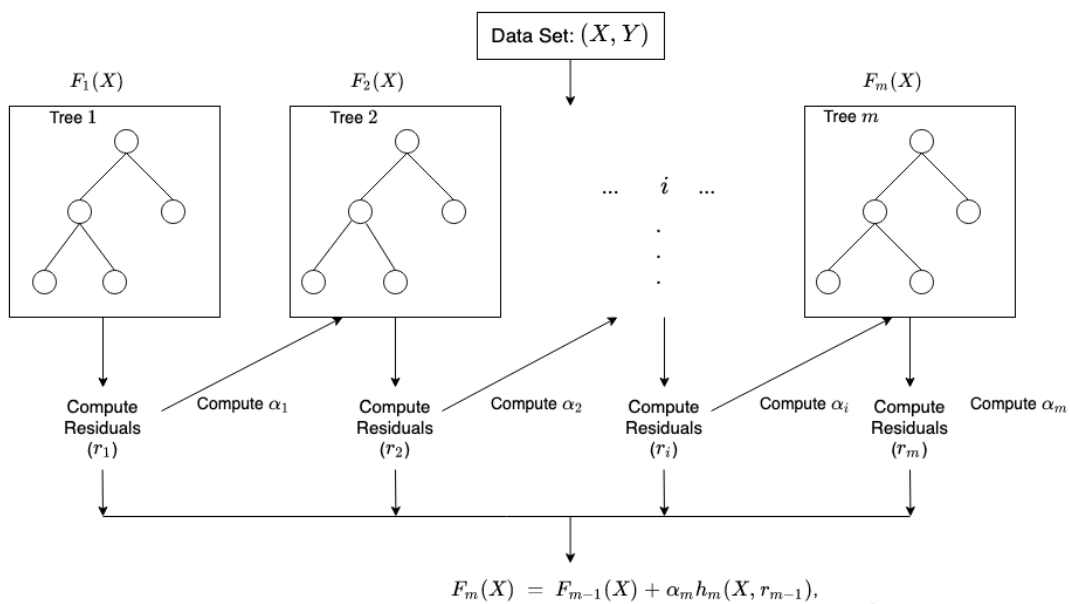
(IBM, 2023)

2.1.9.6 XGBoost (eXtreme Gradient Boosting)

XGBoost เป็นตัวอย่างของ Gradient Boosted Trees ที่ได้รับความนิยมและมีประสิทธิภาพสูงใน การทำนายตัวแปรเป้าหมายจากการรวมค่าประมาณการชุดของตัวแบบพยากรณ์ที่เรียกว่า "Weak Learners" หรือตัวแบบพยากรณ์ที่อ่อนแอกว่า เป็นตัวแบบพยากรณ์ย่อย

Gradient Boosting สำหรับการวิเคราะห์เชิงถดถอย ตัวแบบพยากรณ์ย่อยคือ Regression Trees โดยแต่ละ Regression Tree จะทำการเชื่อมข้อมูลเข้าสู่หนึ่งใน Leaf ของต้นไม้ โดยในแต่ละ Leaf จะมีค่าคะแนน (Score) โดย XGBoost จะลดค่าฟังก์ชันที่เรียกว่า Objective

Function ที่มีการ Regularize (L1 และ L2) ซึ่งรวมกันไปในการประมาณค่าของฟังก์ชัน Convex Loss (โดยอ้างอิงความแตกต่างระหว่างค่าที่คาดการณ์และค่าเป้าหมาย) และ Penalty Term สำหรับความซับซ้อนของตัวแบบพยากรณ์ (ฟังก์ชันของ Regression Tree) กระบวนการฝึกอบรมจะเป็นรูปแบบการทำซ้ำทีละขั้นตอนโดยเพิ่มต้นไม้ใหม่ที่ทำนายค่าความผิดพลาดของต้นไม้ก่อนหน้า ซึ่ง จะทำการรวมกับต้นไม้ก่อนหน้าเพื่อทำการทำนายสุดท้าย

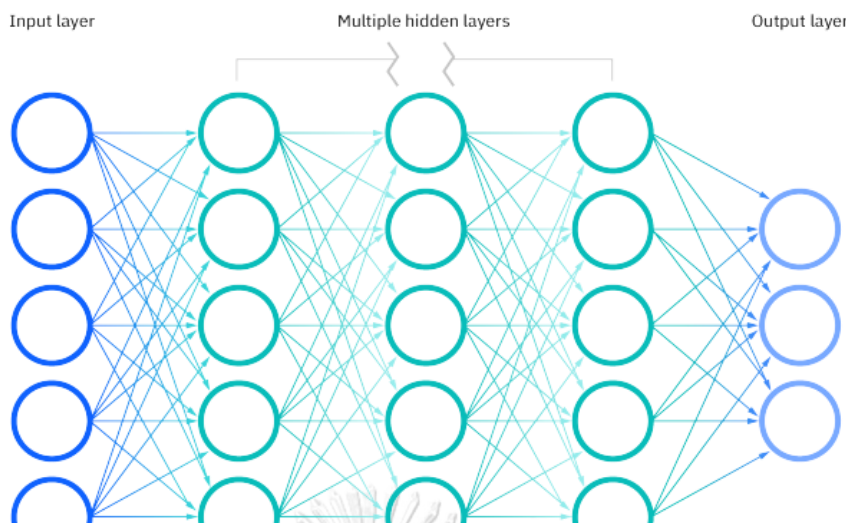


รูปที่ 22 หลักการทำงานของ XGBoost
(AWS Developer Guide, 2023)

จุฬาลงกรณ์มหาวิทยาลัย

2.1.9.7 โครงข่ายประสาทเทียม (Artificial Neural Network)

เครือข่ายประสาทเทียม หรือ Artificial Neural Networks (ANNs) เป็นส่วนหนึ่งของการเรียนรู้ของเครื่องและเป็นส่วนสำคัญของอัลกอริทึมการเรียนรู้เชิงลึก (Deep Learning) ชื่อและโครงสร้างของเครือข่ายประสาทเทียมมีต้นแบบมาจากสมองมนุษย์ โดยจำลองวิธีการส่งสัญญาณระหว่างเซลล์ประสาท เครือข่ายประสาทเทียมประกอบด้วยชั้นโหนดที่ประกอบด้วยชั้นป้อนข้อมูล (Input Layer) ชั้นซ่อน (Hidden Layer) หนึ่งชั้นหรือหลายชั้น และชั้นผลลัพธ์ (Output Layer) แต่ละโหนดหรือเซลล์ประสาทเทียมจะเชื่อมต่อกัน โดยมีน้ำหนัก (Weight) และ Threshold แต่ละโหนด หากผลลัพธ์ของโหนดใดเกินค่า Threshold ที่กำหนดไว้ โหนดนั้นจะถูกเปิดใช้งานและส่งข้อมูลไปยังชั้นถัดไปของเครือข่าย มิฉะนั้น ข้อมูลจะไม่ถูกส่งผ่านไปยังชั้นถัดไปของเครือข่าย โดยมีรูปแบบดังรูปที่ 2-20



รูปที่ 23 ส่วนประกอบและโครงสร้างของตัวแบบพยากรณ์โครงข่ายประสาทเทียม (IBM, 2023)

โครงข่ายประสาทเทียมใช้ชุดข้อมูลฝึกสอนเพื่อเรียนรู้และปรับปรุงความแม่นยำตามเวลา เมื่ออัลกอริทึมเรียนรู้และได้ถูกปรับแต่งให้มีความแม่นยำ โครงข่ายประสาทจะเป็นเครื่องมือที่มีประสิทธิภาพสูงในวิทยาการคอมพิวเตอร์และปัญญาประดิษฐ์ ช่วยให้สามารถจัดหมวดหมู่และจัดกลุ่มข้อมูลได้อย่างรวดเร็ว งานที่เกี่ยวข้องกับการรู้จำเสียงหรือการรู้จำภาพอาจใช้เวลาไม่กี่นาทีเท่านั้นเมื่อเปรียบเทียบกับกระบวนการระบุตัวตนด้วยมนุษย์ผู้เชี่ยวชาญ หนึ่งในโครงข่ายประสาทที่เป็นที่รู้จักดีคือ อัลกอริทึมการค้นหาของ Google

โครงข่ายประสาทเทียมสามารถแบ่งประเภทตามวัตถุประสงค์การใช้งานได้ดังนี้
เปอร์เซปตรอน (Perceptron) เป็นโครงข่ายประสาทเทียมที่เก่าแก่ที่สุด ถูกสร้างขึ้นโดย Frank Rosenblatt เมื่อปี 1958

โครงข่ายประสาทก้ำวหน้า (Feedforward Neural Networks) หรือมัลติเลเยอร์เพอร์เซปตรอน (Multi-layer perceptrons: MLPs) ประกอบด้วยชั้นนำเข้าข้อมูล (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นผลลัพธ์ (Output Layer) แม้ว่าโครงข่ายประสาทเหล่านี้จะเรียกว่า MLPs แต่ส่วนประกอบสำคัญคือ เซลล์ประสาทซิกมอยด์ (Sigmoid Neurons) ไม่ใช่เปอร์เซปตรอน เนื่องจากปัญหาส่วนใหญ่ในความเป็นจริงเป็นปัญหาที่ไม่เป็นเชิงเส้น

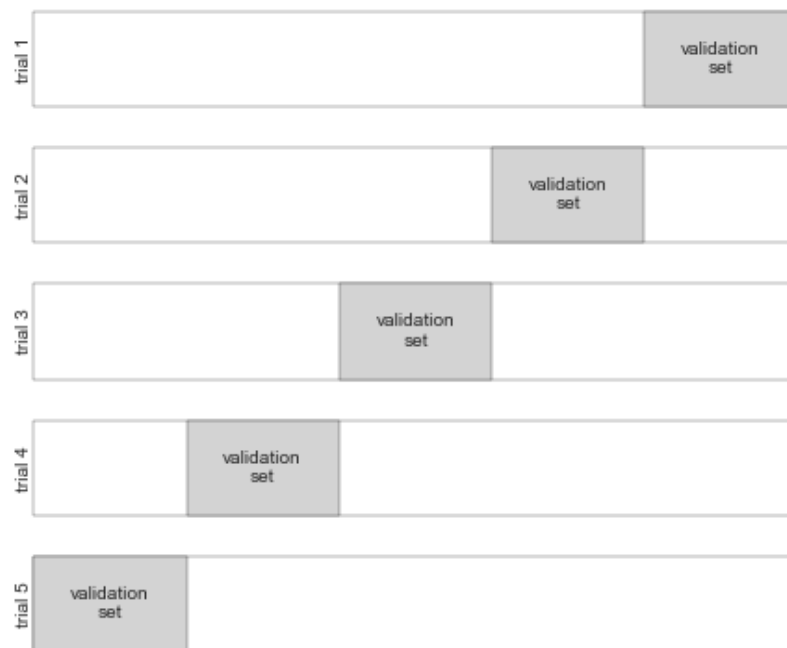
โครงข่ายประสาทคอนโวลูชัน (Convolutional Neural Networks: CNNs) มีลักษณะคล้ายกับโครงข่ายประสาทก้ำวหน้า แต่มักถูกนำมาใช้สำหรับการรู้จำภาพ การรู้จำลวดลาย และ/หรือการประมวลผลภาพคอมพิวเตอร์ โครงข่ายเหล่านี้ใช้หลักการจากการพิจารณาเชิงเส้น

Recurrent Neural Networks อัลกอริทึมการเรียนรู้ชนิดนี้ได้รับการใช้งาน โดยเฉพาะเมื่อใช้ข้อมูลชุดเวลา (Time-series Data) เพื่อทำนายผลลัพธ์ในอนาคต เช่นการทำนายตลาดหุ้นหรือการพยากรณ์ยอดขาย

2.1.9.8 Hyperparameter and Model Validation

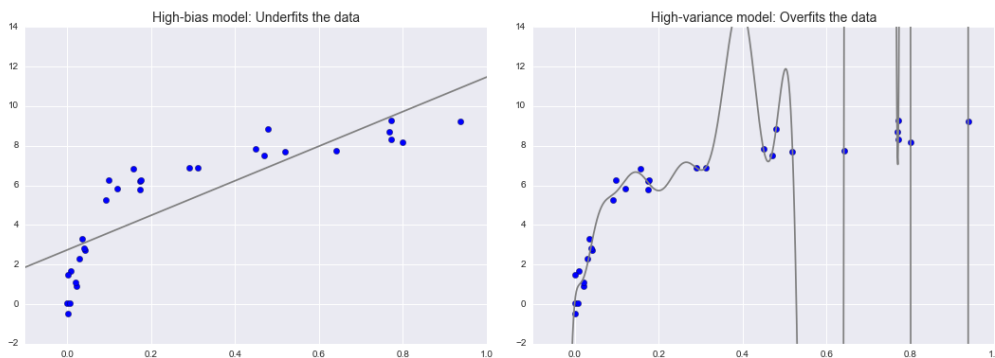
การสร้างตัวแบบพยากรณ์ การกำหนดค่าพารามิเตอร์ของตัวแบบให้เหมาะสมเป็นส่วนที่สำคัญเพื่อให้ได้ผลลัพธ์ที่น่าเชื่อถือและแม่นยำมากที่สุด โดยจะดำเนินการตรวจสอบโมเดลและพารามิเตอร์ว่ามีเหมาะสมกับข้อมูลที่มีหรือไม่ โดยวิธีการต่าง ๆ เช่น

- การแบ่งชุดข้อมูลเป็น 2 ส่วน ได้แก่ ชุดข้อมูลฝึกสอนที่จะเป็นชุดข้อมูลที่ใช้สร้างตัวแบบพยากรณ์ และชุดข้อมูลที่ตัวแบบพยากรณ์ยังไม่เคยเห็น (Unseen Data Set) เพื่อทำการวัดผลตัวแบบพยากรณ์จากการเปรียบเทียบผลพยากรณ์กับข้อมูลที่ตัวแบบพยากรณ์ยังไม่เคยเห็นเพื่อหาความคลาดเคลื่อนที่จะใช้เป็นตัวชี้วัดความแม่นยำของตัวแบบพยากรณ์ ซึ่งหากไม่ทำการแบ่งชุดข้อมูล ผลลัพธ์ที่ได้จะพบว่าตัวแบบพยากรณ์มีความแม่นยำ 100% ทุกครั้ง ทำให้ไม่ทราบความแม่นยำที่จะใช้ในการพยากรณ์ข้อมูลไปข้างหน้า แต่ในงานวิจัยฉบับนี้จะทำการแย่งข้อมูลออกเป็น 3 ส่วนดังที่ได้กล่าวมาในข้างต้นในหัวข้อ 2.1.3
- การทำ Cross-validation เป็นเทคนิคที่ใช้ในการประเมินความสามารถของตัวแบบพยากรณ์ โดยการแบ่งข้อมูลออกเป็นชุดฝึกฝน (Training Set) และชุดทดสอบ (Test Set) หลายชุด เพื่อทำการประเมินตัวแบบพยากรณ์ในแต่ละครั้ง วิธีการสำหรับ Cross-validation ที่ได้รับความนิยมมากที่สุดคือ K-fold Cross-validation ตัวอย่างดังรูปที่ 24 ซึ่งมีขั้นตอนการทำงานตั้งแต่ การแบ่งชุดข้อมูลออกเป็น K ชุดเท่าๆ กัน จากนั้นในแต่ละรอบ เลือกใช้หนึ่งชุดเป็นชุดทดสอบ และใช้ชุดอื่นๆ เป็นชุดฝึกฝน (ฝึกตัวแบบพยากรณ์ด้วยชุดฝึกฝนและทำนายผลลัพธ์บนชุดทดสอบ) แล้วทำการบันทึกผลลัพธ์การทำนายในแต่ละรอบ และคำนวณค่าเฉลี่ยของความแม่นยำที่ได้จากการทำนายในทุกรอบเพื่อประเมินความสามารถของตัวแบบพยากรณ์

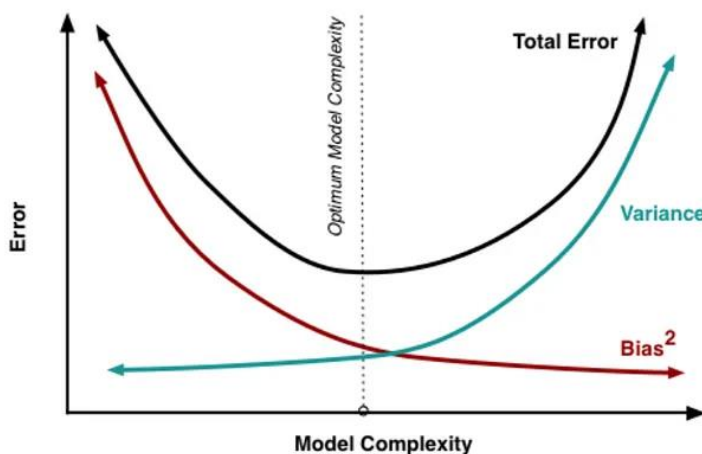


รูปที่ 24 ตัวอย่างรูปแบบของ k-fold Cross-validation ที่ $k = 5$

การเลือกตัวแบบพยากรณ์ที่เหมาะสมที่สุด (Selecting the Best Model) ที่ตัวแบบพยากรณ์ไม่ซับซ้อนเกินไป มีข้อมูลปัจจัยที่เพียงพอ สามารถพิจารณาได้จาก Bias (ความถูกต้องในการแสดงผลลัพธ์ของตัวแบบพยากรณ์) และ Variance (ความยืดหยุ่นของตัวแบบพยากรณ์ในการปรับตัวเข้ากับข้อมูล) ดังตัวอย่างในรูปที่ 25 พบว่าตัวแบบพยากรณ์ในรูปซ้ายไม่สามารถปรับรูปแบบให้เข้ากับลักษณะของข้อมูลชุดฝึกสอน เพราะมีค่า Bias ที่สูง จึงไม่สามารถจับความซับซ้อนหรือลักษณะพิเศษของข้อมูลได้อย่างถูกต้อง ทำให้เกิดความผิดพลาดเมื่อเจอข้อมูลใหม่ หรือเรียกอีกอย่างว่า Underfitting ในขณะที่ตัวแบบพยากรณ์ในรูปขวาสามารถปรับตัวเข้ากับลักษณะข้อมูลชุดฝึกสอนได้ทุกตัว เพราะมีค่า Variance ที่สูง ทำให้ตัวแบบพยากรณ์จะมีความสามารถในการฝึกสอนข้อมูลซับซ้อน แต่อาจทำให้เกิดความผิดพลาดสูงเมื่อเจอข้อมูลใหม่ หรือเรียกอีกอย่างว่า Overfitting ดังนั้นตัวแบบพยากรณ์ที่เหมาะสมที่สุดคือตัวแบบพยากรณ์ที่อยู่ในจุดสมดุลระหว่าง Bias และ Variance ดังรูปที่ 26



รูปที่ 25 การเปรียบเทียบระหว่างตัวแบบพยากรณ์ที่มีค่า Bias สูง และตัวแบบที่มีค่า Variance สูง



รูปที่ 26 กราฟแสดงจุดสมดุลระหว่าง Bias และ Variance

การปรับค่า Hyperparameter เพื่อให้ได้ผลลัพธ์ที่แม่นยำที่สุดซึ่งสามารถทำได้ด้วยการทดสอบทีละค่าหรือเรียกอีกอย่างว่าการทำ Grid Search ซึ่งเป็นการทดลองทุกค่า Hyperparameter ในช่วงที่กำหนด (Grid) โดยกระบวนการ Grid Search จะทำการสร้าง Grid หรือตารางที่ประกอบด้วยค่า Hyperparameter ที่ต้องการทดลอง แล้วทดสอบตัวแบบพยากรณ์ในแต่ละคอมบิเนชันของพารามิเตอร์ทั้งหมดใน Grid โดยประเมินผลลัพธ์ตามเกณฑ์ที่กำหนด เช่น ค่าประสิทธิภาพของตัวแบบพยากรณ์ หรือค่าความคลาดเคลื่อนระหว่างค่าที่ทดลองและค่าจริง เพื่อเลือกค่าพารามิเตอร์ที่ให้ผลลัพธ์ที่ดีที่สุด (VanderPlas, 2016)

2.2 งานวิจัยที่เกี่ยวข้อง

เนื่องจากงานวิจัยการพยากรณ์ราคาเม็ดพลาสติก PP ขณะนี้ยังไม่แพร่หลายจึงทำการอ้างอิงการพยากรณ์จากสินค้าที่มีลักษณะคล้ายคลึงกันแทน เช่น เม็ดพลาสติกชนิดอื่น ยาง และน้ำมัน เนื่องจากเป็นสินค้าปิโตรเคมีเช่นเดียวกับพลาสติกจึงมีคุณสมบัติคล้ายคลึงกัน

2.2.1 งานวิจัยการพยากรณ์ราคาเม็ดพลาสติกชนิดอื่น

ณัฐกมล (2550) ได้ทำการวิจัยกรณีศึกษาบริษัทพลาสติกพีวีซี และหนังเทียม โดยพบว่าทางบริษัทมีปัญหาในเรื่องของระดับสินค้าในคลังที่เกิดการ Understock และ Overstock ซึ่งมีสาเหตุมาจากตัวแบบพยากรณ์ที่ทางบริษัทใช้นั้นมีความแม่นยำไม่เพียงพอเพราะเป็นตัวแบบพยากรณ์เชิงคุณภาพ (Qualitative Forecasting) ซึ่งเป็นการพยากรณ์ที่อาศัยประสบการณ์ของพนักงานในบริษัทในการทำนาย ในส่วนของวิธีการแก้ไขปัญหาดังกล่าวจะใช้โครงข่ายประสาทเทียมเพื่อพยากรณ์ยอดขายของสินค้าตัวอย่างแทนตัวแบบพยากรณ์เก่าเพื่อเป็นการเพิ่มความแม่นยำในการพยากรณ์ ปริมาณความต้องการซึ่งจะช่วยให้เพิ่มผลกำไรให้บริษัทได้ โดยในการพยากรณ์จะใช้ข้อมูล อัตราดอกเบี้ย อัตราว่างงาน ดัชนีราคาผู้บริโภค ราคาน้ำมัน ดัชนีมวลรวมประชาชาติ อัตราการบริโภคภายในครัวเรือน อัตราการผลิตไฮดรอกซีอะครีเลต อัตราการนำเข้า และอัตราการส่งออก มาสร้างตัวแบบพยากรณ์เพื่อทำการพยากรณ์ยอดขายในอดีตแล้วนำมาเปรียบเทียบกับข้อมูลจริงโดยประเมินจากค่าร้อยละของความผิดพลาด (Percentage Error: PE) และค่าเฉลี่ยผิดพลาดยกกำลังสอง (Mean Squared Error : MSE) ผลลัพธ์ที่ได้พบว่าตัวแบบการพยากรณ์โครงข่ายประสาทเทียมนั้นมีความแม่นยำมากกว่าตัวแบบเดิมที่ทางบริษัทใช้ โดยในตัวแบบโครงข่ายประสาทเทียมมีค่าร้อยละของความผิดพลาดเท่ากับ 1.09 และค่าเฉลี่ยผิดพลาดยกกำลังสองเท่ากับ 18.78 ในขณะที่ตัวแบบพยากรณ์แบบเก่ามีค่าร้อยละของความผิดพลาดเท่ากับ 5.163 และค่าเฉลี่ยผิดพลาดยกกำลังสองเท่ากับ 29.165 และเมื่อทางบริษัทได้นำตัวแบบโครงข่ายประสาทเทียมไปปรับใช้ให้การวางแผนการผลิตสอดคล้องกับการพยากรณ์ยอดขายทำให้บริษัทสามารถลดปัญหาสินค้าคงคลังทำให้สามารถลดค่าใช้จ่ายได้ถึง 2,254,000 บาทหรือประมาณ 28 เปอร์เซ็นต์

วรรณิภา และคณะ (2552) ได้ทำการศึกษารูปแบบการพยากรณ์ราคาเม็ดพลาสติกพอลิเอทิลีนโดยมีวัตถุประสงค์เพื่อสร้างแบบจำลองในการพยากรณ์ระดับราคาและระดับการเปลี่ยนแปลงราคาเม็ดพลาสติกพอลิเอทิลีน 3 ชนิดคือ HDPE LLDPE และ LDPE เพื่อนำตัวแบบการพยากรณ์ไปใช้ประโยชน์ในการดำเนินธุรกิจ เช่น การวางแผนจัดซื้อเม็ดพลาสติกเพื่อรองรับกับสถานการณ์ราคาเม็ดพลาสติกที่มีความผันผวน ขั้นตอนในการศึกษามีดังนี้ (1) การเก็บข้อมูลราคาเม็ดพลาสติก ราคาน้ำมันดิบและตัวเลขทางเศรษฐกิจตั้งแต่เดือนมกราคม พ.ศ. 2543 จนถึง เดือนพฤษภาคม พ.ศ. 2552

(2) นำข้อมูลที่รวบรวมมาทดสอบว่าเป็นปัจจัยที่เกี่ยวข้องกับราคาเม็ดพลาสติกหรือไม่โดยการหาค่าสหสัมพันธ์ (3) สร้างตัวแบบการพยากรณ์ราคาเม็ดพลาสติกด้วยวิธีการวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Regression Analysis) (4) ประเมินประสิทธิภาพของตัวแบบการพยากรณ์ ผลการศึกษาพบว่าระดับราคาเม็ดพลาสติกทั้ง 3 ชนิดนั้นมีความสัมพันธ์กับปัจจัยนำเข้าได้แก่ ระดับสินค้าคงคลังทางธุรกิจ ยอดค้าปลีกที่ไม่รวมยานยนต์ ยอดขายบ้านใหม่ จำนวนการขออนุญาตสร้างบ้านใหม่ ดัชนีความเชื่อมั่นผู้บริโภค ดัชนีราคาผู้บริโภคและผู้ผลิต ดัชนี S&P500 ในส่วนของปัจจัยตามจะประกอบไปด้วย อัตราการว่างงานและปริมาณการผลิตภาคอุตสาหกรรม และปัจจัยที่เกิดขึ้นพร้อมๆ กับราคาเม็ดพลาสติก ได้แก่ ราคาน้ำมันดิบ ปริมาณการสั่งซื้อสินค้าคงทนและดัชนี S&P GSCI Agriculture Index จากการประเมินประสิทธิภาพของตัวแบบการพยากรณ์ให้ผลลัพธ์ดังนี้ Adjusted $R^2 > 0.9$ ค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Squared of Error: MSE) มีค่าอยู่ระหว่าง 0.010 – 0.011 และ ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) มีค่าอยู่ระหว่าง 1.125% – 1.170% จากผลการศึกษาค้นคว้าครั้งนี้ทางผู้จัดทำได้เห็นว่าผลลัพธ์จากการพยากรณ์ในงานวิจัยฉบับนี้มีความแม่นยำที่สูงมากและสามารถนำวิธีการวิเคราะห์การถดถอยเชิงพหุคูณมาประยุกต์ใช้กับงานวิจัยฉบับนี้ได้

ปฏิมาพร (2556) ได้ทำการวิจัยการพยากรณ์ยอดขายสำหรับผู้แทนจำหน่ายเม็ดพลาสติก เพื่อให้แผนกขายและการตลาดสามารถวางแผนการสั่งซื้อสินค้าจากผู้ผลิตได้ดีขึ้นเพื่อป้องกันการขาดทุน โดยในงานวิจัยฉบับนี้จะใช้ข้อมูลยอดขายรายเดือน ตั้งแต่เดือนมกราคม พ.ศ. 2547 จนถึงเดือนธันวาคม พ.ศ. 2553 ในการสร้างตัวแบบพยากรณ์ และใช้ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2554 จนถึง เดือนมิถุนายน พ.ศ. 2555 ในการหาค่า MAPE ของแต่ละตัวแบบเพื่อใช้ในการเปรียบเทียบความแม่นยำระหว่างตัวแบบ โดยวิธีการพยากรณ์ที่ใช้ในงานวิจัยฉบับนี้ได้แก่ นาอ็พ ค่าเฉลี่ยเคลื่อนที่ปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์ บอซซ์-เจนกินส์ โครงข่ายประสาทเทียม และวิธีพันธุกรรมและโครงข่ายประสาทเทียม ซึ่งในงานวิจัยฉบับนี้มีสินค้าอยู่ 5 อย่างโดยผลลัพธ์ความแม่นยำของแต่ละตัวแบบในแต่ละสินค้าเป็นดังตารางที่ 2

ตารางที่ 2 ผลลัพธ์ความแม่นยำจากการพยากรณ์จากงานวิจัยของปฏิมาพร (2556)

รหัสสินค้า	MAPE (%)						
	นาอียฟ	ค่าเฉลี่ยเคลื่อนที่	ปรับเรียบแบบโฮลท์และวินเทอร์	บอกซ์-เจนกินส์	โครงข่ายประสาทเทียม 1 ชั้นซ่อน	โครงข่ายประสาทเทียม 2 ชั้นซ่อน	วิธีพันธุกรรมและโครงข่ายประสาทเทียม
1101	10.14	15.15	15.17	15.08	9.35	6.61	6.23
2101	13.48	15.46	16.46	14.91	8.21	9.89	8.52
2301	37.28	37.97	43.24	27.87	20.37	21.23	20.66
1202	30.75	22.98	22.66	24.99	14.82	16.46	13.84
2302	32.66	39.46	48.53	32.09	27.56	13.27	20.77

จากผลลัพธ์ความแม่นยำของการพยากรณ์ทำให้สามารถสรุปได้ว่าตัวแบบโครงข่ายประสาทเทียม และตัวแบบพันธุกรรมและโครงข่ายประสาทเทียม เป็นตัวแบบที่มีความแม่นยำมากที่สุด

Adisak and Vichai (2018) ได้ทำการวิจัยเพื่อสร้างตัวแบบการพยากรณ์ทำนายราคาของเม็ดพลาสติกพอลิไวนิลคลอไรด์ (Polyvinyl Chloride: PVC) ราคาที่ได้รับมาจะมีผลจากปัจจัย 2 อย่าง ได้แก่ราคาน้ำมันดิบ และอัตราแลกเปลี่ยน ซึ่งได้ทำการรวบรวมข้อมูลราคาเม็ดพลาสติกและปัจจัยเป็นระยะเวลา 10 ปีตั้งแต่เดือนมกราคม ค.ศ. 2008 จนถึง เดือนธันวาคม ค.ศ. 2017 โดยในงานวิจัยฉบับนี้จะใช้ตัวแบบการพยากรณ์ ARIMA เพื่อพยากรณ์ราคาเม็ดพลาสติกโดยจะใช้ราคาน้ำมันดิบและอัตราแลกเปลี่ยนเป็นตัวแปรอิสระ ขั้นตอนในการทำงานวิจัยฉบับนี้ประกอบไปด้วย (1) ศึกษาและรวบรวมข้อมูลที่เกี่ยวข้องกับปัจจัยที่มีผลต่อราคาเม็ดพลาสติก (2) ทำการวิเคราะห์ตัวแบบการพยากรณ์ ARIMA เพื่อพยากรณ์ราคาเม็ดพลาสติก (3) วิเคราะห์ตัวแบบการพยากรณ์ ARIMA สำหรับราคาเม็ดพลาสติกด้วยอนุกรมเวลาของตัวแปรอิสระ (4) การพยากรณ์ตัวแปรอิสระ (5) ทำการพยากรณ์ราคาเม็ดพลาสติก PVC จากนั้นวัดประสิทธิภาพของตัวแบบโดยใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination: R^2) ผลการวิจัยแสดงให้เห็นว่า ARIMA (1,1,1) เป็นตัวแบบการพยากรณ์ที่เหมาะสมในการพยากรณ์ราคาเม็ดพลาสติก PVC โดยมีค่า MAPE อยู่ที่ 0.511% และสำหรับ R^2 มีค่าอยู่ที่ 0.923 ผลลัพธ์จาก

งานวิจัยฉบับนี้ทางผู้จัดทำได้เห็นว่าการพยากรณ์ด้วยวิธี ARIMA มีความแม่นยำที่ใกล้เคียงกับวิธีถดถอยเชิงพหุคูณ ทำให้ในงานวิจัยฉบับนี้จะนำวิธี ARIMA มาใช้ในการพยากรณ์ด้วยเช่นกัน

Kalinggo and Zulkarnain (2020) ได้ทำการวิจัยการพยากรณ์ข้อมูลอนุกรมเวลาของราคาเม็ดพลาสติก HDPE โดยได้ทำการรวบรวมข้อมูลราคาเม็ดพลาสติก HDPE ตั้งแต่เดือนมกราคม ค.ศ. 2014 จนถึง เดือนธันวาคม ค.ศ. 2019 โดยในงานวิจัยฉบับนี้มีจุดประสงค์เพื่อการเปรียบเทียบความแม่นยำระหว่างตัวแบบพยากรณ์ ARIMA ตัวแบบพยากรณ์โครงข่ายประสาทเทียม (Neural Network) โดยมีขั้นตอนการวิจัยได้แก่ (1) กำหนดหัวข้อ ปัญหา วัตถุประสงค์ และขอบเขตการดำเนินงาน (2) ศึกษาและรวบรวมทฤษฎีที่เกี่ยวข้อง (3) เก็บรวบรวมข้อมูล (4) ทดสอบลักษณะข้อมูล (5) ทำการพยากรณ์ด้วยตัวแบบ ARIMA และตัวแบบ Neural Network (6) เปรียบเทียบความแม่นยำระหว่าง 2 ตัวแบบจากค่า MAPE (7) ทำการสรุปผลและข้อเสนอแนะในการวิจัยต่อไปในอนาคต ผลลัพธ์ที่ได้คือในงานวิจัยนี้ตัวแบบพยากรณ์ ARIMA ให้ค่าความแม่นยำที่สูงกว่าตัวแบบ Neural Network โดยตัวแบบ ARIMA ที่ดีที่สุดคือ (1,2,0) ให้ค่า MAPE เท่ากับ 4.3% ในขณะที่ตัวแบบ Neural Network ที่ดีที่สุดให้ค่า MAPE เท่ากับ 9.2% จากงานวิจัยฉบับนี้ทำให้สามารถสรุปได้ว่าตัวแบบที่ใช้ในพยากรณ์ไม่จำเป็นต้องใช้ตัวแบบที่ซับซ้อนเสมอไปเพราะในงานวิจัยฉบับนี้ตัวแบบ ARIMA มีความแม่นยำที่สูงมากเหมาะแก่การนำมาใช้พยากรณ์ราคาเม็ดพลาสติก แต่เนื่องจากรูปแบบของข้อมูลที่มีแค่แนวโน้ม ไม่ได้เหมือนกับพยากรณ์ราคาเม็ดพลาสติก PP ที่ข้อมูลมีลักษณะทั้งแนวโน้มและฤดูกาล ทำให้ยังไม่สามารถสรุปได้ว่าควรนำตัวแบบ SARIMA มาใช้แทนตัวแบบ Neural Network นอกจากนั้นหากการสร้างตัวแบบผสมอาจทำให้มีความเป็นไปได้ที่จะได้ตัวแบบที่มีความแม่นยำสูงกว่าตัวแบบเดี่ยว

2.2.2 งานวิจัยเกี่ยวกับน้ำมันปาล์ม

กิตติภูมิ และคณะ (2562) ใช้วิธีการพยากรณ์ข้อมูลอนุกรมเวลา วิธีปรับเรียบเอกซ์โพเนนเชียลหลายแบบในการพยากรณ์ราคาน้ำมันปาล์ม ราคาน้ำมันปาล์มดิบ และการผลิตน้ำมันปาล์มดิบในประเทศไทย ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมีนาคม พ.ศ. 2561 วิธีการที่ใช้ในการพยากรณ์ ได้แก่ Double Exponential Smoothing (DES) Multiplicative Holt-Winters (MHW) Additive Holt-Winters (AHW) Improved Additive Holt-Winters (IAHW) และวิธี Extended Additive Holt-Winters (EAHW) โดยได้ทำการเก็บข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2548 ถึงเดือนธันวาคม พ.ศ. 2560 รวมเป็นระยะเวลา 13 ปี ซึ่งรวบรวมจากฐานข้อมูลของสำนักงานเศรษฐกิจการเกษตรและกรมการค้าภายใน ในงานวิจัยฉบับนี้จะทำการคัดเลือกตัวแบบการพยากรณ์ที่มีความ

แม่นยำมากที่สุดโดยจะเลือกตัวแบบการพยากรณ์ที่มีค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำที่สุด

ผลการวิจัยแสดงให้เห็นว่าในการพยากรณ์ราคาน้ำมันปาล์มตัวแบบการพยากรณ์ DES EAHW AHW MHW และ IAHW ให้ค่า MAPE เท่ากับ 10.29% 10.43% 11.79% 11.82% และ 12.26% ตามลำดับ เนื่องจากค่า MAPE จากตัวแบบการพยากรณ์ DES และ EAHW ไม่มีความแตกต่างกันอย่างมีนัยสำคัญจึงสามารถสรุปได้ว่าตัวแบบการพยากรณ์ที่แม่นยำที่สุดที่ใช้พยากรณ์ราคาน้ำมันปาล์มคือ DES และ EAHW

ในส่วนของการพยากรณ์ราคาน้ำมันปาล์มดิบตัวแบบการพยากรณ์ DES EAHW AHW IAHW และ MHW ให้ค่า MAPE เท่ากับ 7.08% 7.14% 8.54% 8.75% และ 9.13% ตามลำดับ เนื่องจากค่า MAPE จากตัวแบบการพยากรณ์ DES และ EAHW ไม่มีความแตกต่างกันอย่างมีนัยสำคัญจึงสามารถสรุปได้ว่าตัวแบบการพยากรณ์ที่แม่นยำที่สุดที่ใช้พยากรณ์ราคาน้ำมันปาล์มดิบคือ DES และ EAHW

สุดท้ายในการพยากรณ์ปริมาณการผลิตน้ำมันปาล์มดิบตัวแบบการพยากรณ์ EAHW IAHW AHW MHW และ DES ให้ค่า MAPE เท่ากับ 10.88% 10.88% 11.07% 11.75% และ 14.19% ตามลำดับ เนื่องจากค่า MAPE จากตัวแบบการพยากรณ์ IAHW และ EAHW ไม่มีความแตกต่างกันอย่างมีนัยสำคัญจึงสามารถสรุปได้ว่าตัวแบบการพยากรณ์ที่แม่นยำที่สุดที่ใช้พยากรณ์ราคาน้ำมันปาล์มดิบคือ IAHW และ EAHW

2.2.3 งานวิจัยเกี่ยวกับน้ำมันดิบ

Ahmed and Shabri (2014) ได้ทำการวิจัยในการพยากรณ์ราคาน้ำมันดิบประเภท West Texas Intermediate (WTI) ซึ่งได้ทำการรวบรวมข้อมูลแบบรายวันจากหน่วยงาน Energy Information Administration (EIA) ในงานวิจัยนี้จะใช้ตัวแบบ Support Vector Regression (SVR) ตัวแบบ Autoregressive Integrated Moving Average (ARIMA) และตัวแบบ Generalized Autoregressive Conditional Heteroskedasticity (GARCH) ในส่วนของการวัดความแม่นยำของตัวแบบพยากรณ์ในงานวิจัยฉบับนี้จะใช้ค่า ค่ารากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) โดยผลลัพธ์ที่ได้พบว่าตัวแบบ SVM มีค่าความแม่นยำสูงที่สุดโดยมีค่า RMSE และ MAE เท่ากับ 0.8684 และ 0.6304 ตามลำดับ ในขณะที่ตัวแบบ ARIMA ให้ค่า RMSE และ MAE เท่ากับ 0.9856 และ 0.7204 ตามลำดับและตัวแบบ GARCH ให้ค่า RMSE และ MAE เท่ากับ 1.0134 และ 0.7392 ตามลำดับ

Chai (2018) ได้ทำการสร้างตัวแบบพยากรณ์แบบผสมในการพยากรณ์ราคาน้ำมันดิบ WTI จากการผสม 5 ตัวแบบ เริ่มจากตัวแบบ Product Partition Model-K-Means (PPM-KM) ที่ใช้ในการตรวจหาจุดเปลี่ยนของลำดับข้อมูลราคาน้ำมันดิบ ตัวแบบต่อมาที่ใช้ผสมด้วยคือตัวแบบ Time-varying Transition Probability Markov Regime Switching (TVTP-MRS) เพื่อทำการระบุการเปลี่ยนแปลงระดับข้อมูลราคาน้ำมันดิบ (Regime-switching Characteristic) จากนั้นใช้ตัวแบบ Bayesian Model Averaging (BMA) เพื่อการหาค่าพยากรณ์หลักในแต่ละ Regime และในตัวแบบสุดท้ายที่จะนำมาผสมคือตัวแบบ Time-varying Parameter Structure Time Series Model (TVP-STSM) มีหน้าที่ในการแยกองค์ประกอบของข้อมูล (Decompose) ตรวจสอบความสัมพันธ์ความแปรปรวนใน Regime และทำการพยากรณ์ราคาน้ำมันดิบ โดยจะนำตัวแบบผสมนี้มาทำการเปรียบเทียบความแม่นยำกับตัวแบบ ARIMA ด้วยค่า MAPE จากผลการพยากรณ์พบว่าตัวแบบผสมให้ค่า MAPE เท่ากับ 1.3451% ในขณะที่ตัวแบบ ARIMA ให้ค่า MAPE เท่ากับ 13.0162% จึงสามารถสรุปได้ว่าตัวแบบผสมมีความแม่นยำที่สูงกว่าตัวแบบพยากรณ์ทั่วไปมาก

Li (2019) พบว่าลักษณะข้อมูลของปริมาณการใช้พลังงานนั้นมีความผันผวนไม่แน่นอนสูงจึงได้ทำการสร้างตัวแบบพยากรณ์ Artificial Neural Network ตัวแบบ Linear Regression และตัวแบบพยากรณ์แบบผสมระหว่างตัวแบบ K-Means และ Artificial Neural Network เพื่อทำการพยากรณ์ปริมาณการใช้พลังงานคือก๊าซและน้ำมันซึ่งประกอบไปด้วย 2 ขั้นตอนหลักเริ่มจากการแบ่งกลุ่ม (Clustering) ด้วยวิธี K-Means เพื่อแยกชุดข้อมูลเป็นชุดข้อมูลย่อยได้เป็น 4 ชุดข้อมูลย่อย จากนั้นในขั้นตอนถัดมาจะทำการพยากรณ์ในแต่ละชุดข้อมูลย่อยด้วยตัวแบบ Artificial Neural Network หลังจากนั้นจะนำผลลัพธ์ที่ได้จากการพยากรณ์ในแต่ละชุดข้อมูลย่อยของแต่ละตัวแบบพยากรณ์มาเปรียบเทียบหาตัวแบบที่มีความแม่นยำสูงสุดของแต่ละชุดข้อมูลย่อยซึ่งจะประเมินจากค่า RMSE โดยผลลัพธ์ที่ได้พบว่าชุดข้อมูลที่ 1 มีค่า RMSE ของตัวแบบพยากรณ์แบบผสม ตัวแบบพยากรณ์ Artificial Neural Network ตัวแบบ Linear Regression เท่ากับ 3.62 3.84 และ 5.78 ตามลำดับ ชุดข้อมูลที่ 2 มีค่า RMSE ของตัวแบบพยากรณ์แบบผสม ตัวแบบพยากรณ์ Artificial Neural Network ตัวแบบ Linear Regression เท่ากับ 4.48 5.26 และ 5.09 ตามลำดับ ชุดข้อมูลที่ 3 มีค่า RMSE ของตัวแบบพยากรณ์แบบผสม ตัวแบบพยากรณ์ Artificial Neural Network ตัวแบบ Linear Regression เท่ากับ 6.88 7.14 และ 7.92 ตามลำดับ ชุดข้อมูลที่ 4 มีค่า RMSE ของตัวแบบพยากรณ์แบบผสม ตัวแบบพยากรณ์ Artificial Neural Network ตัวแบบ Linear Regression เท่ากับ 5.02 5.38 และ 6.27 ตามลำดับ ทำให้สามารถสรุปได้ว่าตัวแบบพยากรณ์แบบผสมให้ค่าความแม่นยำที่สูงที่สุด

Keerthan et al. (2019) ได้ใช้ตัวแบบการพยากรณ์การเรียนรู้ด้วยเครื่อง 5 ตัวแบบ ได้แก่ SVR (Linear, RBF, Polynomial) Random Forest Regression และ Linear Regression ในการทำงานวิจัยการพยากรณ์ราคาน้ำมันดิบ โดยมีสาเหตุมาจากการที่ราคาน้ำมันดิบที่มีมูลค่าเพิ่มขึ้นส่งผลให้ในประเทศอินเดียประสบปัญหาหาค่า LPG (Liquified Petroleum Gas) ปีโตรเลียม และน้ำมันดีเซลเพิ่มขึ้น โดยมีขั้นตอนดังต่อไปนี้ (1) รวบรวมข้อมูล ค้นหาปัจจัยที่เกี่ยวข้อง ทำความสะอาดข้อมูล (Data Cleaning) (2) คำนวณค่าสหสัมพันธ์ของปัจจัย (Correlation) (3) คำนวณค่า Bias และค่า Variance ของข้อมูล (4) สร้างตัวแบบพยากรณ์จากชุดข้อมูล Training ซึ่งในงานวิจัยฉบับนี้ จะแบ่งข้อมูลแบ่งออกเป็น 2 ส่วนได้แก่ ชุดข้อมูล Training ในสัดส่วน 80% และชุดข้อมูล Testing ในสัดส่วน 20% (5) คัดเลือกตัวแบบที่มีความแม่นยำสูงที่สุดจากการพยากรณ์แล้วเทียบผลลัพธ์กับข้อมูลชุด Testing ผลลัพธ์จากการพยากรณ์สามารถสรุปได้ว่าตัวแบบ Linear Regression ให้ความแม่นยำสูงที่สุดโดยให้ค่าเท่ากับ 85% ในขณะที่ตัวแบบ SVR แบบ Linear RBF และ Polynomial ให้ค่าความแม่นยำเท่ากับ 0% 20% และ 0% ตามลำดับ และตัวแบบ Random Forest มีความแม่นยำเท่ากับ 60%

Manickavasagam (2020) ได้ทำการวิจัยการพยากรณ์ราคาน้ำมันดิบแบบ Brent และ WTI ด้วยตัวแบบพยากรณ์แบบผสมระหว่างตัวแบบการเรียนรู้ด้วยเครื่องและตัวแบบ Nature Inspired Algorithms ดังต่อไปนี้ (1) ตัวแบบ MARSplines-IPSO-BPNN ในตัวแบบนี้จะประกอบไปด้วย 3 ส่วน ได้แก่ Multivariate Adaptive Regression Splines (MARSplines) ซึ่งเป็นส่วนที่ใช้ในการหาว่าปัจจัยใดที่มีผลกระทบต่อราคาน้ำมันดิบแล้วจะนำตัวแปรเหล่านั้นไปใช้ในขั้นตอนต่อมาคือ Improved Particle Swarm Optimization (IPSO) ที่ทำหน้าที่ในการหาค่าประมาณที่ดีที่สุดของพารามิเตอร์ที่นำไปในส่วนสุดท้ายคือ Backpropagation Neural Network (BPNN) จากนั้นจะได้ตัวแบบที่ใช้พยากรณ์ราคาน้ำมันดิบ (2) ตัวแบบ MARSplines-FPA-BPNN ในตัวแบบนี้ประกอบไปด้วย 3 ส่วน โดยจะต่างจากตัวแบบพยากรณ์แรกในส่วนที่ทำหน้าที่ในการหาค่าพารามิเตอร์ซึ่งในตัวแบบนี้จะใช้ Flower Pollination Algorithms (FPA) แทน IPSO นอกจากนั้นอีก 2 ส่วนประกอบที่เหลือจะเหมือนกับตัวแบบแรก โดยในการพยากรณ์ในงานวิจัยฉบับนี้ได้ทำการแบ่งข้อมูลออกเป็น 2 ช่วงคือ ชุดข้อมูล Training ในสัดส่วน 80% เพื่อใช้ในการสร้างตัวแบบพยากรณ์และชุดข้อมูล Testing ในสัดส่วน 20% เพื่อใช้ในการเปรียบเทียบความแม่นยำระหว่างตัวแบบซึ่งจะใช้ค่า MSE MAPE และ Theil U-statistic ในการเปรียบเทียบ โดยผลลัพธ์จากการพยากรณ์ทำให้สามารถสรุปได้ว่าตัวแบบ MARSplines-FPA-BPNN ให้ค่าความแม่นยำสูงที่สุด โดยแบ่งเป็น 2 กรณีได้แก่ (1) พยากรณ์ราคาน้ำมัน WTI ตัวแบบ MARSplines-FPA-BPNN ให้ค่า MSE MAPE และ Theil U เท่ากับ 0.000002

0.0808 และ 0.0019 ตามลำดับ และตัวแบบ MARSplines-IPSO-BPNN ให้ค่า MSE MAPE และ Theil U เท่ากับ 0.000016 0.2933 และ 0.0050 ตามลำดับ (2) พยากรณ์ราคาน้ำมันดิบ Brent ตัวแบบ MARSplines-FPA-BPNN ให้ค่า MSE MAPE และ Theil U เท่ากับ 0.000004 0.1722 และ 0.0031 ตามลำดับ และตัวแบบ MARSplines-IPSO-BPNN ให้ค่า MSE MAPE และ Theil U เท่ากับ 0.000015 0.3916 และ 0.0060 ตามลำดับ

Haque and Shaik (2021) ได้ทำการพยากรณ์ราคาน้ำมันดิบในช่วงเวลาตั้งแต่วันที่ 10 กุมภาพันธ์ ค.ศ. 2020 จนถึงวันที่ 27 เมษายน ค.ศ. 2020 โดยสาเหตุหลักในการทำวิจัยครั้งนี้ เนื่องจากสถานการณ์การแพร่ระบาดของโรคโควิด-19 ส่งผลให้ราคาน้ำมันดิบผันผวนสูงและราคาน้ำมันดิบส่งผลโดยตรงต่อเศรษฐกิจเช่น การผลิตและการขนส่ง ดังนั้นการมีตัวแบบพยากรณ์เพื่อใช้พยากรณ์ราคาน้ำมันดิบในช่วงนี้จึงเป็นสิ่งสำคัญ โดยในงานวิจัยฉบับนี้จะใช้ตัวแบบพยากรณ์ 2 ตัวแบบได้แก่ ตัวแบบ Autoregressive Integrated Moving Average (ARIMA) และตัวแบบ Generalized Autoregressive Conditional Heteroskedasticity (GARCH) และได้มีการรวบรวมข้อมูลราคาน้ำมันดิบประเภท WTI ตั้งแต่วันที่ 3 มกราคม ค.ศ. 1986 จนถึงวันที่ 27 เมษายน ค.ศ. 2020 โดยข้อมูลในช่วงตั้งแต่วันที่ 3 มกราคม ค.ศ. 1986 จนถึงวันที่ 10 กุมภาพันธ์ ค.ศ. 2020 จะถูกนำมาใช้ในการสร้างตัวแบบพยากรณ์ ในส่วนที่เหลือจะถูกนำมาใช้ในการเปรียบเทียบความแม่นยำระหว่างตัวแบบโดยจะประเมินความแม่นยำจากค่า ค่ารากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error: RMSE) และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) และจากผลการพยากรณ์พบว่าตัวแบบ ARIMA ให้ผลลัพธ์ที่ใกล้เคียงตัวแบบ GARCH โดยตัวแบบ ARIMA ให้ค่า RMSE เท่ากับ 10.18 และค่า MAE เท่ากับ 3.94 ในขณะที่ตัวแบบ GARCH ให้ค่า RMSE เท่ากับ 10.20 และค่า MAE เท่ากับ 3.95

2.2.4 งานวิจัยเกี่ยวกับยาง

Zahari (2018) ได้เห็นถึงความสำคัญในการพยากรณ์ราคายางธรรมชาติจากการที่ราคายางธรรมชาติมีความไม่แน่นอนซึ่งจะส่งผลต่อผู้ผลิต ผู้ทำการค้า ผู้บริโภค และผู้ที่เกี่ยวข้องในการผลิตยางธรรมชาติ ดังนั้นเพื่อช่วยในการตัดสินใจของแต่ละฝ่ายให้เป็นไปได้ง่ายขึ้นจึงได้จัดทำตัวแบบพยากรณ์ราคายางธรรมชาติ ในงานวิจัยฉบับนี้ได้ทำการรวบรวมข้อมูลราคายางธรรมชาติจาก Malaysian Rubber Board ตั้งแต่เดือนมกราคม ค.ศ. 2000 จนถึงเดือนธันวาคม ค.ศ. 2015 จะเป็นชุดข้อมูล Training Data Set และช่วงเดือน มกราคม ค.ศ. 2016 จนถึงเดือนธันวาคม ค.ศ. 2016 จะเป็นชุดข้อมูล Cross Validation Data Set และใช้ตัวแบบ ARIMA ในการพยากรณ์ โดยผลลัพธ์ที่ได้พบว่าตัวแบบที่เหมาะสมในการพยากรณ์ชุดข้อมูลชุดนี้คือตัวแบบ ARIMA (1,1,0)

Jong (2020) ได้ทำการสร้างตัวแบบผสมระหว่างตัวแบบ ARIMA กับตัวแบบ Support Vector Regression (SVR) ในการพยากรณ์ราคาขายทรงกรมชาติรายวัน แล้วจะนำไปเปรียบเทียบกับตัวแบบพยากรณ์ที่แยกเดี่ยวได้แก่ ตัวแบบ ARIMA และตัวแบบ SVR ซึ่งจากผลลัพธ์ที่ได้พบว่าตัวแบบผสมมีความแม่นยำสูงที่สุดโดยมีค่า MAE MFE MAPE และ RMSE เท่ากับ 2.2909 0.0631 0.3535 4.5556 และ 0.9976 ตามลำดับ ส่วนตัวแบบ SVR ให้ค่า MAE MFE MAPE และ RMSE เท่ากับ 3.1755 0.1662 0.489 3.9522 และ 0.9981 ตามลำดับ และตัวแบบ ARIMA ให้ค่า MAE MFE MAPE และ RMSE เท่ากับ 2.4918 0.1168 0.3831 3.2144 และ 0.1441 ตามลำดับ



บทที่ 3

ขั้นตอนการดำเนินงาน

ในงานวิจัยฉบับนี้จะทำการสร้างตัวแบบการพยากรณ์บนโปรแกรม R-Studio โดยใช้ภาษา R และบนแพลตฟอร์ม Google Colab โดยใช้ภาษา Python ในการพยากรณ์รวมถึงวัดค่า MAPE เพื่อเลือกตัวแบบพยากรณ์ที่มีความแม่นยำมากที่สุด โดยจะมีขั้นตอนดังต่อไปนี้

3.1 รวบรวมข้อมูลและกำหนดขอบเขตการวิจัย

งานวิจัยฉบับนี้ได้รวบรวมข้อมูลราคาเม็ดพลาสติก PP มาจากสถาบันพลาสติกซึ่งเป็นข้อมูลราคาเฉลี่ยจากผู้ผลิตแต่ละบริษัท และข้อมูลปัจจัยที่มีผลต่อราคาเม็ดพลาสติก PP โดยข้อมูลทั้ง 2 อย่างจะทำการเก็บรวบรวมมาตั้งแต่ เดือนมกราคม พ.ศ. 2554 ถึง เดือนธันวาคม พ.ศ. 2565 รวมเป็นระยะเวลา 12 ปี และจะทำการแบ่งข้อมูลออกเป็น 3 ช่วงได้แก่

1. ชุดข้อมูลฝึกสอน (Training Data Set) 7 ปี ตั้งแต่เดือนมกราคม พ.ศ. 2554 ถึงเดือนธันวาคม พ.ศ. 2560 โดยเป็นชุดข้อมูลที่ใช้ในการสร้างตัวแบบการพยากรณ์
2. ชุดข้อมูลตรวจสอบ (Cross Validation Data Set) 2 ปี 6 เดือน ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563 โดยเป็นชุดข้อมูลที่ใช้ในการทดสอบความแม่นยำและเปรียบเทียบความแม่นยำของแต่ละตัวแบบการพยากรณ์เพื่อเลือกตัวแบบการพยากรณ์ที่แม่นยำที่สุด
3. ชุดข้อมูลทดสอบ (Testing Data Set) 2 ปี 6 เดือน ตั้งแต่เดือนกรกฎาคม พ.ศ. 2563 ถึงเดือนธันวาคม พ.ศ. 2566 โดยเป็นชุดข้อมูลที่ใช้กับตัวแบบการพยากรณ์ที่แม่นยำที่สุดที่ได้จากชุดข้อมูล Cross Validation Data Set เพื่อนำมาวัดความผิดพลาดที่เป็นอิสระ (Unbiased Error) จากข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen Data)

และผู้วิจัยได้ทบทวนวรรณกรรมเพื่อศึกษาตัวแปรอิสระที่มีอิทธิพลต่อการบริโภคน้ำมันสำเร็จรูปจากวรรณกรรมดังแสดงในตารางที่ 3 ในส่วนของข้อมูลปัจจัยที่เกี่ยวข้องเมื่ออ้างอิงจากบทที่ 2 พบว่าจะมีการใช้ปัจจัยตามตารางที่ 3

ตารางที่ 3 ปัจจัยที่ถูกนำมาพิจารณาในแต่ละงานวิจัยจากบทที่ 2

ผู้แต่ง	ปี	ข้อมูลที่ใช้การพยากรณ์	ปัจจัยที่นำมาพิจารณา
วรรณิภา และคณะ	2552	ราคาเม็ดพลาสติก พอลิเอทิลีน	ระดับสินค้าคงคลังทางธุรกิจ ยอดค้าปลีกที่ไม่รวมยานยนต์ ยอดขายบ้านใหม่ ราคาน้ำมันดิบ จำนวนการขออนุญาตสร้างบ้านใหม่ ดัชนี ความเชื่อมั่นผู้บริโภค ดัชนีราคาผู้บริโภคและผู้ผลิต ดัชนี S&P500 อัตราการว่างงาน ปริมาณการผลิตภาคอุตสาหกรรม ปริมาณการสั่งซื้อสินค้าคงทน ดัชนี S&P GSCI Agriculture Index
ณัฐกมล	2560	ยอดขายพลาสติกพีวีซี	อัตราดอกเบี้ย อัตราว่างงาน ดัชนีราคา ผู้บริโภค ราคาน้ำมัน ดัชนีมวลรวม ประชาชาติ อัตราการบริโภคภายใน ครัวเรือน อัตราการผลิตไฮสเตรนซ์ อัตราการนำเข้า และอัตราการส่งออก
Adisak and Vichai	2561	ราคาของเม็ดพลาสติก พอลิไวนิลคลอไรด์	ราคาน้ำมันดิบ อัตราแลกเปลี่ยน

นอกจากปัจจัยที่อ้างอิงจากบทที่ 2 ทางผู้จัดทำได้เพิ่มปัจจัยที่เกี่ยวข้องกับราคาเม็ดพลาสติก PP จากการรวบรวมข้อมูลจากหน่วยงานภาครัฐ ได้แก่ (1.) ดัชนีการผลิตเม็ดพลาสติก PP เพราะเป็นการสื่อถึงปริมาณการผลิตเม็ดพลาสติกซึ่งจะส่งผลต่อราคาเม็ดพลาสติก (2.) สถิติการส่งออกและการนำเข้าของโพลิโพรพิลีนชั้นปฐมภูมิเพราะมีผลโดยตรงต่อการส่งผลิตเม็ดพลาสติกทำให้มีผลต่อราคาด้วยเช่นกัน (3.) สถิติการผลิตรถยนต์และรถจักรยานยนต์ เพราะรถยนต์และรถจักรยานยนต์มีการใช้พลาสติกเป็นชิ้นส่วนจำนวนมากทำให้จำนวนรถยนต์และรถจักรยานยนต์ที่ถูกผลิตมีผลโดยตรงต่อราคาเม็ดพลาสติก ดังนั้นเมื่อทำการคัดเลือกปัจจัยจากงานวิจัยในบทที่ 2 มารวมกับปัจจัยที่เพิ่มเข้ามาจะได้ผลลัพธ์ดังตารางที่ 4

ตารางที่ 4 ข้อมูลราคาและปัจจัยที่เกี่ยวข้องกับราคาเม็ดพลาสติก PP

ปัจจัย	ความถี่	หน่วย
ราคาเม็ดพลาสติก PP	รายเดือน	บาท/ก.ก.
ดัชนีการผลิตเม็ดพลาสติก PP		-
สถิติการส่งออกของโพลิโพรพิลีนชั้นปฐมภูมิ		ล้านตัน
สถิติการนำเข้าของโพลิโพรพิลีนชั้นปฐมภูมิ		ล้านตัน
ราคาน้ำมันดิบ Brent		ดอลลาร์สหรัฐ
ราคาน้ำมันดิบ WTI		ดอลลาร์สหรัฐ
ดัชนีราคาผู้บริโภค(CPI)		-
ดัชนีราคาผู้ผลิต(PPI)		-
สถิติการผลิตรถยนต์		พันคัน
สถิติการผลิตรถจักรยานยนต์		พันคัน

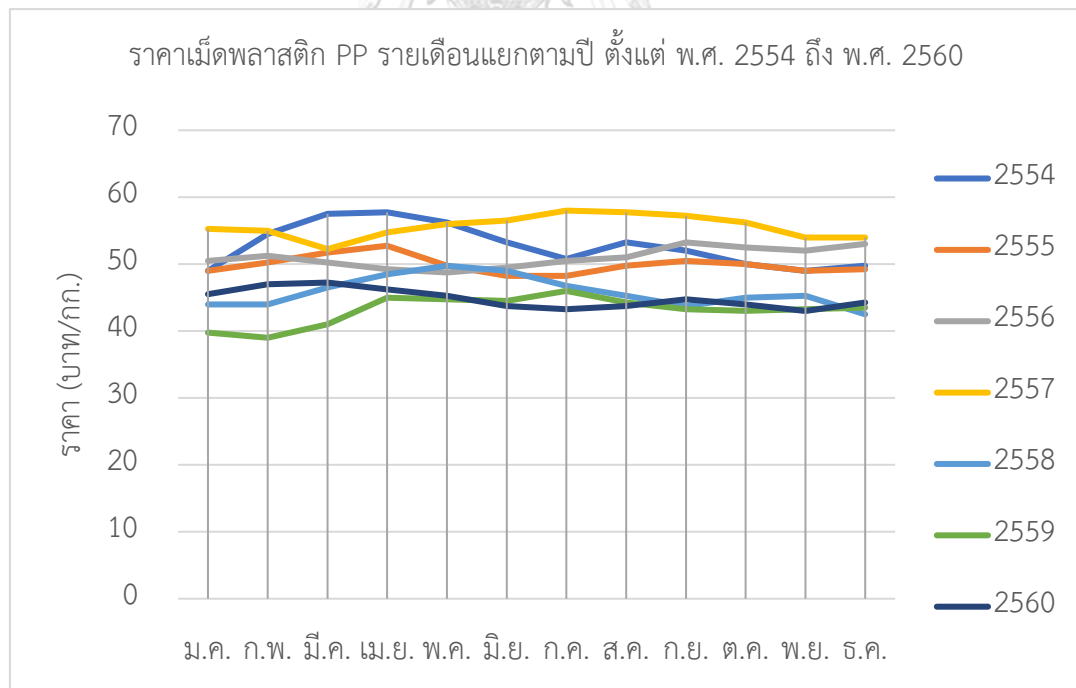
สุดท้ายจะเป็นการนำข้อมูลที่รวบรวมมาเข้าสู่กระบวนการทำความสะอาดข้อมูล (Data Cleaning) ได้แก่ ตรวจสอบค่าผิดปกติ (Outlier) ตรวจสอบความสม่ำเสมอของระยะห่างระหว่างข้อมูลแต่ละตัว ตรวจสอบข้อมูลที่ขาดหายไป (Missing Data) ว่าเกิดขึ้นในชุดข้อมูลหรือไม่ เป็นต้น เพื่อให้ข้อมูลสามารถนำไปใช้งานในการวิเคราะห์และการพยากรณ์ได้

3.2 วิเคราะห์ข้อมูล

ในขั้นตอนนี้จะทำการนำข้อมูลราคาเม็ดพลาสติก PP ในส่วนของชุดข้อมูลฝึกสอนมาพล็อตกราฟเพื่อดูรูปแบบลักษณะของข้อมูลจะพบว่าข้อมูลมีแนวโน้มไปในทิศทางที่ลดลงดังรูปที่ 27 และพล็อตกราฟรายปีเพื่อพิจารณาลักษณะทางฤดูกาล (Seasonal) จะพบว่าข้อมูลมีรูปแบบการเคลื่อนที่คล้ายคลึงกันในแต่ละปีจึงมีข้อสังเกตว่าข้อมูลอาจมีลักษณะฤดูกาลดังรูปที่ 28

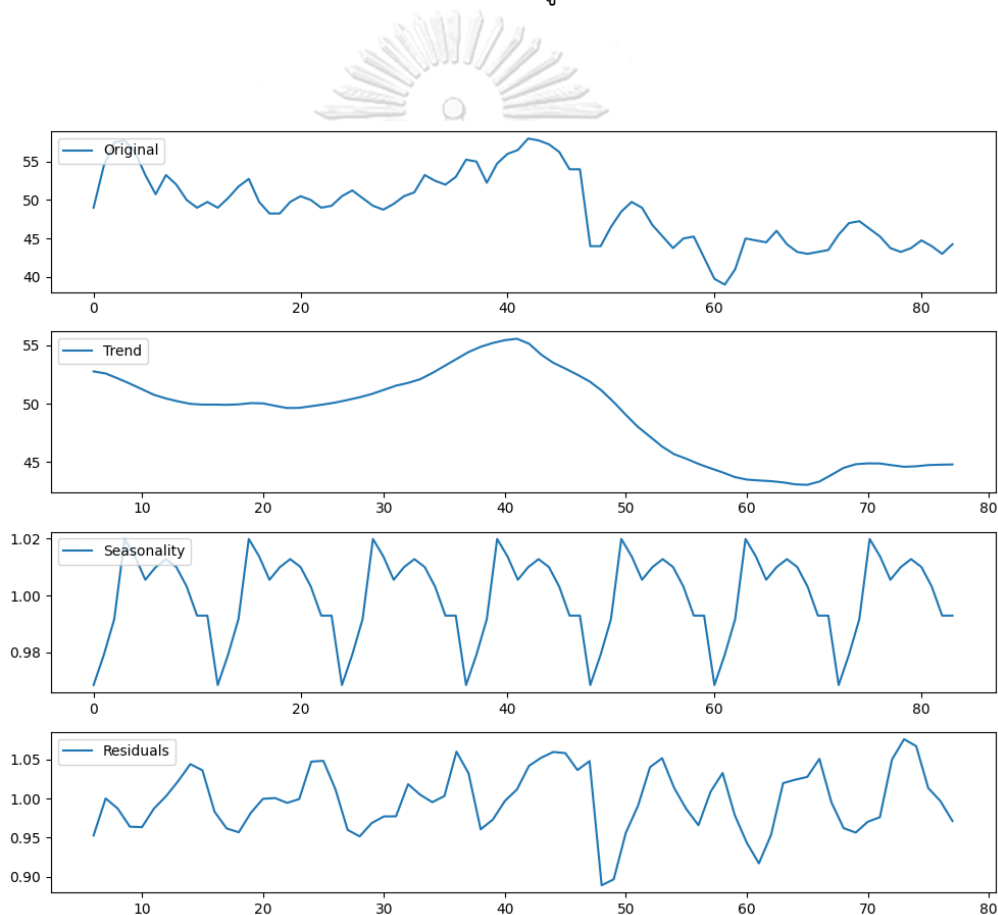


รูปที่ 27 ราคาเม็ดพลาสติก PP ในประเทศไทยตั้งแต่เดือนมกราคม พ.ศ. 2554
ถึง เดือนธันวาคม พ.ศ. 2560 (ที่มา : สถาบันพลาสติก)



รูปที่ 28 ราคาเม็ดพลาสติก PP รายเดือนแยกตามปี ตั้งแต่ พ.ศ. 2554 ถึง พ.ศ. 2560

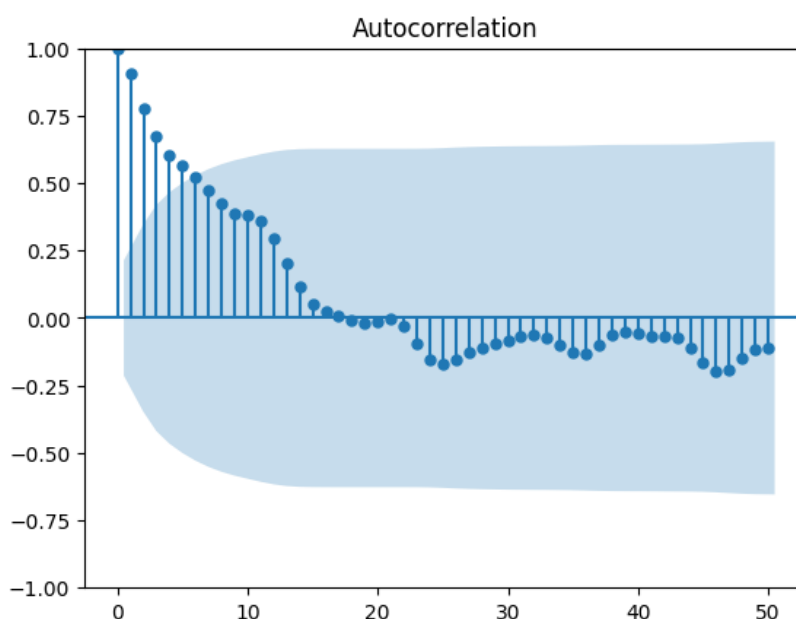
ลำดับถัดมาจะเป็นการพิสูจน์ลักษณะของข้อมูลซึ่งจะประกอบไปด้วย 4 ส่วนหลักได้แก่ แนวโน้ม (Trend) ฤดูกาล (Seasonal) วัฏจักร (Cyclical) และความผิดปกติ (Irregular) โดยนำข้อมูลราคาเม็ดพลาสติก PP มาทำการวิเคราะห์ข้อมูลแบบแยกองค์ประกอบ (Decomposition Method) ในรูปแบบคูณ (Multiplicative) โดยจะใช้ภาษา Python ผ่าน Google Colab ด้วยคำสั่ง `seasonal_decompose` Library `statsmodels.tsa.seasonal` เพื่อแยกองค์ประกอบข้อมูลแล้วนำลักษณะข้อมูลที่ได้มาเลือกตัวแบบการพยากรณ์ดังรูปที่ 29 ซึ่งสามารถสรุปได้ว่าข้อมูลราคาเม็ดพลาสติก PP มีลักษณะเป็นฤดูกาลและมีแนวโน้ม โดยดูได้จากในส่วนของ Seasonal ในรูปที่มีลักษณะวนซ้ำกลับมาที่เดิม และในส่วนของ Trend ในรูปพบว่ามีแนวโน้มในทางลดลง



รูปที่ 29 โครงสร้างองค์ประกอบข้อมูลราคาเม็ดพลาสติก PP

สุดท้ายจะเป็นการตรวจสอบรูปแบบของอนุกรมเวลาข้อมูลด้วยการใช้ฟังก์ชันอัตสหสัมพันธ์ (Autocorrelation Function, ACF) โดยจะใช้ Library `statsmodels.graphics.tsaplots` คำสั่ง `plot_acf` ในการดำเนินงาน

จากการตรวจสอบรูปแบบอนุกรมเวลาของข้อมูลราคาเม็ดพลาสติก PP โดยใช้ ACF ดังแสดงในรูปที่ 30 พบว่าการลดลงของค่า ACF อย่างช้า ๆ ในขณะที่ Lag เพิ่มขึ้นนั้นแสดงให้ว่าข้อมูลมีรูปแบบแนวโน้มปรากฏอยู่ และการที่ ACF มีลักษณะการเคลื่อนไหวเป็นรูปเส้นโค้งเปลือกหอย (Scalloped Shape) ปรากฏอยู่ แสดงว่าข้อมูลมีรูปแบบฤดูกาลปรากฏอยู่ด้วย (Hyndman และ Athanasopoulos, 2018)



รูปที่ 30 กราฟฟังก์ชันอัตโนมัติสหสัมพันธ์ (ACF) ของราคาเม็ดพลาสติก PP

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

3.3 สร้างตัวแบบการพยากรณ์จากชุดข้อมูลฝึกสอน

ในขั้นตอนนี้จะเป็นการสร้างตัวแบบการพยากรณ์ราคาเม็ดพลาสติก PP จากชุดข้อมูลฝึกสอน ซึ่งวิธีการในการพยากรณ์แบ่งออกเป็น 3 แบบได้แก่ วิธีการพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data Method) วิธีการพยากรณ์ข้อมูลเชิงสาเหตุ (Casual Method) และวิธีการพยากรณ์ด้วยการเรียนรู้ด้วยเครื่อง (Machine Learning Method)

3.3.1 วิธีการพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data Method)

การพยากรณ์ข้อมูลอนุกรมเวลาจะเป็นการนำข้อมูลในอดีตมาพยากรณ์ข้อมูลในอนาคต ซึ่งจากการพิจารณาลักษณะข้อมูลพบว่าข้อมูลราคาเม็ดพลาสติก PP รายเดือนในประเทศไทยมีรูปแบบเป็นแนวโน้มและฤดูกาล โดยในงานวิจัยนี้จะใช้เลือกใช้ตัวแบบการปรับเรียบเอกซ์โพเนนเชียลแบบโฮลต์และวินเทอร์ (Holt-Winters Exponential Smoothing) หรือตัวแบบ Holt-Winters และตัว

แบบบอซ-เจนกินส์แบบมีฤดูกาล (Box-Jenkins: Seasonal Autoregressive Integrated Moving Average, SARIMA) หรือ SARIMA ซึ่งจะใช้แพลตฟอร์ม Google Colab ด้วยภาษา Python เป็นเครื่องมือในการสร้างตัวแบบพยากรณ์

3.3.1.1 วิธีปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์ (Holt – Winters Exponential Smoothing Method)

ชุดข้อมูลฝึกสอนของราคาเม็ดพลาสติก PP รายเดือนในประเทศไทยมีลักษณะรูปแบบแนวโน้มในทิศทางลดลง ดังนั้นจึงเลือกใช้วิธีปรับเรียบเอกซ์โพเนนเชียลแบบโฮลท์และวินเทอร์แบบคูณ (Multiplicative Holt – Winters Exponential Smoothing Method) โดยขั้นตอนดำเนินการได้แก่

1. ทำการสร้างตัวแบบพยากรณ์จากชุดข้อมูลฝึกสอน จากภาษา Python บน Google Colab ด้วย Library statsmodels.tsa.holtwinters คำสั่ง ExponentialSmoothing โดยค่าพารามิเตอร์สำหรับตัวแบบพยากรณ์ที่ต้องทำการปรับเพื่อที่จะให้ได้ผลลัพธ์การพยากรณ์ที่เหมาะสม ได้แก่ ค่าคงที่ปรับเรียบสำหรับค่าระดับ (α) ค่าคงที่ปรับเรียบสำหรับค่าแนวโน้ม (β) และค่าคงที่ปรับเรียบสำหรับค่าฤดูกาล (γ)
2. ทำการพยากรณ์จากตัวแบบในข้อ 1. ที่เป็นระยะเวลาล่วงหน้า 30 เดือนให้เท่ากับจำนวนข้อมูลของชุดข้อมูลตรวจสอบ
3. นำผลที่ได้จากการพยากรณ์มาเทียบกับชุดข้อมูลตรวจสอบ เพื่อนำไปใช้ในการหา MAPE ของตัวแบบ
4. พิจารณาว่าเกิดการ Overfit หรือ Underfit หรือไม่
5. นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด

3.3.1.2 วิธี Seasonal Autoregressive Integrated Moving Average

ชุดข้อมูลฝึกสอนของราคาเม็ดพลาสติก PP รายเดือนในประเทศไทยมาสร้างตัวแบบการพยากรณ์ SARIMA โดยมีขั้นตอนดังนี้

1. ทำการสร้างตัวแบบพยากรณ์จากชุดข้อมูลฝึกสอน จากภาษา Python บน Google Colab ด้วย Library statsmodels.tsa.statespace ในการสร้างตัวแบบพยากรณ์ โดยจะพิจารณาตัวแบบที่เหมาะสมสำหรับข้อมูลด้วยค่า Corrected for Akaike's Information Criterion (AICc) ที่ต่ำที่สุด

2. วิเคราะห์ตรวจสอบค่าความคลาดเคลื่อนของตัวแบบพยากรณ์ (Residual) เช่น ค่า ACF PACF ของค่าความคลาดเคลื่อนจากตัวแบบพยากรณ์จะต้องไม่มีรูปแบบแนวโน้มและฤดูกาล และลักษณะการกระจายตัวของข้อมูลค่าความคลาดเคลื่อนต้องมีการแจกแจงแบบปกติ (Normal Distribution)
3. ทำการพยากรณ์เป็นระยะเวลาล่วงหน้า 30 เดือนให้เท่ากับจำนวนข้อมูลของชุดข้อมูลตรวจสอบ
4. นำผลที่ได้จากการพยากรณ์มาเทียบกับชุดข้อมูลตรวจสอบเพื่อนำไปใช้ในการหา MAPE ของตัวแบบ
5. พิจารณาว่าเกิดการ Overfit หรือ Underfit หรือไม่
6. นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด

3.3.2 วิธีการพยากรณ์ข้อมูลเชิงสาเหตุ (Casual Method)

ตัวแบบการพยากรณ์ที่ใช้สำหรับการพยากรณ์เชิงสาเหตุในงานวิจัยฉบับนี้คือตัวแบบการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Model) โดยในงานวิจัยฉบับนี้ได้ทำการรวบรวมข้อมูลปัจจัยต่างๆที่คาดว่าจะส่งผลกับราคาเม็ดพลาสติก PP โดยได้ทำการรวบรวมข้อมูลปัจจัยตั้งแต่เดือน มกราคม พ.ศ. 2554 จนถึงเดือน ธันวาคม พ.ศ. 2566 แล้วทำการแบ่งเป็น 3 ช่วง เช่นเดียวกับการแบ่งช่วงข้อมูลราคาเม็ดพลาสติก PP โดยขั้นตอนการพยากรณ์ด้วยวิธีถดถอยพหุคูณมีดังนี้

1. รวบรวมปัจจัยที่เกี่ยวข้องกับราคาเม็ดพลาสติก PP โดยศึกษาจากงานวิจัยที่เกี่ยวข้อง และข้อมูลเกี่ยวกับผลิตภัณฑ์สุดท้ายของเม็ดพลาสติก PP จะได้ปัจจัยดังเช่น ดัชนีการผลิตเม็ดพลาสติก PP สถิติการส่งออกและการนำเข้าของโพลิโพรพิลีนขั้นปฐมภูมิ ราคาน้ำมันดิบ Brent และ WTI ดัชนีราคาผู้บริโภค (CPI) ดัชนีราคาผู้ผลิต (PPI) และ สถิติการผลิตรถยนต์และรถจักรยานยนต์ รวมเป็น 9 ปัจจัย
2. นำข้อมูลแต่ละปัจจัยในช่วง Training Data Set มาทำการสร้างสมการถดถอยด้วยวิธีการถดถอยแบบขั้นตอน (Stepwise Regression) กำหนด Stepwise Selection Term (α) = 0.05 เพื่อหาปัจจัยที่ไม่เกี่ยวข้องและทำการทดสอบค่า p-value โดยแต่ละปัจจัยต้องมีค่า p-value ที่ต่ำกว่า 0.05 ถึงจะถือว่าปัจจัยนั้นมีผลต่อราคาเม็ดพลาสติก PP

3. หาค่า VIF (Variance Inflation Factor) ของแต่ละปัจจัยหากปัจจัยใดมีค่าเกิน 10 ให้ตัดทิ้งเพราะเป็นการบ่งชี้ว่าปัจจัยนั้นส่งผลต่อปัจจัยอื่น เนื่องจากต้องการให้ปัจจัยแต่ละตัวเป็นอิสระต่อกัน
4. ทำการทดสอบบนโปรแกรม Minitab เวอร์ชัน 18 จะได้ค่าสัมประสิทธิ์ของแต่ละปัจจัย และค่า R-Squared adjust ของตัวแบบถดถอยพหุคูณ และทดสอบสมมติฐานของการถดถอยเชิงเส้นโดยพิจารณาจากค่าความคลาดเคลื่อน (ประชากรค่าความผิดพลาดมีค่าเฉลี่ยเท่ากับศูนย์ มีความแปรปรวนคงที่ มีการแจกแจงแบบปกติ และมีอิสระต่อกัน)
5. เมื่อได้ตัวแบบถดถอยพหุคูณที่มีค่า Adjusted R-Squared ที่เหมาะสม ให้ทำการพยากรณ์ล่วงหน้าเป็นระยะเวลา 30 เดือน เพื่อที่จะสามารถนำไปเปรียบเทียบกับชุดข้อมูลตรวจสอบ
6. นำผลที่ได้จากการพยากรณ์มาเทียบกับชุดข้อมูลตรวจสอบเพื่อนำไปใช้ในการหา MAPE ของตัวแบบ
7. พิจารณาว่าเกิดการ Overfit หรือ Underfit หรือไม่
8. นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด

3.3.3 วิธีพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง

งานวิจัยฉบับนี้จะทำการสร้างตัวแบบการพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่องโดยใช้ตัวแบบดังนี้ ตัวแบบ Support Vector Regression ตัวแบบ Decision Tree ตัวแบบ Random Forest ตัวแบบ XGBoost ตัวแบบ Artificial Neural Network และตัวแบบ Hybrid

3.3.3.1 การปรับไฮเปอร์พารามิเตอร์ของตัวแบบการเรียนรู้ของเครื่อง

เทคนิคการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบการเรียนรู้ของเครื่องที่ใช้ในงานวิจัยฉบับนี้คือ Grid Search CV โดยใช้ Python Library sklearn.model_selection คำสั่ง GridSearchCV การทำ Grid Search CV จะเริ่มจากการกำหนดค่าไฮเปอร์พารามิเตอร์ที่ต้องการจะทดลองในการสร้างตัวแบบการเรียนรู้ของเครื่องแต่ละตัวแบบ จากนั้นทำการดำเนินการพยากรณ์และประเมินผลลัพธ์ของแต่ละตัวแบบพยากรณ์ทุก Combination ของแต่ละค่าไฮเปอร์พารามิเตอร์ โดยใช้การแบ่งชุดข้อมูลตรวจสอบออกเป็นหลายชุด (Cross Validation) หลังจากที่ทำทำการประเมินครบทุก Combination ผลลัพธ์ตัวแบบสุดท้ายจะเลือกจากตัวแบบที่มีค่าความคลาดเคลื่อน หรือ MAPE ต่ำที่สุด

ข้อมูลตัวแปรอิสระที่จะนำมาใช้ในงานวิจัยฉบับนี้จะเลือกใช้ตัวแปรอิสระทั้งหมดในการสร้างตัวแบบพยากรณ์ดังตารางที่ 3-2 โดยตัวแปรอิสระมีลักษณะ หน่วย รวมถึงขนาดที่แตกต่าง

กันจึงต้องทำการปรับสเกลของแต่ละตัวแปรอิสระด้วยการทำ Feature Scaling ด้วยการใช้ Python Library sklearn.preprocessing คำสั่ง StandardScaler แต่กลุ่มตัวแบบ Ensemble Learning เช่น ตัวแบบ Decision Tree ตัวแบบ Random Forest ตัวแบบ XGBoost จะไม่ได้ทำ Feature Scaling

3.3.3.2 วิธีพยากรณ์ด้วยตัวแบบ Support Vector Regression (SVR)

ตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่องตัวแรกที่จะนำมาใช้คือตัวแบบ SVR โดยการพยากรณ์ในหัวข้อ 3.3.3 ถึง 3.3.5 จะดำเนินการด้วยภาษา python บน Google Colab และจะมีการนำเข้า Library ที่จำเป็นในการสร้างตัวแบบ เช่น pandas numpy matplotlib.pyplot sklearn.svm sklearn.model_selection sklearn.metrics เป็นต้น โดยในส่วนของ SVR จะมีขั้นตอนการสร้างตัวแบบดังนี้

1. กำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ SVR ในการทดสอบใน Grid Search CV

ตารางที่ 5 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ SVR

ไฮเปอร์พารามิเตอร์	ค่าที่ใช้ทดสอบใน Grid Search CV
Kernel	'linear', 'poly', 'rbf', 'sigmoid'
C	0.01, 0.1, 1, 10, 15, 20, 30, 100, 1000, 3000, 5000
Gamma	'scale', 'auto', 0.001, 0.01, 0.1, 1
Degree	1, 2, 3, 4

- kernel คือ ชนิดของ kernel ที่จะใช้ในตัวแบบพยากรณ์ SVR
 - C คือ Regularization parameter
 - gamma คือ ค่าสัมประสิทธิ์ใน Polynomial, RBF และ Sigmoid kernel
 - degree คือ ระดับ Degree ใน Polynomial kernel
2. สรุปค่าไฮเปอร์พารามิเตอร์ที่ให้ผลลัพธ์ MAPE ที่ดีที่สุดจากข้อ 1. แล้วทำการพยากรณ์ เป็นระยะเวลาล่วงหน้า 30 เดือนให้เท่ากับจำนวนข้อมูลของชุดข้อมูลตรวจสอบ
 3. นำผลที่ได้จากการพยากรณ์มาเทียบกับชุดข้อมูลตรวจสอบ เพื่อนำไปใช้ในการหา MAPE ของตัวแบบ

- นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด

3.3.3.3 วิธีพยากรณ์ด้วยตัวแบบต้นไม้การตัดสินใจ (Decision Tree)

ตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่อง Decision Tree จะทำการพยากรณ์ในหัวข้อ 3.3.3 ถึง 3.3.5 โดยจะดำเนินการด้วยภาษา python บน Google Colab และจะมีการนำเข้า Library ที่จำเป็นในการสร้างตัวแบบ เช่น pandas numpy matplotlib.pyplot sklearn.tree sklearn.metrics sklearn.model_selection เป็นต้น โดยในส่วนของ Decision Tree จะมีขั้นตอนการสร้างตัวแบบดังนี้

- กำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ Decision Tree ในการทดสอบใน Grid Search CV

ตารางที่ 6 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ Decision Tree

ไฮเปอร์พารามิเตอร์	ค่าที่ใช้ทดสอบใน Grid Search CV
Max Depth	2, 4, 6, 8, 10, 12, 14, 16, 18, 20
Min Samples Split	2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50
Min Samples Leaf	1, 2, 4, 6, 8, 10, 12, 16, 18, 20
Max Features	None, 'sqrt', 'log2'
Splitter	'best', 'random'

- Max Depth คือ ความลึกสูงสุดของต้นไม้การตัดสินใจ กำหนดจำนวนชั้นที่ต้นไม้สามารถสร้างขึ้นได้ ค่าที่สูงขึ้นอาจทำให้ต้นไม้มีความซับซ้อนมากขึ้น ซึ่งอาจทำให้เกิดการ overfitting
- Min Samples Split คือ จำนวนตัวอย่างขั้นต่ำที่จำเป็นต้องใช้ในการแบ่งโหนด (node splitting) ของต้นไม้การตัดสินใจ ค่าที่สูงขึ้นจะทำให้ต้นไม้มีความซับซ้อนลดลงและอาจทำให้เกิดการเรียนรู้ไม่เพียงพอจากชุดข้อมูลฝึกสอน
- Min Samples Leaf คือ จำนวนตัวอย่างขั้นต่ำที่ต้องมีในใบ (leaf) ของต้นไม้การตัดสินใจ ค่าที่สูงขึ้นอาจทำให้ต้นไม้มีความซับซ้อนลดลงและอาจทำให้เกิดการเรียนรู้ไม่เพียงพอจากชุดข้อมูลฝึกสอน

- Max Features คือ การกำหนดจำนวนคุณสมบัติที่ใช้สุ่มที่ถูกนำมาพิจารณาในแต่ละโหนด ในกรณีที่ max features เป็น None ทุกคุณสมบัติจะถูกพิจารณา
 - Splitter คือ พารามิเตอร์ที่กำหนดวิธีการแบ่งแยกโหนดในต้นไม้การตัดสินใจ มีค่าให้เลือกอยู่สองแบบคือ "best" และ "random" การเลือก "best" จะทำการค้นหาค่าที่ดีที่สุดสำหรับการแบ่งโหนดในแต่ละชั้น ในขณะที่การเลือก "random" จะทำการสุ่มค่าในการแบ่งโหนด
2. สรุปลค่าไฮเปอร์พารามิเตอร์ที่ให้ผลลัพธ์ MAPE ที่ดีที่สุดจากข้อ 1. แล้วทำการพยากรณ์ เป็นระยะเวลาล่วงหน้า 30 เดือนให้เท่ากับจำนวนข้อมูลของชุดข้อมูลตรวจสอบ
 3. นำผลที่ได้จากการพยากรณ์มาเทียบกับชุดข้อมูลตรวจสอบ เพื่อนำไปใช้ในการหา MAPE ของตัวแบบ
 4. นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด

3.3.3.4 วิธีพยากรณ์ด้วยตัวแบบป่าสุ่ม (Random Forest)

ตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่อง Random Forest จะทำการการพยากรณ์ในหัวข้อ 3.3.3 ถึง 3.3.5 โดยจะดำเนินการด้วยภาษา python บน Google Colab และจะมีการนำเข้า Library ที่จำเป็นในการสร้างตัวแบบ เช่น pandas numpy matplotlib.pyplot sklearn.ensemble sklearn.model_selection sklearn.metrics เป็นต้น โดยในส่วนของ Random Forest จะมีขั้นตอนการสร้างตัวแบบดังนี้

1. กำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ Random Forest ในการทดสอบใน Grid Search CV

ตารางที่ 7 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ Random Forest

ไฮเปอร์พารามิเตอร์	ค่าที่ใช้ทดสอบใน Grid Search CV
N_estimator	10, 20, 30, 40, 50, 100, 200, 250, 300, 350
Max Depth	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15
Min Samples Split	2, 3, 4, 5, 6, 7, 8, 9
Max Features	'sqrt', 'log2'

- $N_{estimator}$ คือ จำนวนต้นไม้ตัดสินใจ (Decision Tree) ในป่าสุ่ม (Random Forest) หรือจำนวนตัวอย่างที่ถูกสุ่มใช้ในการสร้างต้นไม้แต่ละต้น ค่าที่สูงขึ้นจะทำให้ Random Forest มีความซับซ้อนมากขึ้น แต่ก็อาจทำให้กระบวนการฝึกสอนช้าลง และเสี่ยงในการเกิด overfitting
 - Max Depth คือ ความลึกสูงสุดของต้นไม้ตัดสินใจ (Decision Tree) แต่ละต้นในป่าสุ่ม (Random Forest) ค่าที่สูงขึ้นจะทำให้ต้นไม้มีความซับซ้อนและมีความสามารถในการเรียนรู้รายละเอียดมากขึ้น แต่อาจทำให้เกิด overfitting
 - Min Samples Split คือ จำนวนตัวอย่างขั้นต่ำที่จำเป็นต้องใช้ในการแบ่งโหนด (node splitting) ของต้นไม้การตัดสินใจ หากจำนวนตัวอย่างในโหนดหนึ่งต่ำกว่า Min Samples Split แล้วต้นไม้จะไม่แยกส่วนเพิ่มต่อไป ค่า Min Samples Split ที่สูงขึ้นจะช่วยลดความซับซ้อนของต้นไม้และลดความเสี่ยงในการเกิด overfitting แต่ก็อาจทำให้มีความไม่แม่นยำลดลง
 - Max Features คือ การกำหนดจำนวนคุณสมบัติที่ใช้สุ่มที่ถูกนำมาพิจารณาในแต่ละโหนด ในกรณีที่ max features เป็น None ทุกคุณสมบัติจะถูกพิจารณา
2. สรุปค่าไฮเปอร์พารามิเตอร์ที่ให้ผลลัพธ์ MAPE ที่ดีที่สุดจากข้อ 1. แล้วทำการพยากรณ์ เป็นระยะเวลาล่วงหน้า 30 เดือนให้เท่ากับจำนวนข้อมูลของชุดข้อมูลตรวจสอบ
 3. นำผลที่ได้จากการพยากรณ์มาเทียบกับชุดข้อมูลตรวจสอบเพื่อนำไปใช้ในการหา MAPE ของตัวแบบ
 4. นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด

3.3.3.5 วิธีพยากรณ์ด้วยตัวแบบ XGBoost

ตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่อง XGBoost จะทำการการพยากรณ์ในหัวข้อ 3.3.3 ถึง 3.3.5 โดยจะดำเนินการด้วยภาษา python บน Google Colab และจะมีการนำเข้า Library ที่จำเป็นในการสร้างตัวแบบ เช่น pandas numpy matplotlib.pyplot xgboost sklearn.metrics sklearn.model_selection เป็นต้น โดยในส่วนของ XGBoost จะมีขั้นตอนการสร้างตัวแบบดังนี้

1. กำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ XGBoost ในการทดสอบใน Grid Search CV

ตารางที่ 8 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ XGBoost

ไฮเปอร์พารามิเตอร์	ค่าที่ใช้ทดสอบใน Grid Search CV
N_estimator	400, 500, 600, 700, 800
Max Depth	7, 8, 9, 10, 11, 13
Learning Rate	0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8
Subsample	0.1, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1
Gamma	0.7, 0.8, 0.9

- N_estimator คือ จำนวนต้นไม้ตัดสินใจ (Decision Tree) ใน XGBoost หรือจำนวนตัวอย่างที่ถูกสุ่มใช้ในการสร้างต้นไม้แต่ละต้น ค่าที่สูงขึ้นจะทำให้ XGBoost มีความซับซ้อนมากขึ้น แต่ก็อาจทำให้กระบวนการฝึกสอนช้าลง และเสี่ยงในการเกิด overfitting
 - Max Depth คือ ความลึกสูงสุดของต้นไม้ตัดสินใจ (Decision Tree) แต่ละต้นใน XGBoost ค่าที่สูงขึ้นจะทำให้ต้นไม้มีความซับซ้อนและมีความสามารถในการเรียนรู้รายละเอียดมากขึ้น แต่อาจทำให้เกิด overfitting
 - Learning Rate คือ พารามิเตอร์ที่กำหนดอัตราการเรียนรู้ของตัวแบบพยากรณ์ XGBoost ค่าที่ต่ำทำให้ใช้เวลาในการฝึกฝนนานขึ้น แต่ช่วยลดการเกิด overfitting
 - Subsample คือ สัดส่วนของตัวอย่างที่ถูกเลือกในแต่ละรอบฝึกสอน ค่าที่น้อยจะช่วยลดความซับซ้อนและให้กับตัวแบบพยากรณ์ XGBoost
 - Gamma คือ Regularization Hyperparameter
2. สรุปค่าไฮเปอร์พารามิเตอร์ที่ให้ผลลัพธ์ MAPE ที่ดีที่สุดจากข้อ 1. แล้วทำการพยากรณ์ เป็นระยะเวลาล่วงหน้า 30 เดือนให้เท่ากับจำนวนข้อมูลของชุดข้อมูลตรวจสอบ
 3. นำผลที่ได้จากการพยากรณ์มาเทียบกับชุดข้อมูลตรวจสอบเพื่อนำไปใช้ในการหา MAPE ของตัวแบบ
 4. นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด

3.3.3.6 วิธีพยากรณ์ด้วยตัวแบบโครงข่ายประสาทเทียม (Artificial Neural Network: ANN)

ตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่อง ANN จะทำการการพยากรณ์ในหัวข้อ 3.3.3 ถึง 3.3.5 โดยจะดำเนินการด้วยภาษา python บน Google Colab และจะมีการนำเข้า Library ที่จำเป็นในการสร้างตัวแบบ เช่น pandas numpy matplotlib.pyplot sklearn.model_selection keras.models keras.layers และ keras.wrappers.scikit_learn เป็นต้น โดยในส่วนของ ANN จะมีขั้นตอนการสร้างตัวแบบดังนี้

1. กำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบ ANN ในการทดสอบใน Grid Search CV

ตารางที่ 9 ค่าไฮเปอร์พารามิเตอร์ที่ใช้ทดสอบใน Grid Search CV ของตัวแบบ ANN

ไฮเปอร์พารามิเตอร์	ค่าที่ใช้ทดสอบใน Grid Search CV
Batch Size	8, 16, 32
Epochs	100, 200, 250, 300, 350
Activation	'relu', 'tanh', 'elu', 'selu'
Optimizer	'adam', 'rmsprop'
Hidden Units	100, 200, 250, 300, 350
Dropout Rate	0.2, 0.4, 0.6, 0.8, 1.0

- Batch Size คือ จำนวนข้อมูลที่ถูกแบ่งใช้ในการหาพารามิเตอร์เพื่อสร้างตัวแบบพยากรณ์ในแต่ละครั้ง การใช้ข้อมูลทั้งหมดในชุดข้อมูลฝึกสอนเพื่อหาพารามิเตอร์ที่เดียวไม่สามารถทำได้ เนื่องจากข้อจำกัดของหน่วยประมวลผล ดังนั้น จึงทำการแบ่งชุดข้อมูลเป็นกลุ่ม และใช้ เพื่อกำหนดขนาดของกลุ่มข้อมูลที่จะใช้ในการหาค่าพารามิเตอร์ของตัวแบบพยากรณ์
- Epochs คือ จำนวนรอบการฝึกสอนทั้งหมดที่ตัวแบบพยากรณ์ จะถูกฝึกสอนในข้อมูลทั้งหมด ในแต่ละรอบ (epoch) ตัวแบบพยากรณ์จะใช้ Batch Size ในการหาค่าพารามิเตอร์ของตัวแบบพยากรณ์
- Activation คือ ฟังก์ชันที่ใช้ในการคำนวณผลลัพธ์ของแต่ละชั้น (Layer) ได้แก่ ชั้นซ่อน และชั้นข้อมูลออกของตัวแบบพยากรณ์ โดยที่ 'relu', 'tanh', 'elu',

'selu' ย่อมาจาก Rectified Linear Unit, Hyperbolic Tangent, Exponential Linear Unit, Scaled Exponential Linear Unit Sigmoid ตามลำดับ

- Optimizer คือ อัลกอริทึมที่ใช้ในการปรับแต่งพารามิเตอร์ของตัวแบบพยากรณ์ ในระหว่างการฝึกสอน แต่ละอัลกอริทึมมีวิธีการปรับแต่งที่แตกต่างกัน และอาจมีผลต่อความแม่นยำและการฝึกสอนของตัวแบบพยากรณ์
 - Hidden Units คือ จำนวนของหน่วยซ่อนในชั้นซ่อน
 - Dropout Rate คือ อัตราส่วนของโหนดที่ถูกสุ่มลบออกจากการฝึกสอนในแต่ละชั้น (Layer) ของตัวแบบพยากรณ์ การใช้ Dropout สามารถช่วยลดปัญหา overfitting
2. สรุปค่าไฮเปอร์พารามิเตอร์ที่ให้ผลลัพธ์ MAPE ที่ดีที่สุดจากข้อ 1. แล้วทำการพยากรณ์ เป็นระยะเวลาล่วงหน้า 30 เดือนให้เท่ากับจำนวนข้อมูลของชุดข้อมูลตรวจสอบ
 3. นำผลที่ได้จากการพยากรณ์มาเทียบกับชุดข้อมูลตรวจสอบเพื่อนำไปใช้ในการหา MAPE ของตัวแบบ
 4. นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด

3.3.4 วิธีพยากรณ์ด้วยตัวแบบพยากรณ์ผสม (Hybrid Forecasting Model)

ตัวแบบพยากรณ์ผสมเป็นแบบการพยากรณ์ที่เกิดจากการนำตัวแบบพยากรณ์เดี่ยวในหัวข้อ 3.3.1 ถึง 3.3.3 มารวมกัน โดยมีจุดประสงค์คือเพื่อเพิ่มความแม่นยำในการพยากรณ์ราคาเม็ดพลาสติก PP มากขึ้น วิธีการสร้างตัวแบบพยากรณ์ผสมเป็นไปตามรูปที่ 31 โดยมีรายละเอียดขั้นตอนดังนี้

1. สร้างตัวแบบพยากรณ์เดี่ยวแล้วทำการพยากรณ์จากชุดข้อมูลฝึกสอนไปข้างหน้า 30 เดือนจะได้ผลลัพธ์เป็นค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลทดสอบ หรือ \hat{y}_t พร้อมทั้งหาค่าพารามิเตอร์ของตัวแบบพยากรณ์ที่ทำให้ผลลัพธ์การพยากรณ์มีความแม่นยำมากที่สุด โดยจะแบ่งเป็น 2 กลุ่ม กลุ่มแรก คือ ตัวแบบการถดถอยเชิงเส้นพหุคูณ และตัวแบบพยากรณ์ข้อมูลอนุกรมเวลา ได้แก่ ตัวแบบ SARIMA ตัวแบบ Holt-Winters และกลุ่มที่ 2 คือ ตัวแบบพยากรณ์การเรียนรู้ของเครื่อง ได้แก่ ตัวแบบ SVR ตัวแบบ Decision Tree ตัวแบบ Random Forest ตัวแบบ XG Boost และตัวแบบ ANN

2. คัดเลือกตัวแบบพยากรณ์ในกลุ่มแรกที่มีความแม่นยำมากที่สุดอันดับแรก และตัวแบบพยากรณ์ในกลุ่มที่ 2 ที่มีความแม่นยำมากที่สุด 3 อันดับแรก มาทำการสร้างตัวแบบผสม โดยจะแบ่งตัวแบบพยากรณ์ผสมเป็น 2 กลุ่ม ตัวแบบพยากรณ์ผสมกลุ่มแรก คือ การผสมตัวแบบพยากรณ์เดี่ยว ระหว่างตัวแบบพยากรณ์กลุ่มแรกและตัวแบบพยากรณ์กลุ่มที่ 2 และตัวแบบพยากรณ์ผสมกลุ่มที่ 2 คือ การผสมตัวแบบพยากรณ์เดี่ยว ระหว่างตัวแบบพยากรณ์กลุ่มที่ 2 ด้วยกันเอง โดยสามารถสรุปได้ดังตารางที่ 10 ซึ่งในตัวแบบพยากรณ์แบบผสมจะประกอบไปด้วยตัวแบบพยากรณ์เดี่ยว 2 ตัว โดยจะเรียกตัวแบบพยากรณ์ตัวแรกว่าตัวแบบพยากรณ์หลักและตัวแบบพยากรณ์เดี่ยวตัวถัดมาเป็นตัวแบบพยากรณ์รอง โดยในการคัดเลือกตัวแบบพยากรณ์ต้องมีค่า MAPE ที่ต่ำกว่า 10% และตัวแบบพยากรณ์หลักกับตัวแบบพยากรณ์รองจะไม่เป็นตัวแบบพยากรณ์ชนิดเดียวกัน เพราะจะคล้ายกับรูปแบบ Ensemble Learning ที่เกิดจากตัวแบบพยากรณ์เดี่ยวประเภทเดียวกันมารวมกัน

ตารางที่ 10 สรุปการแบ่งกลุ่มในการสร้างตัวแบบพยากรณ์ผสม

	กลุ่มที่ 1	กลุ่มที่ 2
ตัวแบบพยากรณ์เดี่ยว	ตัวแบบการถดถอยเชิงเส้นพหุคูณ ตัวแบบพยากรณ์ข้อมูลอนุกรมเวลา	ตัวแบบการเรียนรู้ของเครื่อง
ตัวแบบพยากรณ์ผสม	ตัวแบบพยากรณ์เดี่ยวจากกลุ่มที่ 1 + ตัวแบบพยากรณ์เดี่ยวจากกลุ่มที่ 2	ตัวแบบพยากรณ์เดี่ยวจากกลุ่มที่ 2 + ตัวแบบพยากรณ์เดี่ยวจากกลุ่มที่ 2

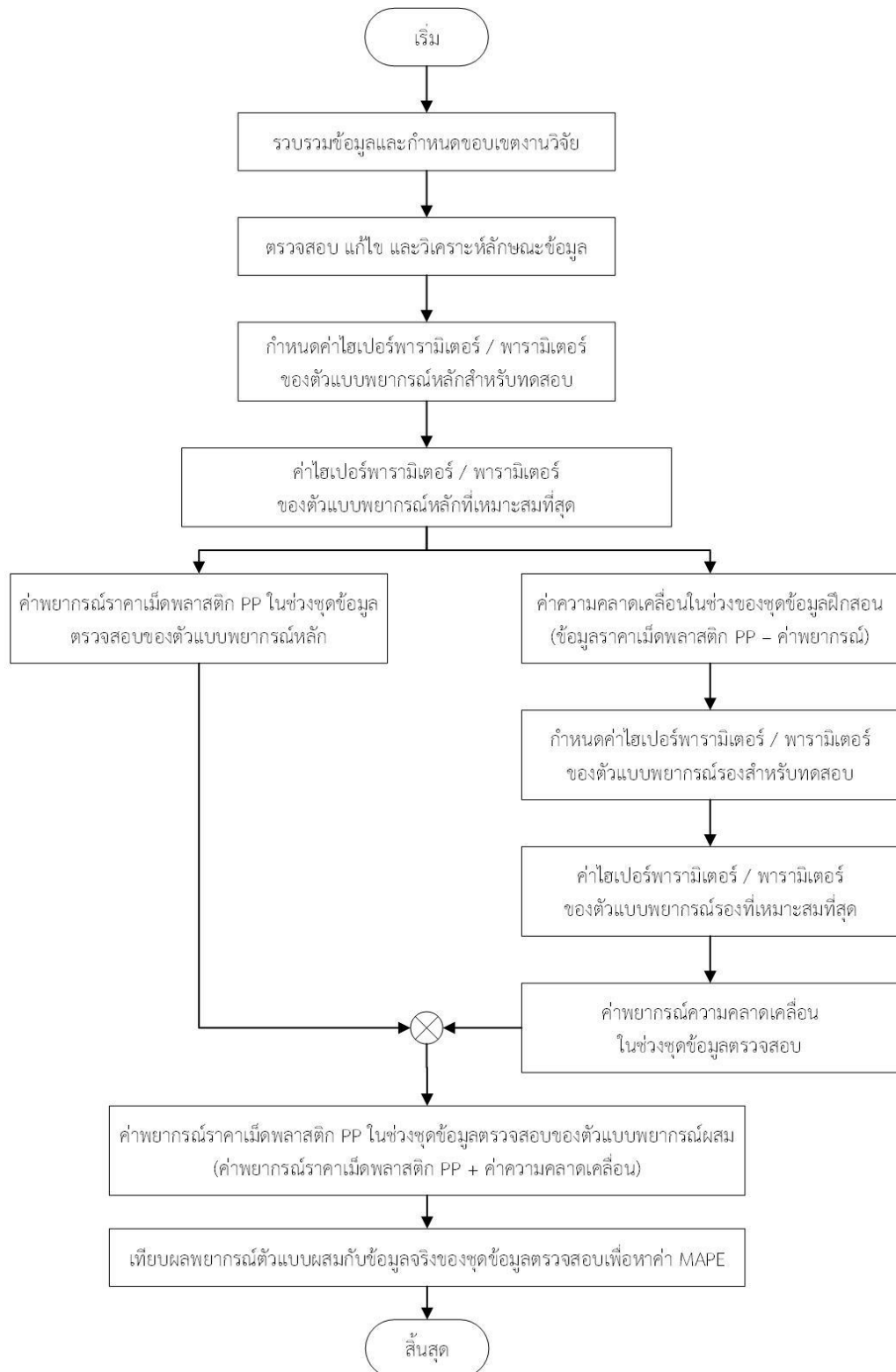
3. ใช้ค่าพารามิเตอร์จากตัวแบบพยากรณ์หลักมาทำการค่าความคลาดเคลื่อนในช่วงของชุดข้อมูลฝึกสอน
4. ใช้ค่าความคลาดเคลื่อนในช่วงของชุดข้อมูลฝึกสอนมาทำการพยากรณ์ไปข้างหน้า 30 เดือนด้วยตัวแบบพยากรณ์รอง จะได้ผลลัพธ์เป็นค่าพยากรณ์ความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบ หรือ $\hat{\epsilon}_t$
5. รวมค่าพยากรณ์ความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบจากตัวแบบพยากรณ์รอง เข้ากับค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบจากตัวแบบ

พยากรณ์หลัก จะได้ผลลัพธ์เป็นค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูล
ตรวจสอบของตัวแบบพยากรณ์ผสม หรือ \hat{y}

$$\hat{y} = \hat{y}_t + \hat{\epsilon}_t \quad (3.1)$$

- นำค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์
ผสมมาเทียบกับชุดข้อมูล Cross Validation Data Set เพื่อนำไปใช้ในการหา MAPE
ของตัวแบบ
- นำค่า MAPE ที่ได้ไปเปรียบเทียบกับตัวแบบอื่นเพื่อหาตัวแบบที่แม่นยำที่สุด





รูปที่ 31 ขั้นตอนของวิธีพยากรณ์ด้วยตัวแบบพยากรณ์ผสม

3.4 การเปรียบเทียบความแม่นยำของตัวแบบการพยากรณ์ โดยเทียบกับชุดข้อมูล Cross Validation Data Set

ขั้นตอนนี้จะเป็นการนำผลลัพธ์การพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบจากตัวแบบพยากรณ์ทั้งหมดตั้งแต่หัวข้อ 3.3.1 จนถึง 3.3.4 มาเทียบกับชุดข้อมูล Cross Validation Data Set เพื่อหาค่า MAPE ของแต่ละตัวแบบ แล้วนำค่า MAPE ของแต่ละตัวแบบการพยากรณ์มาเปรียบเทียบว่าตัวแบบใดที่ให้ค่า MAPE ต่ำที่สุด โดยจะถือว่าตัวแบบที่มีค่า MAPE ที่ต่ำที่สุดจะเป็นตัวแบบที่เหมาะสมที่สุดที่จะนำไปพยากรณ์ราคาเม็ดพลาสติก PP

3.5 การวัดความแม่นยำกับข้อมูล Testing Data Set

ขั้นตอนนี้จะเป็นการนำตัวแบบการพยากรณ์ที่เหมาะสมที่สุดจากขั้นตอนที่ 3.4 มาพยากรณ์ทำการในช่วงตั้งแต่ เดือนกรกฎาคม พ.ศ. 2563 ถึงเดือนธันวาคม พ.ศ. 2566 ซึ่งเปรียบเสมือนข้อมูลที่ยังไม่พบเจอทำให้ถือว่าเป็น Unbiased Error จากนั้นนำผลการพยากรณ์ไปวัดความแม่นยำกับชุดข้อมูล Testing Data Set จะได้ค่า MAPE ที่บ่งบอกถึงประสิทธิภาพของตัวแบบนั้น ๆ

บทที่ 4

ผลการศึกษา

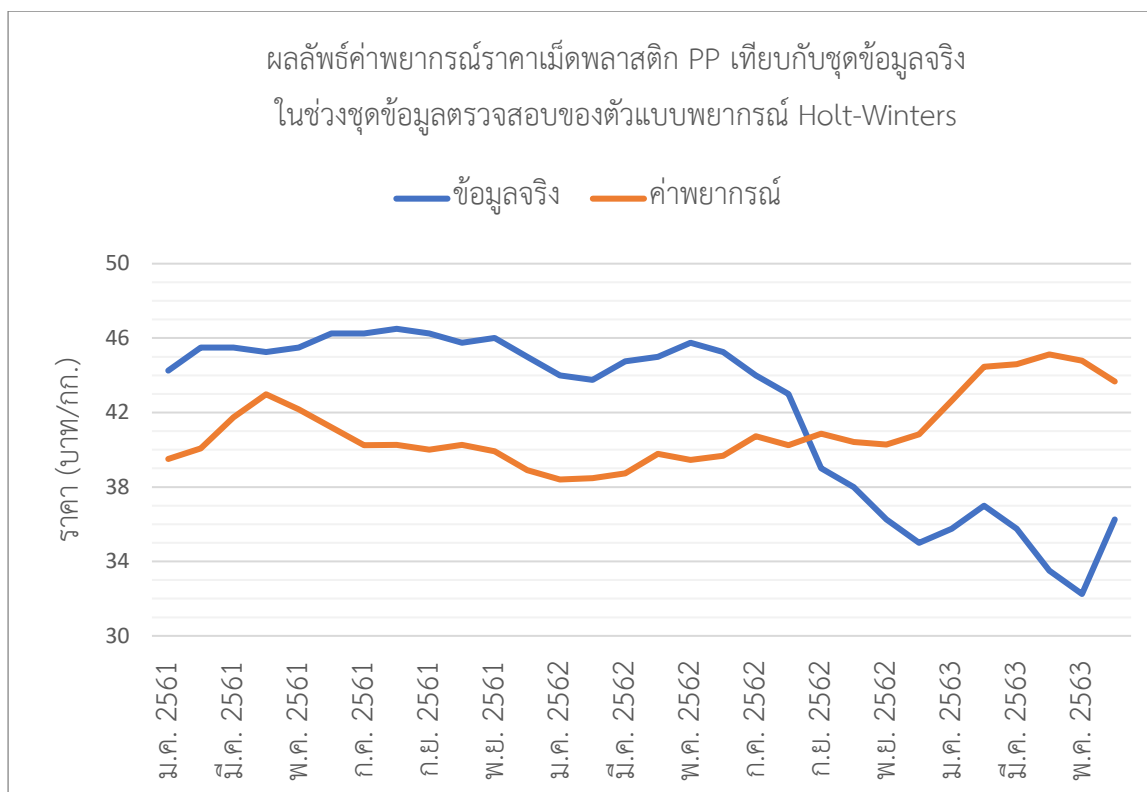
งานวิจัยฉบับนี้ได้ทำการศึกษาการพยากรณ์ราคาเม็ดพลาสติก PP ในประเทศไทย โดยใช้ตัวแบบพยากรณ์เชิงสาเหตุ คือ ตัวแบบการถดถอยเชิงเส้นพหุคูณ ตัวแบบพยากรณ์ข้อมูลอนุกรมเวลา ได้แก่ ตัวแบบ SARIMA ตัวแบบ Holt-Winters ตัวแบบพยากรณ์การเรียนรู้ของเครื่อง ได้แก่ ตัวแบบ SVR ตัวแบบ Decision Tree ตัวแบบ Random Forest ตัวแบบ XG Boost และตัวแบบ ANN และตัวแบบการพยากรณ์แบบผสม โดยได้ผลลัพธ์ดังนี้

4.1 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ Holt-Winters

จากการดำเนินการพยากรณ์ด้วยตัวแบบพยากรณ์ Holt-Winters ในรูปแบบ Multiplicative ตามหัวข้อ 3.3.1.1 ทำให้ได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบการพยากรณ์ ได้แก่ α β และ γ รวมถึงค่า MAPE ของผลลัพธ์จากการพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบตั้งในตารางที่ 11 ส่วนกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ Holt-Winters แสดงในรูปที่ 32

ตารางที่ 11 ค่าพารามิเตอร์ของตัวแบบพยากรณ์ Holt-Winters และค่า MAPE

พยากรณ์	พารามิเตอร์ของตัวแบบ Holt-Winters			MAPE
	α	β	γ	
ราคาเม็ดพลาสติก PP	0.9999999	8.571431e-10	2.491541e-10	13.99 %



รูปที่ 32 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ Holt-Winters

4.2 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SARIMA

จากการดำเนินการพยากรณ์ด้วยตัวแบบพยากรณ์ SARIMA ตามหัวข้อ 3.3.1.2 ทำให้ได้ผลลัพธ์ตัวแบบการพยากรณ์ที่เหมาะสมหลังจากการทำ Stepwise เพื่อหาค่า AIC ที่ต่ำที่สุดคือ SARIMA(1,0,1)(2,1,0)₁₂ โดยมีค่า AIC อยู่ที่ 328.352 ดังรูปที่ 33 ถัดมาเมื่อทำการทดสอบ z- test กับสัมประสิทธิ์ (Coefficients) ของแต่ละตัวแบบพบว่าสัมประสิทธิ์ทุกตัว จากทุกตัวแบบมีนัยสำคัญภายใต้ระดับความเชื่อมั่น 95% ดังแสดงในรูปที่ 34 จากนั้นทำการวิเคราะห์ตรวจสอบค่าความคลาดเคลื่อนของตัวแบบพยากรณ์ SARIMA (Diagnostic Checking) ได้ผลลัพธ์ดังรูปที่ 35 กล่าวคือค่า ACF PACF ของค่าความคลาดเคลื่อนจากตัวแบบพยากรณ์ไม่มีรูปแบบแนวโน้มและฤดูกาล และลักษณะการกระจายตัวของข้อมูลค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ (Normal Distribution) ทำให้สามารถสรุปได้ว่าตัวแบบพยากรณ์ดังกล่าวเหมาะสมแก่การนำไปพยากรณ์ราคาเม็ดพลาสติก PP และได้ผลลัพธ์ความแม่นยำ MAPE อยู่ที่ 8.09% ของผลลัพธ์จากการพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ ส่วนกราฟแสดงผลลัพธ์ค่าพยากรณ์

ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ SARIMA แสดงในรูปที่ 36 และได้ผลลัพธ์ค่า MAPE อยู่ที่ 8.09%

```

Performing stepwise search to minimize aic
ARIMA(1,0,1)(0,1,1)[12] intercept : AIC=inf, Time=1.78 sec
ARIMA(0,0,0)(0,1,0)[12] intercept : AIC=444.711, Time=0.07 sec
ARIMA(1,0,0)(1,1,0)[12] intercept : AIC=340.928, Time=0.76 sec
ARIMA(0,0,1)(0,1,1)[12] intercept : AIC=inf, Time=0.95 sec
ARIMA(0,0,0)(0,1,0)[12] : AIC=447.265, Time=0.03 sec
ARIMA(1,0,0)(0,1,0)[12] intercept : AIC=343.387, Time=0.06 sec
ARIMA(1,0,0)(2,1,0)[12] intercept : AIC=331.810, Time=0.52 sec
ARIMA(1,0,0)(2,1,1)[12] intercept : AIC=inf, Time=2.59 sec
ARIMA(1,0,0)(1,1,1)[12] intercept : AIC=inf, Time=1.73 sec
ARIMA(0,0,0)(2,1,0)[12] intercept : AIC=422.261, Time=0.58 sec
ARIMA(2,0,0)(2,1,0)[12] intercept : AIC=329.657, Time=0.73 sec
ARIMA(2,0,0)(1,1,0)[12] intercept : AIC=339.481, Time=0.25 sec
ARIMA(2,0,0)(2,1,1)[12] intercept : AIC=inf, Time=1.88 sec
ARIMA(2,0,0)(1,1,1)[12] intercept : AIC=inf, Time=1.27 sec
ARIMA(3,0,0)(2,1,0)[12] intercept : AIC=330.851, Time=0.87 sec
ARIMA(2,0,1)(2,1,0)[12] intercept : AIC=329.677, Time=0.90 sec
ARIMA(1,0,1)(2,1,0)[12] intercept : AIC=328.463, Time=0.74 sec
ARIMA(1,0,1)(1,1,0)[12] intercept : AIC=338.562, Time=0.25 sec
ARIMA(1,0,1)(2,1,1)[12] intercept : AIC=inf, Time=2.13 sec
ARIMA(1,0,1)(1,1,1)[12] intercept : AIC=inf, Time=1.52 sec
ARIMA(0,0,1)(2,1,0)[12] intercept : AIC=362.183, Time=2.57 sec
ARIMA(1,0,2)(2,1,0)[12] intercept : AIC=329.981, Time=0.86 sec
ARIMA(0,0,2)(2,1,0)[12] intercept : AIC=347.384, Time=0.86 sec
ARIMA(2,0,2)(2,1,0)[12] intercept : AIC=inf, Time=2.87 sec
ARIMA(1,0,1)(2,1,0)[12] : AIC=328.352, Time=0.39 sec
ARIMA(1,0,1)(1,1,0)[12] : AIC=337.117, Time=0.13 sec
ARIMA(1,0,1)(2,1,1)[12] : AIC=inf, Time=2.22 sec
ARIMA(1,0,1)(1,1,1)[12] : AIC=inf, Time=0.83 sec
ARIMA(0,0,1)(2,1,0)[12] : AIC=375.602, Time=0.41 sec
ARIMA(1,0,0)(2,1,0)[12] : AIC=331.015, Time=0.33 sec
ARIMA(2,0,1)(2,1,0)[12] : AIC=329.349, Time=0.71 sec
ARIMA(1,0,2)(2,1,0)[12] : AIC=329.596, Time=0.89 sec
ARIMA(0,0,0)(2,1,0)[12] : AIC=439.078, Time=0.39 sec
ARIMA(0,0,2)(2,1,0)[12] : AIC=356.234, Time=1.16 sec
ARIMA(2,0,0)(2,1,0)[12] : AIC=329.537, Time=0.81 sec
ARIMA(2,0,2)(2,1,0)[12] : AIC=inf, Time=2.52 sec

Best model: ARIMA(1,0,1)(2,1,0)[12]
Total fit time: 37.669 seconds

```

รูปที่ 33 ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบ SARIMA ของราคาเม็ดพลาสติก PP

Best model summary:

```

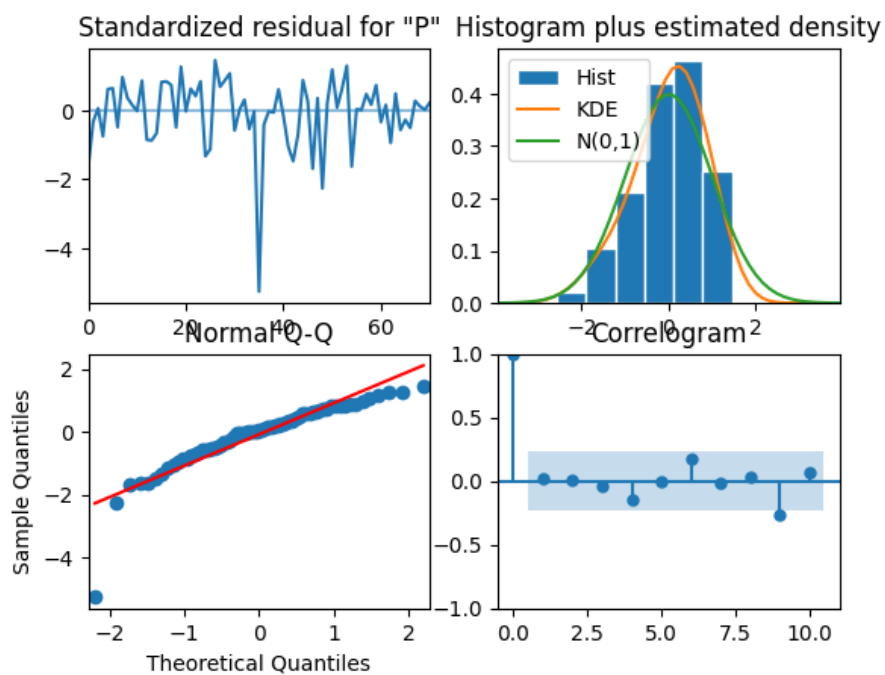
=====
SARIMAX Results
=====
Dep. Variable:          y      No. Observations:      84
Model:                SARIMAX(1, 0, 1)x(2, 1, [], 12)  Log Likelihood        -159.176
Date:                  Sun, 25 Jun 2023              AIC                   328.352
Time:                  19:34:39                       BIC                   339.735
Sample:                0                               HQIC                  332.884
                    - 84
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ar.L1         0.8336    0.090     9.276    0.000     0.657    1.010
ma.L1         0.3493    0.163     2.143    0.032     0.030    0.669
ar.S.L12      -0.5255    0.188    -2.798    0.005    -0.893   -0.157
ar.S.L24      -0.4209    0.144    -2.924    0.003    -0.703   -0.139
sigma2        4.3516    0.534     8.151    0.000     3.305    5.398
=====
Ljung-Box (L1) (Q):      0.23  Jarque-Bera (JB):      147.03
Prob(Q):                 0.63  Prob(JB):              0.00
Heteroskedasticity (H): 1.14  Skew:                  -1.68
Prob(H) (two-sided):    0.75  Kurtosis:              9.14
=====

```

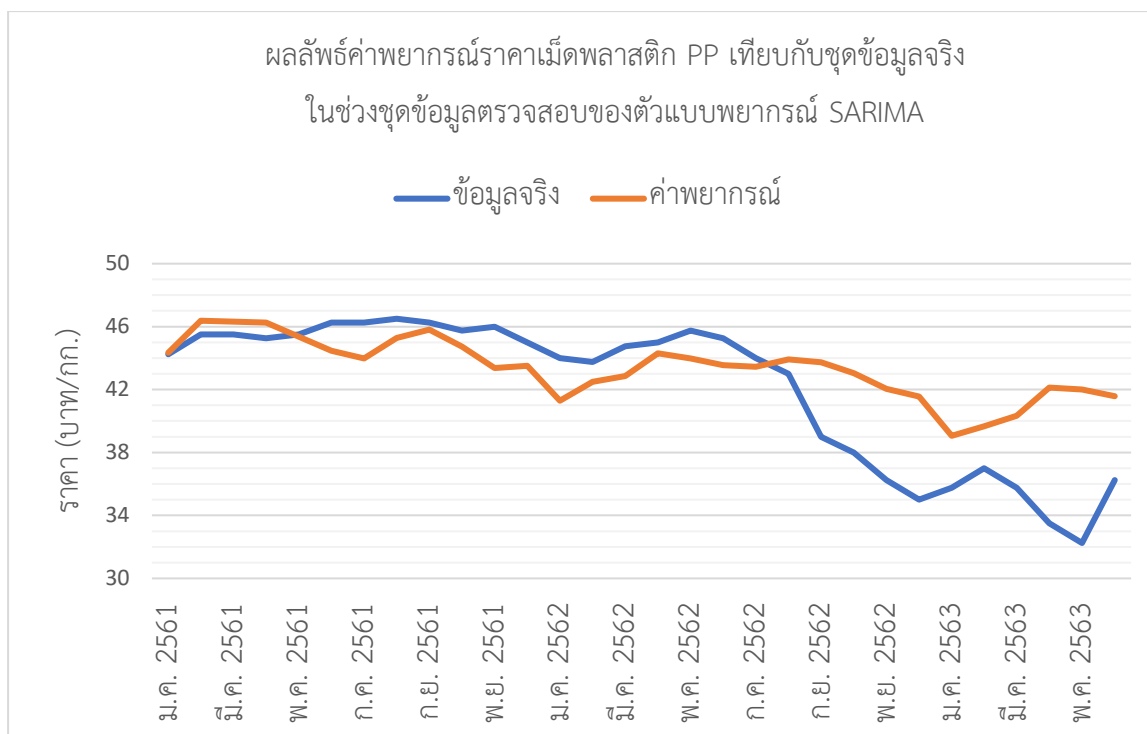
Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

รูปที่ 34 ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบ SARIMA ของราคาเม็ดพลาสติก PP



รูปที่ 35 การวิเคราะห์ตรวจสอบค่าความคลาดเคลื่อนของตัวแบบพยากรณ์ SARIMA (Diagnostic Checking)



รูปที่ 36 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SARIMA

4.3 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์วิเคราะห์การถดถอยเชิงเส้นพหุคูณ

เมื่อนำข้อมูลปัจจัยแต่ละตัวในช่วงชุดข้อมูลฝึกสอนเข้ากระบวนการถดถอยแบบขั้นตอน (Stepwise) จากนั้นพิจารณาค่า p-value ของแต่ละปัจจัยต้องไม่เกิน 0.05 และทำการทดสอบหาค่า VIF ของแต่ละปัจจัยหากปัจจัยใดมีค่าเกิน 10 ให้ตัดทิ้งจะทำให้เหลือปัจจัยอิสระที่ส่งผลต่อราคาเม็ดพลาสติก PP ดังนี้ ค่าดัชนีราคาผู้บริโภค (CPI)ราคาน้ำมันดิบ WTI จำนวนรถยนต์ที่ถูกผลิต รวมเป็น 3 ปัจจัย โดยเมื่อทำการทดสอบบนโปรแกรม Minitab เวอร์ชัน 18 ทำให้ได้สัมประสิทธิ์ของแต่ละปัจจัยและได้ค่า Adjusted R-Squared เท่ากับ 0.7530 ดังรูปที่ 37

Stepwise Selection of Terms

α to enter = 0.05, α to remove = 0.05

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	1377.40	459.13	85.36	0.000
Oil_Price_WTI	1	1194.30	1194.30	222.05	0.000
Consumer_Price_Index	1	45.62	45.62	8.48	0.005
Production_Car	1	64.30	64.30	11.95	0.001
Error	80	430.28	5.38		
Total	83	1807.69			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.31917	76.20%	75.30%	72.61%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.9	12.5	0.31	0.757	
Oil_Price_WTI	0.1781	0.0119	14.90	0.000	1.37
Consumer_Price_Index	0.371	0.127	2.91	0.005	1.41
Production_Car	-0.02346	0.00679	-3.46	0.001	1.07

Regression Equation

Price = 3.9 + 0.1781 Oil_Price_WTI + 0.371 Consumer_Price_Index - 0.02346 Production_Car

รูปที่ 37 ค่าสัมประสิทธิ์และ VIF ของแต่ละปัจจัยจากโปรแกรม Minitab

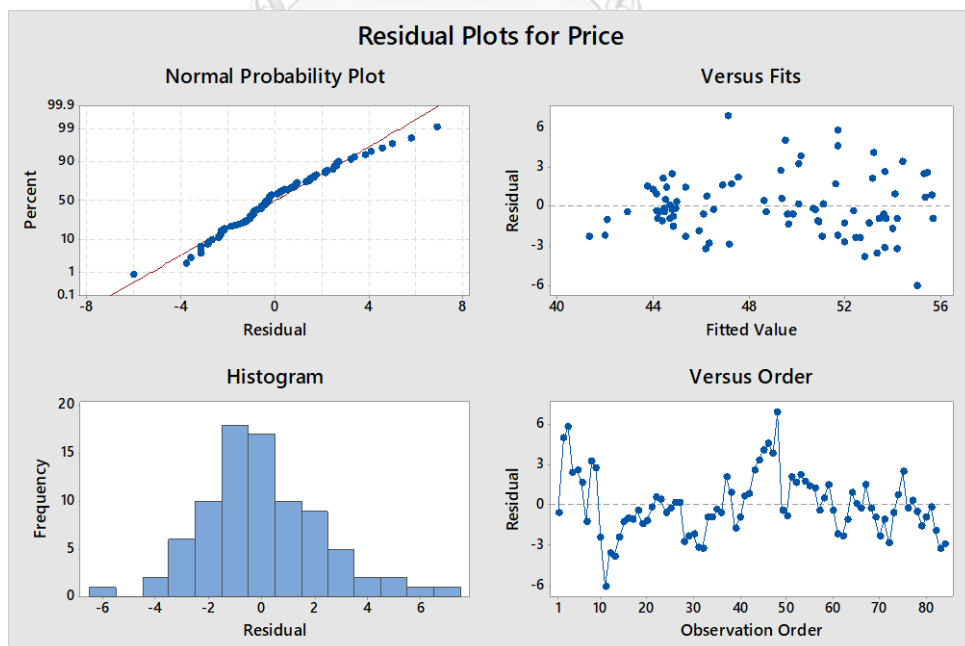
จากค่าสัมประสิทธิ์ที่ได้จะสามารถสร้างตัวแบบถดถอยเชิงพหุคูณจากชุดข้อมูล Training Data Set ได้ดังสมการที่ 4.1

$$\hat{Y} = 3.9 + 0.1781X_1 + 0.371X_2 - 0.02346X_3 \quad (4.1)$$

- โดยที่ \hat{Y} หมายถึง ผลลัพธ์การพยากรณ์
- X_1 หมายถึง ราคาน้ำมันดิบ WTI
- X_2 หมายถึง ค่าดัชนีราคาผู้บริโภค (CPI)
- X_3 หมายถึง จำนวนรถยนต์ที่ถูกผลิต

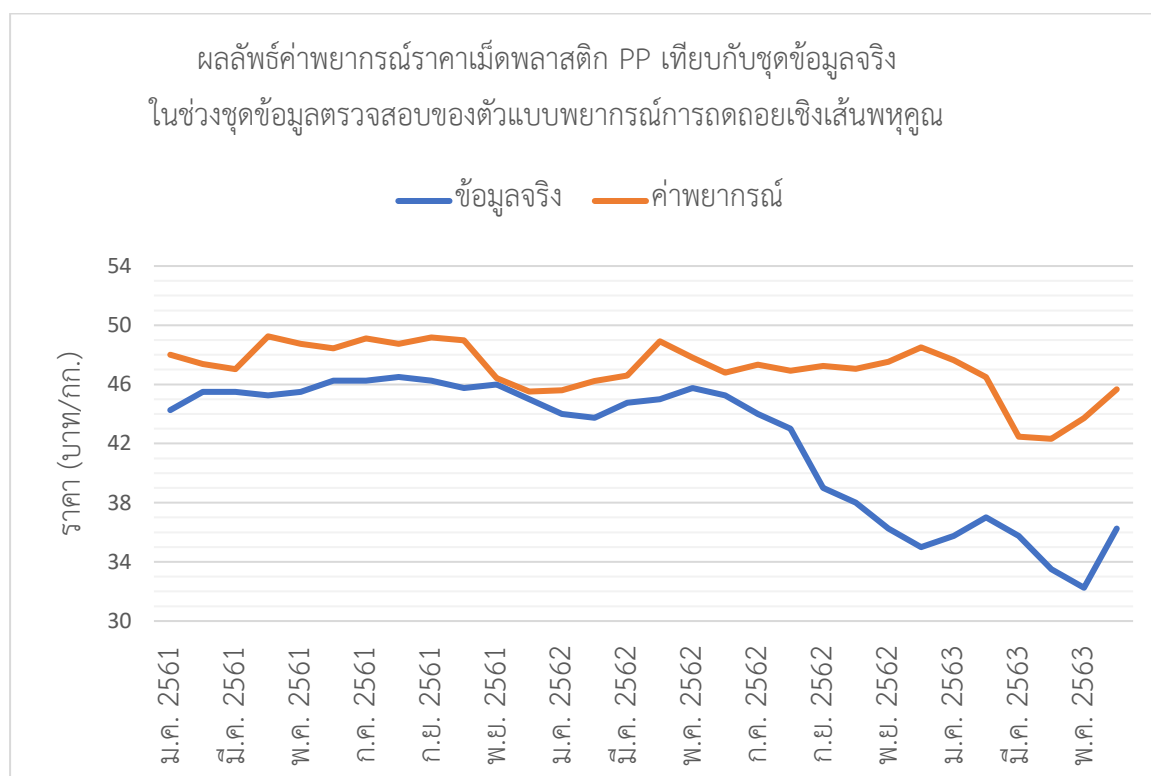
เมื่อพิจารณาค่าสัมประสิทธิ์ของแต่ละปัจจัยพบว่า 1.ดัชนีราคาผู้บริโภคซึ่งสะท้อนให้เห็นถึงภาวะเงินเฟ้อ โดยถ้าค่าดัชนีราคาผู้บริโภคมีค่าสูงหมายความว่าเกิดภาวะเงินเฟ้อซึ่งราคาสินค้าจะมีค่าเพิ่มขึ้นสอดคล้องไปในทิศทางเดียวกับราคาเม็ดพลาสติกที่จะเพิ่มขึ้นด้วยจากปริมาณความต้องการที่เพิ่มขึ้นทำให้สัมประสิทธิ์ควรมีค่าเป็นบวก 2.ราคาน้ำมันดิบ WTI จะมีผลโดยตรงต่อราคาพลาสติกเพราะพลาสติกถูกผลิตมาจากน้ำมันทำให้ราคาน้ำมันและราคาเม็ดพลาสติกมีทิศทางเดียวกันจึงควรมีสัมประสิทธิ์เป็นค่าบวก และตัวแปรสุดท้ายคือจำนวนรถยนต์ที่ถูกผลิตมีทิศทางเปลี่ยนแปลงไปในทางเดียวกับราคาเม็ดพลาสติกเพราะว่าในรถยนต์จะประกอบไปด้วยชิ้นส่วนที่ผลิตมาจากพลาสติกมากมายทำให้เมื่อผลิตรถยนต์เป็นจำนวนมากย่อมใช้เม็ดพลาสติกเป็นจำนวนมากเช่นกันทำให้ราคาเม็ดพลาสติกสูงขึ้นจึงทำให้สัมประสิทธิ์ควรมีค่าเป็นบวกแต่ผลลัพธ์ได้มาเป็นค่าลบ ซึ่งแสดงถึงการที่ข้อมูลมีความสัมพันธ์ในทิศทางตรงข้ามกัน โดยสันนิษฐานได้ว่าพลาสติก PP ไม่ได้ถูกใช้ในรถยนต์ในปริมาณมากเมื่อเทียบกับพลาสติกชนิดอื่น

ขั้นตอนถัดมาคือผลการทดสอบสมมติฐานของการถดถอยเชิงเส้นกับตัวแบบพยากรณ์ที่สร้างขึ้น ซึ่งจะทดสอบด้วยการนำค่าความผิดพลาด (Residual) ไปสร้างกราฟดังแสดงในรูปที่ 38 สามารถสรุปได้ว่าตัวแบบพยากรณ์สอดคล้องกับสมมติฐานของการถดถอย (ประชากรค่าความผิดพลาดมีค่าเฉลี่ยเท่ากับศูนย์ มีความแปรปรวนคงที่ มีการแจกแจงแบบปกติ และมีอิสระต่อกัน)



รูปที่ 38 ผลลัพธ์การทดสอบสมมติฐานการถดถอยของราคาเม็ดพลาสติก PP

ขั้นตอนสุดท้ายคือการนำตัวแบบถดถอยเชิงเส้นพหุคูณทำการพยากรณ์ไปล่วงหน้า 30 เดือน ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563 จะได้ผลลัพธ์ของการพยากรณ์แสดงใน รูปที่ 39 และได้ค่าผลลัพธ์ค่า MAPE อยู่ที่ 13.00%



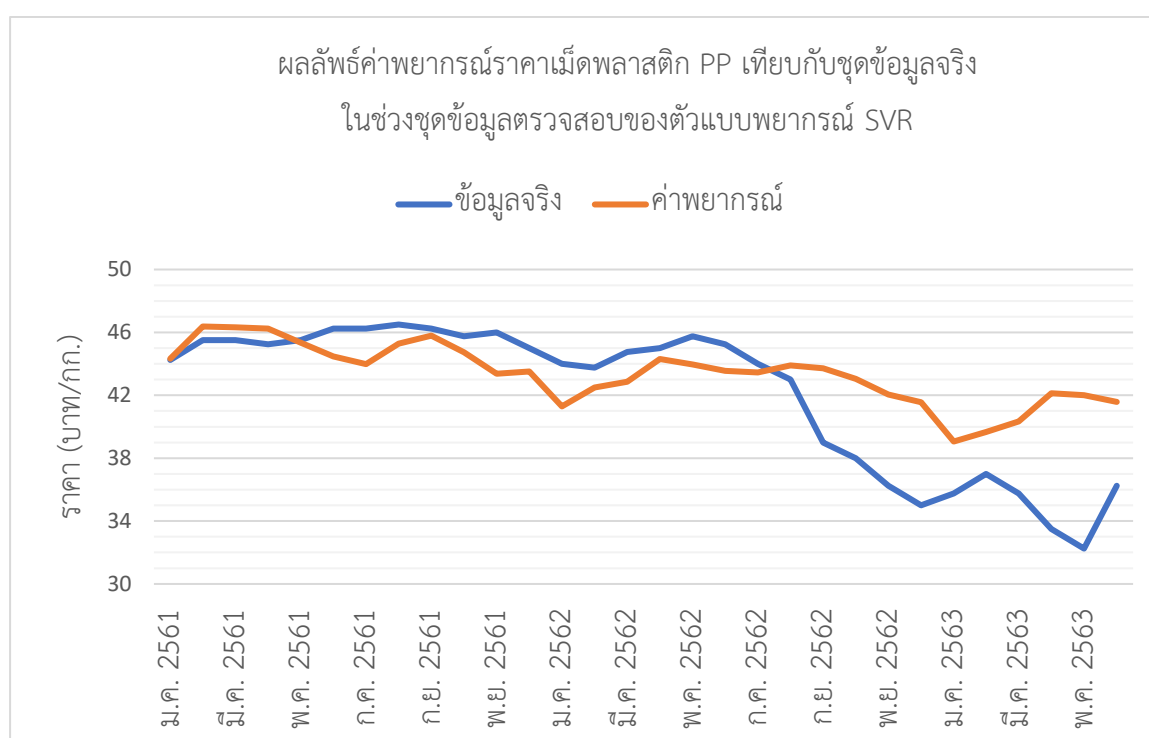
รูปที่ 39 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์วิเคราะห์การถดถอยเชิงเส้นพหุคูณ

4.4 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ Support Vector Regression

จากการดำเนินการพยากรณ์ด้วยตัวแบบพยากรณ์ SVR ตามหัวข้อ 3.3.3.2 ทำให้ได้ผลลัพธ์ตัวแบบการพยากรณ์ที่เหมาะสมหลังจากการทำ Gridsearch CV เพื่อหาค่า Hyperparameter ที่ทำให้ตัวแบบพยากรณ์ได้ผลลัพธ์ที่แม่นยำที่สุดดังตารางที่ 12 และได้ผลลัพธ์ความแม่นยำ MAPE อยู่ที่ 9.56 % ของผลลัพธ์จากการพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ ส่วนกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ SVR แสดงในรูปที่ 40 โดยในการฝึกสอนตัวแบบพยากรณ์ใช้เวลาทั้งสิ้น 13.48 นาที

ตารางที่ 12 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ SVR

พยากรณ์	ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์				MAPE
	Kernel	C	Gamma	Degree	
ราคาเม็ดพลาสติก PP	'rbf'	10	0.01	1	9.56 %



รูปที่ 40 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SVR

4.5 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ Decision Tree

จากการดำเนินการพยากรณ์ด้วยตัวแบบพยากรณ์ Decision Tree ตามหัวข้อ 3.3.3.3 ทำให้ได้ผลลัพธ์ตัวแบบการพยากรณ์ที่เหมาะสมหลังจากการทำ Gridsearch CV เพื่อหาค่า Hyperparameter ที่ทำให้ตัวแบบพยากรณ์ได้ผลลัพธ์ที่แม่นยำที่สุดดังตารางที่ 13 และได้ผลลัพธ์ความแม่นยำ MAPE อยู่ที่ 16.94 % ของผลลัพธ์จากการพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ ส่วนกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบ

กับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ Decision Tree แสดงในรูปที่ 41 โดยในการฝึกสอนตัวแบบพยากรณ์ใช้เวลาทั้งสิ้น 1.55 นาที

ตารางที่ 13 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ Decision Tree

ไฮเปอร์พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์	MAPE
Max Depth	10	16.94 %
Min Samples Split	40	
Min Samples Leaf	2	
Max Features	'sqrt'	
Splitter	'random'	



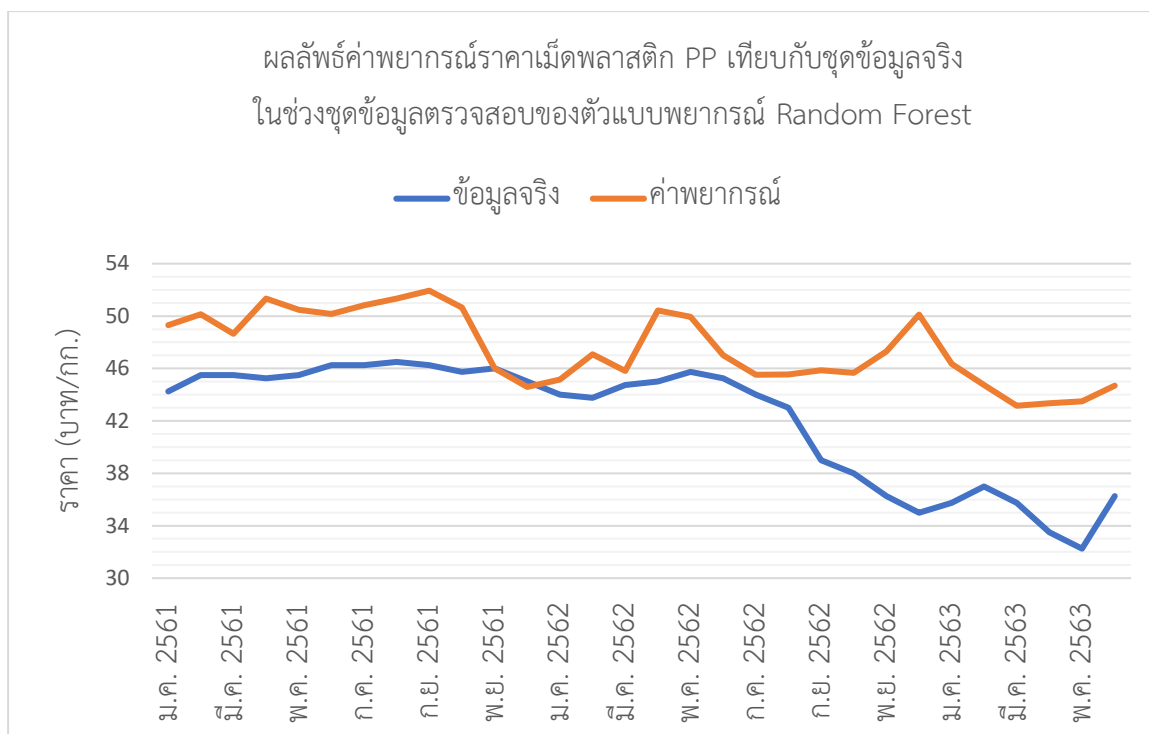
รูปที่ 41 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ Decision Tree

4.6 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ Random Forest

จากการดำเนินการพยากรณ์ด้วยตัวแบบพยากรณ์ Random Forest ตามหัวข้อ 3.3.3.4 ทำให้ได้ผลลัพธ์ตัวแบบการพยากรณ์ที่เหมาะสมหลังจากการทำ Gridsearch CV เพื่อหาค่า Hyperparameter ที่ทำให้ตัวแบบพยากรณ์ได้ผลลัพธ์ที่แม่นยำที่สุดดังตารางที่ 14 และได้ผลลัพธ์ความแม่นยำ MAPE อยู่ที่ 14.10 % ของผลลัพธ์จากการพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ ส่วนกราฟแสดงผลค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ Random Forest แสดงในรูปที่ 42 โดยในการฝึกสอนตัวแบบพยากรณ์ใช้เวลาทั้งสิ้น 35.30 นาที

ตารางที่ 14 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ Random Forest

ไฮเปอร์พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์	MAPE
N_estimator	250	14.10 %
Max Depth	9	
Min Samples Split	4	
Max Features	'log2'	



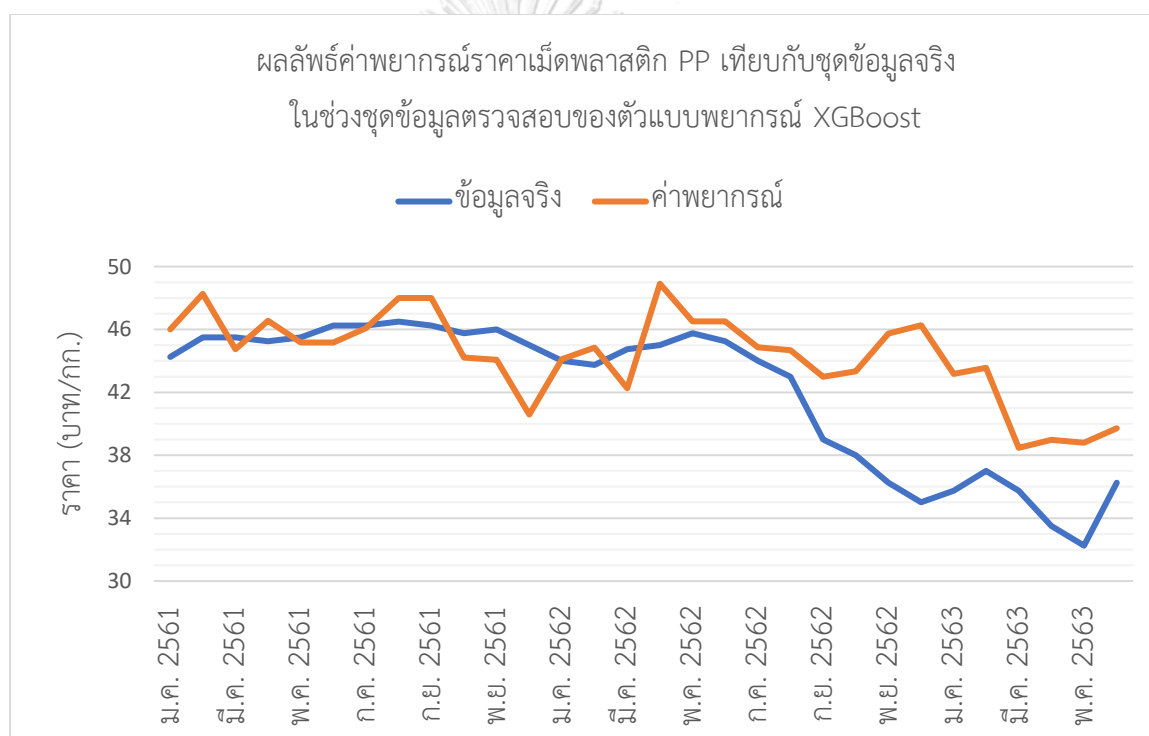
รูปที่ 42 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ Random Forest

4.7 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ XGBoost

จากการดำเนินการพยากรณ์ด้วยตัวแบบพยากรณ์ XGBoost ตามหัวข้อ 3.3.3.5 ทำให้ได้ผลลัพธ์ตัวแบบการพยากรณ์ที่เหมาะสมหลังจากการทำ Gridsearch CV เพื่อหาค่า Hyperparameter ที่ทำให้ตัวแบบพยากรณ์ได้ผลลัพธ์ที่แม่นยำที่สุดดังตารางที่ 15 และได้ผลลัพธ์ค่าความแม่นยำ MAPE อยู่ที่ 8.16 % ของผลลัพธ์จากการพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ ส่วนกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ XGBoost แสดงในรูปที่ 43 โดยในการฝึกสอนตัวแบบพยากรณ์ใช้เวลาทั้งสิ้น 106.5 นาที

ตารางที่ 15 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ XGBoost

ไฮเปอร์พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์	MAPE
N_estimator	700	8.16 %
Max Depth	8	
Learning Rate	0.4	
Subsample	0.7	
Gamma	0.8	



รูปที่ 43 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ XGBoost

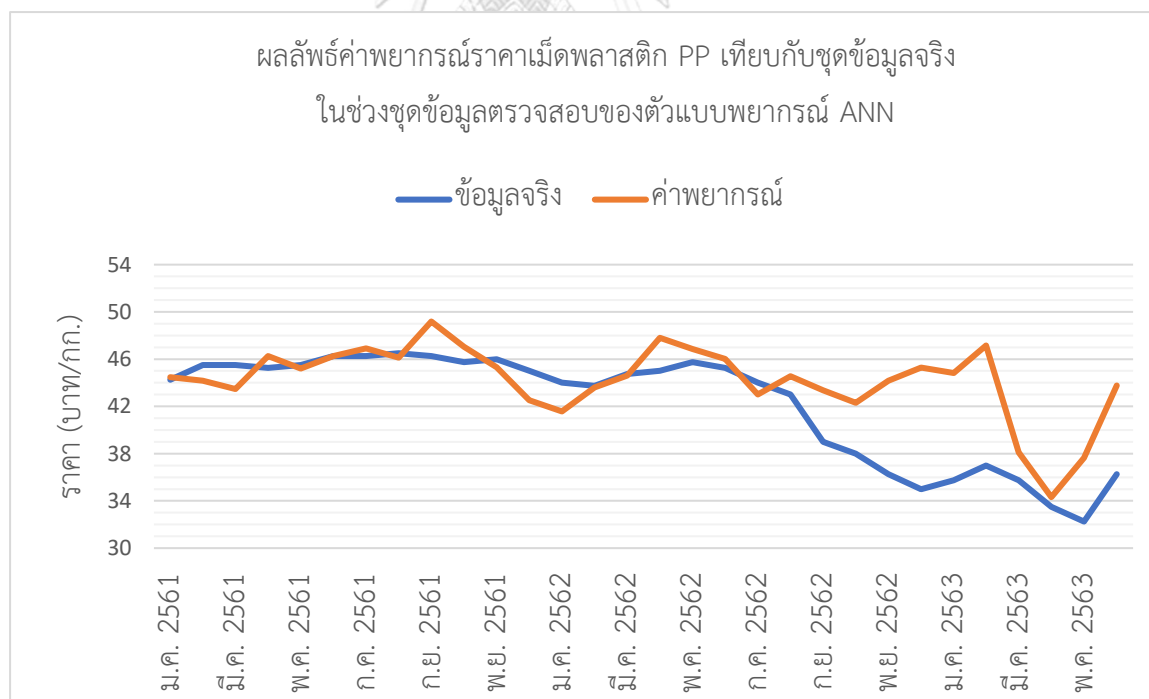
4.8 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ ANN

จากการดำเนินการพยากรณ์ด้วยตัวแบบพยากรณ์ ANN ตามหัวข้อ 3.3.3.6 ทำให้ได้ผลลัพธ์ตัวแบบการพยากรณ์ที่เหมาะสมหลังจากการทำ Gridsearch CV เพื่อหาค่า Hyperparameter ที่ทำให้ตัวแบบพยากรณ์ได้ผลลัพธ์ที่แม่นยำที่สุดดังตารางที่ 16 และได้ผลลัพธ์ความแม่นยำ MAPE อยู่ที่

7.71 % ของผลลัพธ์จากการพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ ส่วนกราฟแสดงผลค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ ANN แสดงในรูปที่ 44 โดยในการฝึกสอนตัวแบบพยากรณ์ใช้เวลาทั้งสิ้น 580.45 นาที

ตารางที่ 16 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมและค่า MAPE ของตัวแบบพยากรณ์ ANN

ไฮเปอร์พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์	MAPE
Batch Size	16	7.71 %
Epochs	300	
Activation	'elu'	
Optimizer	'rmsprop'	
Hidden Units	350	
Dropout Rate	0.6	



รูปที่ 44 กราฟแสดงผลค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ ANN

4.9 เปรียบเทียบผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์เดี่ยว

ผลลัพธ์จากตัวแบบพยากรณ์เดี่ยวที่ใช้พยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริง ในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ตั้งแต่หัวข้อที่ 4.1 ถึง 4.8 พบว่าตัวแบบพยากรณ์ที่มีความแม่นยำสูงที่สุดในตัวแบบพยากรณ์กลุ่มแรกในหัวข้อ 3.3.4 คือ ตัวแบบพยากรณ์ SARIMA มีค่า MAPE เท่ากับ 8.09 % ในส่วนของตัวแบบพยากรณ์ที่มีความแม่นยำสูงที่สุด 3 อันดับแรกในกลุ่มตัวแบบการพยากรณ์การเรียนรู้ด้วยเครื่อง คือ ตัวแบบพยากรณ์ ANN มีค่า MAPE เท่ากับ 7.71 % ตัวแบบพยากรณ์ XGBoost มีค่า MAPE เท่ากับ 8.09 % และตัวแบบพยากรณ์ SVR มีค่า MAPE เท่ากับ 9.56 % ดังที่สรุปในตารางที่ 17

ตารางที่ 17 เปรียบเทียบผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์เดี่ยว

วิธีการพยากรณ์	ตัวแบบพยากรณ์	MAPE
วิธีพยากรณ์ข้อมูลอนุกรมเวลา	ตัวแบบพยากรณ์ SARIMA	8.09 %
	ตัวแบบพยากรณ์ Holt-Winters	14.77 %
วิธีพยากรณ์ข้อมูลเชิงสาเหตุ	ตัวแบบการพยากรณ์ถดถอยเชิงเส้นพหุคูณ	13.00 %
วิธีพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง	ตัวแบบพยากรณ์ SVR	9.56 %
	ตัวแบบพยากรณ์ Decision Tree	16.94 %
	ตัวแบบพยากรณ์ Random Forest	14.10 %
	ตัวแบบพยากรณ์ XGBoost	8.16 %
	ตัวแบบพยากรณ์ ANN	7.71 %

4.10 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ผสม

ตัวแบบพยากรณ์ผสมเกิดจากการนำตัวแบบพยากรณ์เดี่ยว 2 ตัวมารวมกัน โดยสามารถช่วยให้มีความแม่นยำสูงขึ้นกว่าตัวแบบพยากรณ์เดี่ยว ในส่วนของการดำเนินการสร้างตัวแบบพยากรณ์ผสมตามหัวข้อที่ 3.3.4 โดยจะทำการคัดเลือกตัวแบบพยากรณ์เดี่ยวในกลุ่มแรกที่มีความแม่นยำมากที่สุดอันดับแรก และตัวแบบพยากรณ์เดี่ยวในกลุ่มที่ 2 ที่มีความแม่นยำมากที่สุด 3 อันดับแรก มาทำการสร้างตัวแบบผสมโดยจะแบ่งตัวแบบพยากรณ์ผสมเป็น 2 กลุ่ม ดังตารางที่ 18

ตารางที่ 18 สรุปการแบ่งกลุ่มในการสร้างตัวแบบพยากรณ์ผสม

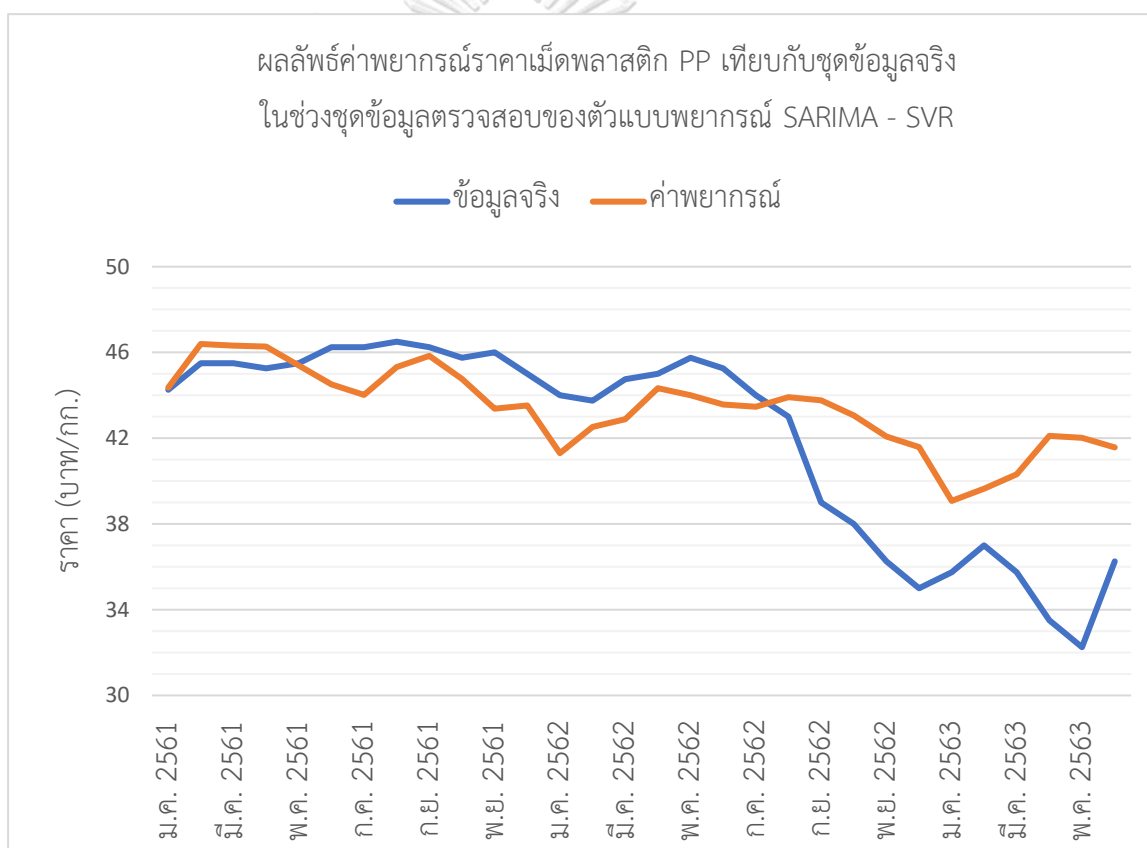
	กลุ่มที่ 1	กลุ่มที่ 2
ตัวแบบพยากรณ์ แบบผสม	ตัวแบบพยากรณ์ SARIMA - SVR	ตัวแบบพยากรณ์ ANN - XG
	ตัวแบบพยากรณ์ SARIMA - XG	ตัวแบบพยากรณ์ ANN - SVR
	ตัวแบบพยากรณ์ SARIMA - ANN	ตัวแบบพยากรณ์ XG - ANN
	ตัวแบบพยากรณ์ SVR - SARIMA	ตัวแบบพยากรณ์ XG - SVR
	ตัวแบบพยากรณ์ XG - SARIMA	ตัวแบบพยากรณ์ SVR - XG
	ตัวแบบพยากรณ์ ANN - SARIMA	ตัวแบบพยากรณ์ SVR - ANN

4.10.1 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SARIMA - SVR

ตัวแบบพยากรณ์ SARIMA – SVR สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ SARIMA และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ SVR ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.2 เช่น ค่าพารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 19 โดยการพยากรณ์ของตัวแบบพยากรณ์ SARIMA – SVR ใช้เวลาทั้งสิ้น 5.30 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 7.18 % ซึ่งกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ SARIMA - SVR แสดงในรูปที่ 45

ตารางที่ 19 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SARIMA - SVR และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	พารามิเตอร์ของตัวแบบพยากรณ์	SARIMA(1,0,1)(2,1,0) ₁₂
ตัวแบบพยากรณ์รอง	Kernel	'sigmoid'
	C	0.01
	Gamma	'scale'
MAPE	7.18 %	



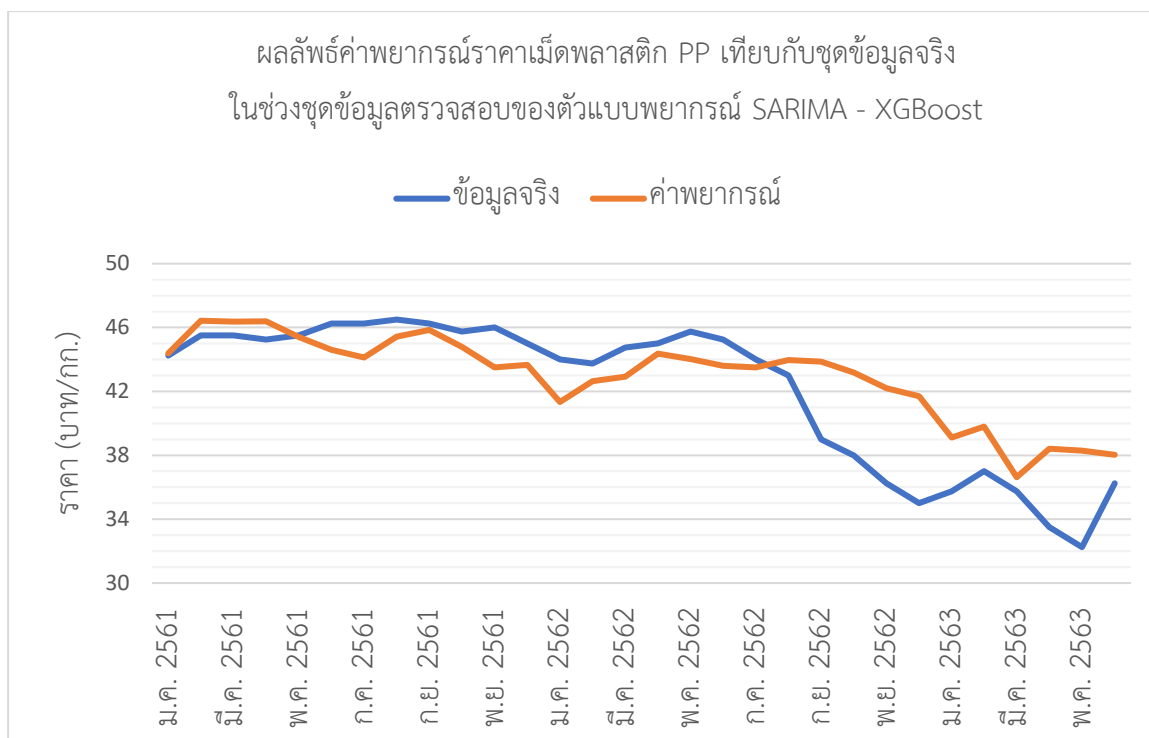
รูปที่ 45 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SARIMA - SVR

4.10.2 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SARIMA – XGBoost

ตัวแบบพยากรณ์ SARIMA – XGBoost สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ SARIMA และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ XGBoost ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.2 เช่นค่าพารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูล ตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 20 โดยการพยากรณ์ของตัวแบบพยากรณ์ SARIMA – XGBoost ใช้เวลาทั้งสิ้น 11.45 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูล ตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลทดสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 5.77 % ซึ่งกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูล ตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ SARIMA - XGBoost แสดงในรูปที่ 46

ตารางที่ 20 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SARIMA – XGBoost และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	พารามิเตอร์ของตัวแบบพยากรณ์	SARIMA(1,0,1)(2,1,0) ₁₂
ตัวแบบพยากรณ์รอง	N_estimator	100
	Max Depth	3
	Learning Rate	0.01
	Subsample	1
	Gamma	0
MAPE	5.77 %	



รูปที่ 46 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SARIMA - XGBoost

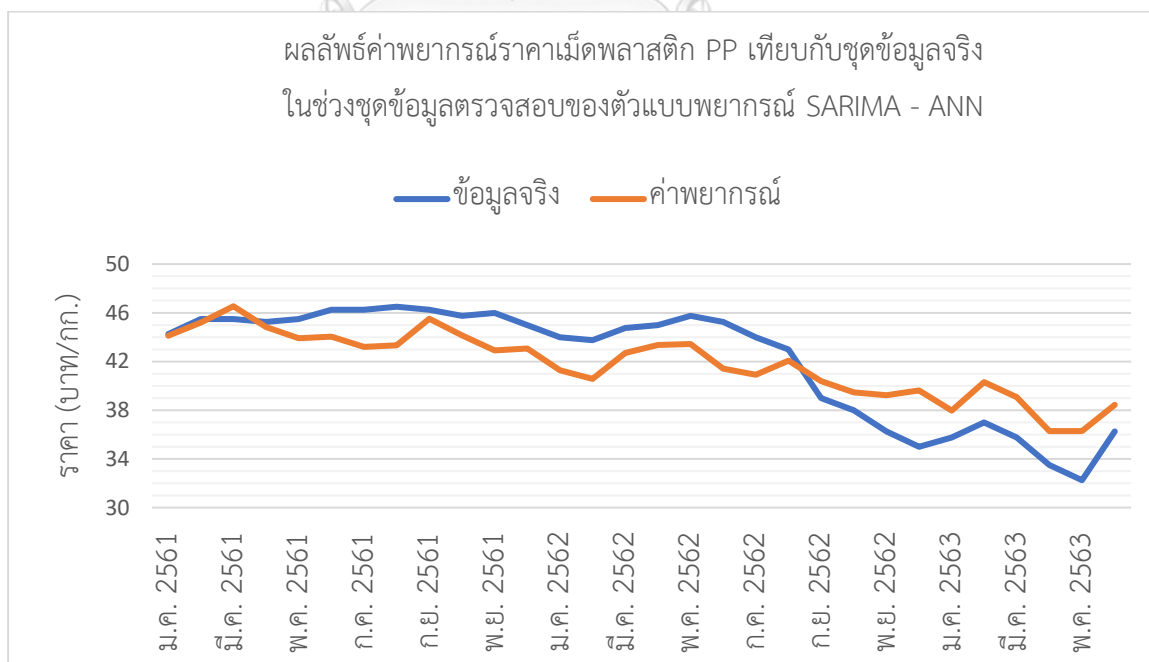
4.10.3 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SARIMA – ANN

ตัวแบบพยากรณ์ SARIMA – ANN สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ SARIMA และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ ANN ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.2 เช่น ค่าพารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบและนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 21 โดยการพยากรณ์ของตัวแบบพยากรณ์ SARIMA – ANN ใช้เวลาทั้งสิ้น 11.45 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 5.54 % ซึ่งกราฟ

แสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ SARIMA - ANN แสดงในรูปที่ 47

ตารางที่ 21 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SARIMA – ANN และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	พารามิเตอร์ของตัวแบบพยากรณ์	SARIMA(1,0,1)(2,1,0) ₁₂
ตัวแบบพยากรณ์รอง	Batch Size	32
	Epochs	100
	Activation	'relu'
	Optimizer	'rmsprop'
	Hidden Units	200
	Dropout Rate	0.4
MAPE	5.54 %	



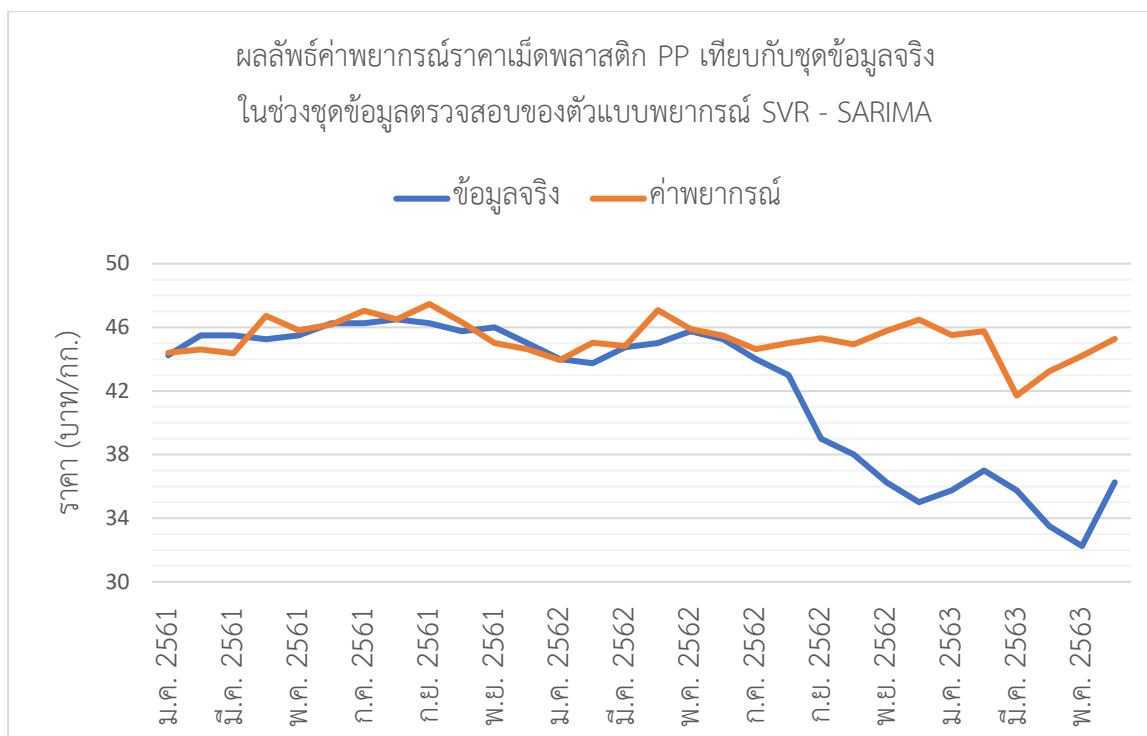
รูปที่ 47 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SARIMA – ANN

4.10.4 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SVR - SARIMA

ตัวแบบพยากรณ์ SVR - SARIMA สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบด้วยด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ SVR และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ SARIMA ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.4 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการทำ stepwise เพื่อหาค่า AIC ที่ต่ำที่สุดซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 22 โดยการพยากรณ์ของตัวแบบพยากรณ์ SVR - SARIMA ใช้เวลาทั้งสิ้น 32 วินาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 9.48 % ซึ่งกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ SVR - SARIMA แสดงในรูปที่ 48

ตารางที่ 22 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SVR - SARIMA และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	Kernel	'rbf'
	C	10
	Gamma	0.01
	Degree	1
ตัวแบบพยากรณ์รอง	พารามิเตอร์ของตัวแบบพยากรณ์	SARIMA(1,0,0)(0,0,0) ₁₂
MAPE	9.48 %	



รูปที่ 48 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SVR - SARIMA

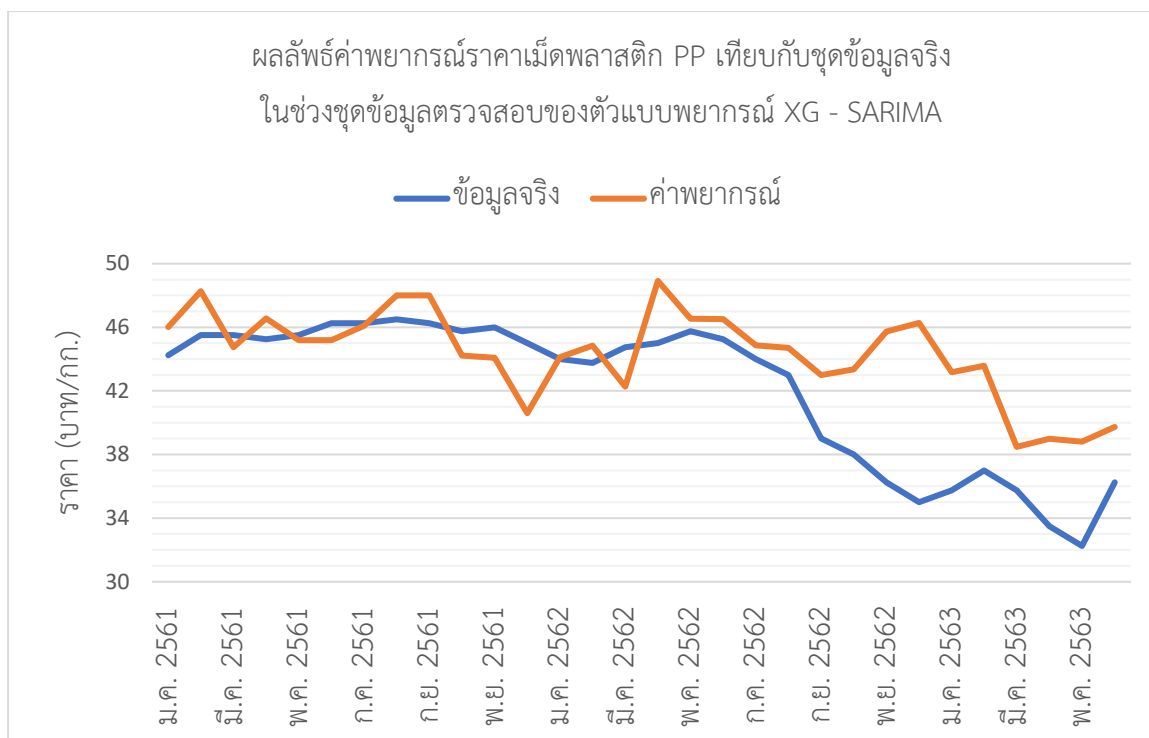
4.10.5 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ XGBoost - SARIMA

ตัวแบบพยากรณ์ XGBoost - SARIMA สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ XGBoost และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ SARIMA ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.7 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการทำ stepwise เพื่อหาค่า AIC ที่ต่ำที่สุดซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองตั้งตารางที่ 23 โดยการพยากรณ์ของตัวแบบพยากรณ์ XGBoost - SARIMA ใช้เวลาทั้งสิ้น 23 วินาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า

MAPE อยู่ที่ 8.16 % ซึ่งกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริง ในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ XGBoost - SARIMA แสดงในรูปที่ 49

ตารางที่ 23 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ XGBoost - SARIMA และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	N_estimator	700
	Max Depth	8
	Learning Rate	0.4
	Subsample	0.7
	Gamma	0.8
ตัวแบบพยากรณ์รอง	พารามิเตอร์ของตัวแบบพยากรณ์	SARIMA(0,0,0)(0,0,0) ₁₂
MAPE	8.16 %	



รูปที่ 49 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ XGBoost - SARIMA

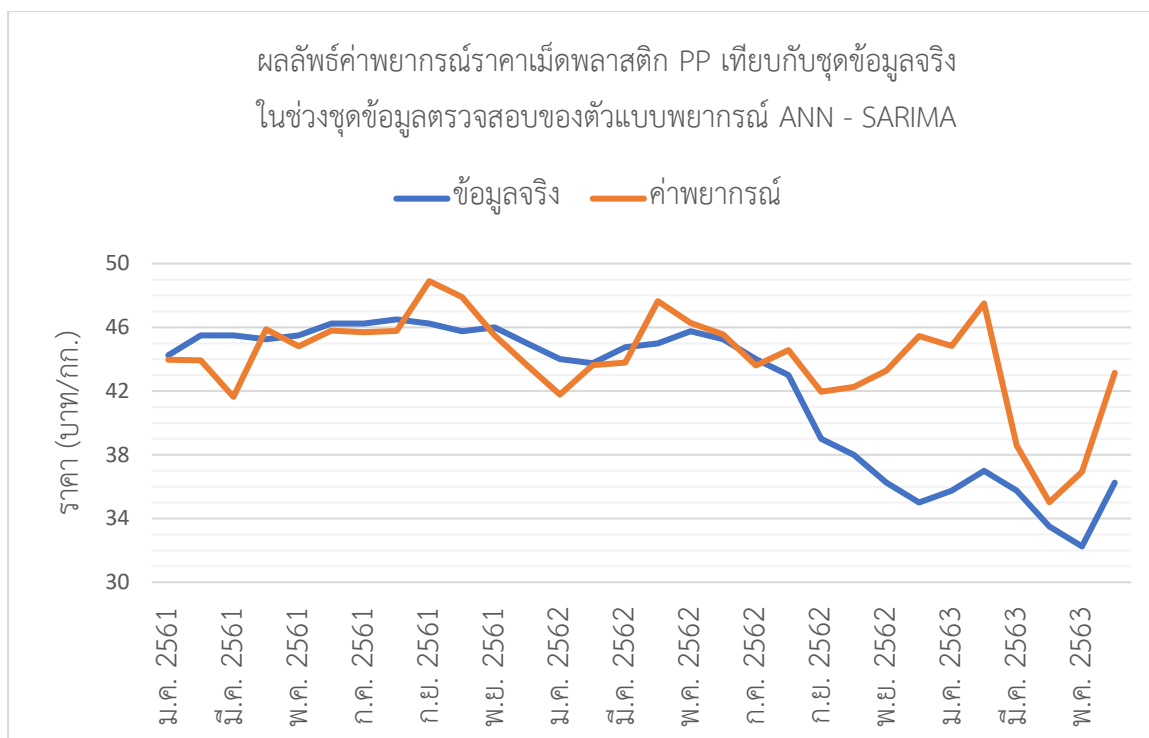
4.10.6 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ ANN - SARIMA

ตัวแบบพยากรณ์ ANN - SARIMA สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ ANN และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ SARIMA ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.8 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการทำ stepwise เพื่อหาค่า AIC ที่ต่ำที่สุดซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 24 โดยการพยากรณ์ของตัวแบบพยากรณ์ ANN - SARIMA ใช้เวลาทั้งสิ้น 53 วินาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลทดสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 7.38

% ซึ่งกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ ANN - SARIMA แสดงในรูปที่ 50

ตารางที่ 24 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ ANN - SARIMA และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	Batch Size	16
	Epochs	300
	Activation	'elu'
	Optimizer	'rmsprop'
	Hidden Units	350
	Dropout Rate	0.6
ตัวแบบพยากรณ์รอง	พารามิเตอร์ของตัวแบบพยากรณ์	SARIMA(0,0,3)(2,0,1) ₁₂
MAPE	7.38 %	



รูปที่ 50 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ ANN - SARIMA

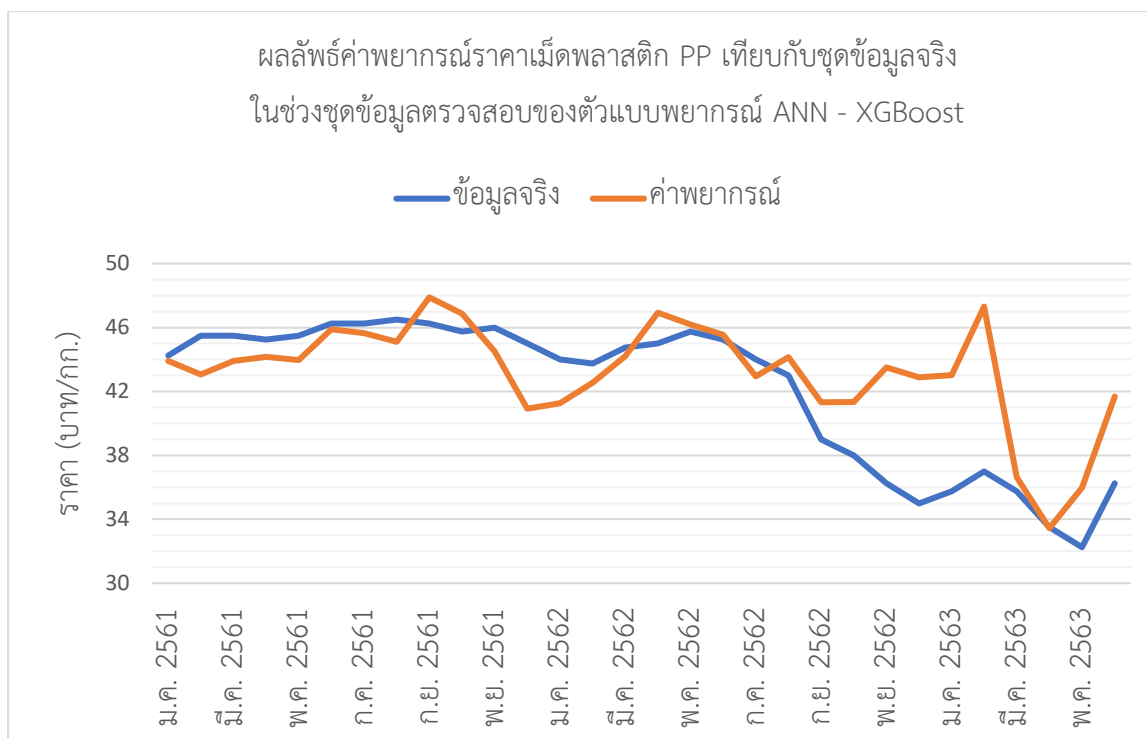
4.10.7 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ ANN - XGBoost

ตัวแบบพยากรณ์ ANN - XGBoost สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ ANN และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ XGBoost ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.8 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 25 โดยการพยากรณ์ของตัวแบบพยากรณ์ ANN - XGBoost ใช้เวลาทั้งสิ้น 20.2 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 6.49 % ซึ่งกราฟ

แสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ ANN - XGBoost แสดงในรูปที่ 51

ตารางที่ 25 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ ANN - XGBoost และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	Batch Size	16
	Epochs	300
	Activation	'elu'
	Optimizer	'rmsprop'
	Hidden Units	350
	Dropout Rate	0.6
ตัวแบบพยากรณ์รอง	N_estimator	100
	Max Depth	3
	Learning Rate	0.1
	Subsample	0.8
	Gamma	0.1
MAPE	6.49 %	



รูปที่ 51 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ ANN - XGBoost

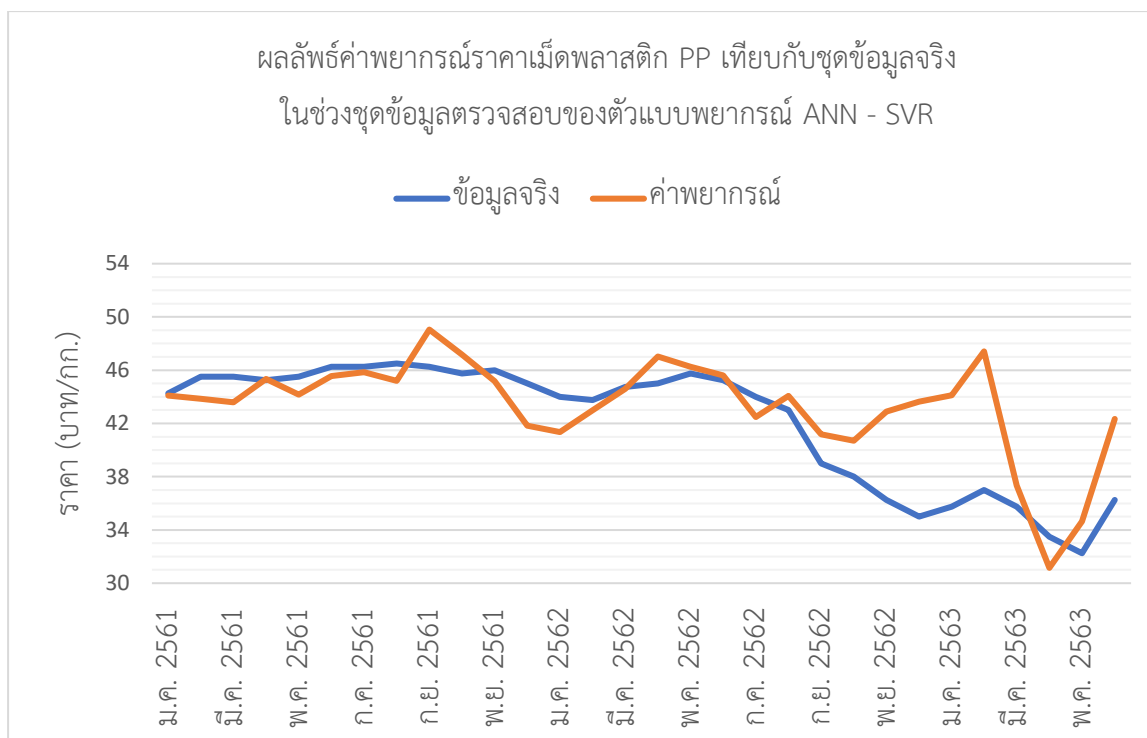
4.10.8 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ ANN - SVR

ตัวแบบพยากรณ์ ANN - SVR สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ ANN และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ SVR ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.8 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 26 โดยการพยากรณ์ของตัวแบบพยากรณ์ ANN - SVR ใช้เวลาทั้งสิ้น 6.10 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 6.60 % ซึ่งกราฟแสดงผล

ลัพท์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ ANN - SVR แสดงในรูปที่ 52

ตารางที่ 26 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ ANN - SVR และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	Batch Size	16
	Epochs	300
	Activation	'elu'
	Optimizer	'rmsprop'
	Hidden Units	350
	Dropout Rate	0.6
ตัวแบบพยากรณ์รอง	Kernel	'poly'
	C	15
	Gamma	0.01
	Degree	2
MAPE	6.60 %	



รูปที่ 52 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ ANN - SVR

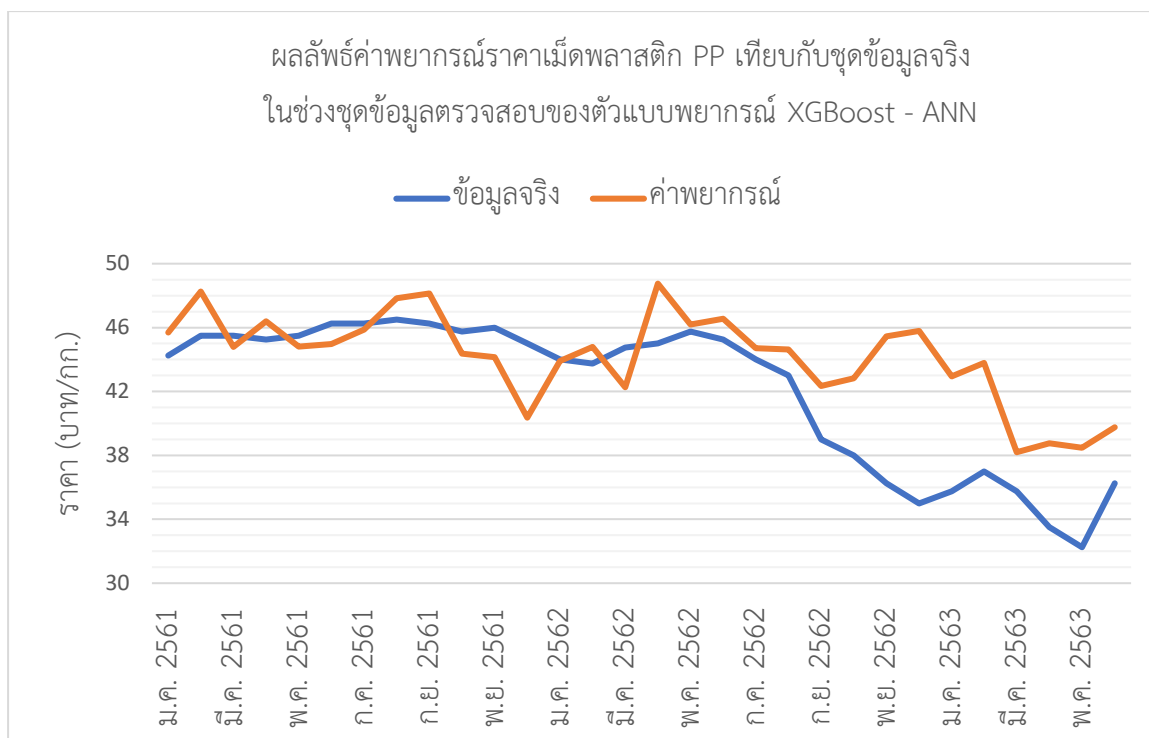
4.10.9 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ XGBoost - ANN

ตัวแบบพยากรณ์ XGBoost - ANN สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ XGBoost และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ ANN ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.7 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 27 โดยการพยากรณ์ของตัวแบบพยากรณ์ XGBoost - ANN ใช้เวลาทั้งสิ้น 600 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 7.87 % ซึ่ง

กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ XGBoost - ANN แสดงในรูปที่ 53

ตารางที่ 27 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ XGBoost - ANN และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	N_estimator	700
	Max Depth	8
	Learning Rate	0.4
	Subsample	0.7
	Gamma	0.8
ตัวแบบพยากรณ์รอง	Batch Size	16
	Epochs	250
	Activation	'tanh'
	Optimizer	'adam'
	Hidden Units	100
	Dropout Rate	0.8
MAPE	7.87 %	



รูปที่ 53 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ ของตัวแบบพยากรณ์ XGBoost - ANN

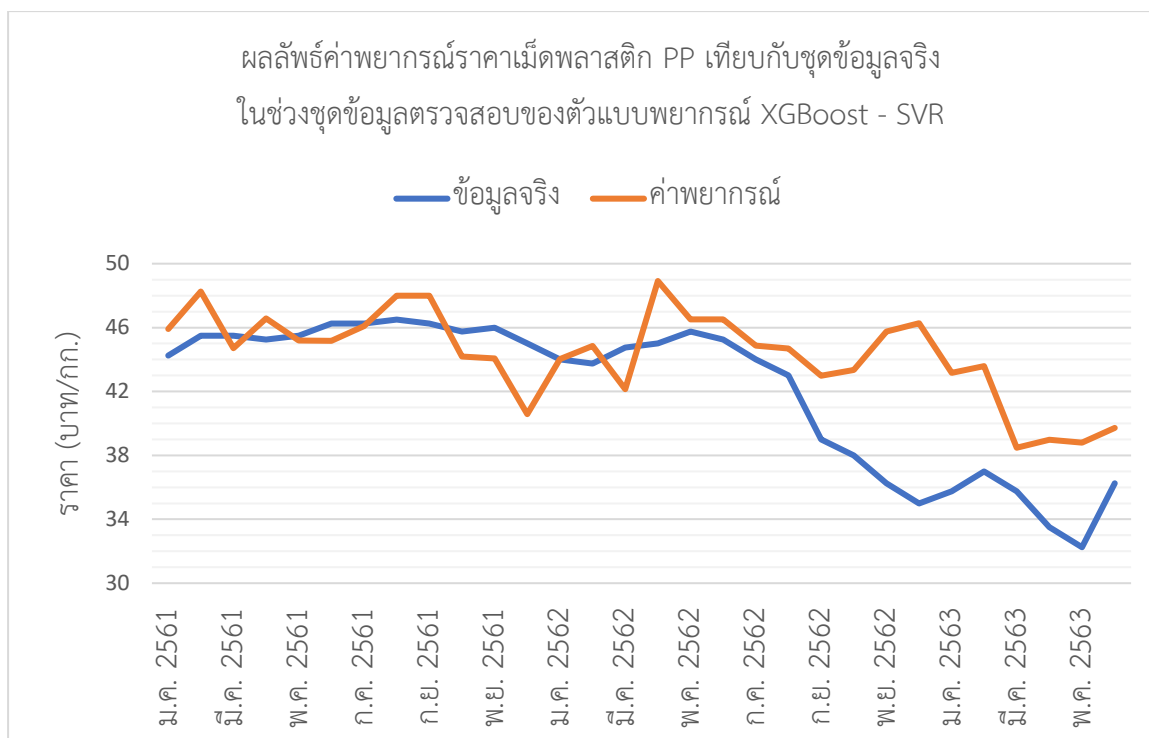
4.10.10 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ XGBoost - SVR

ตัวแบบพยากรณ์ XGBoost - SVR สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ XGBoost และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ SVR ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.7 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 28 โดยการพยากรณ์ของตัวแบบพยากรณ์ XGBoost - SVR ใช้เวลาทั้งสิ้น 6.50 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 8.16 % ซึ่ง

กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ XGBoost - SVR แสดงในรูปที่ 54

ตารางที่ 28 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ XGBoost - SVR และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	N_estimator	700
	Max Depth	8
	Learning Rate	0.4
	Subsample	0.7
	Gamma	0.8
ตัวแบบพยากรณ์รอง	Kernel	'rbf'
	C	1
	Gamma	1
	Degree	1
MAPE	8.16 %	



รูปที่ 54 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ XGBoost - SVR

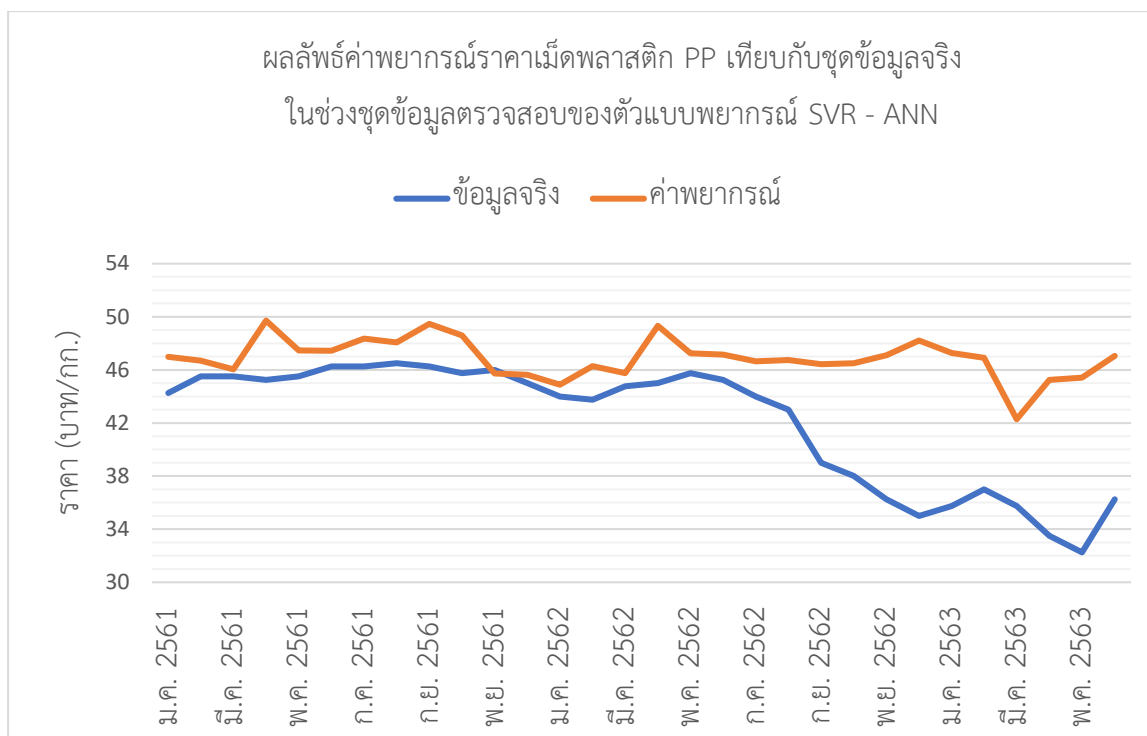
4.10.11 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SVR - ANN

ตัวแบบพยากรณ์ SVR - ANN สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ SVR และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ ANN ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.4 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 29 โดยการพยากรณ์ของตัวแบบพยากรณ์ SVR - ANN ใช้เวลาทั้งสิ้น 650 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 12.78 % ซึ่งกราฟแสดงผล

ลัพท์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ SVR - ANN แสดงในรูปที่ 55

ตารางที่ 29 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SVR - ANN และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	Kernel	'rbf'
	C	10
	Gamma	0.01
	Degree	1
ตัวแบบพยากรณ์รอง	Batch Size	32
	Epochs	100
	Activation	'tanh'
	Optimizer	'rmsprop'
	Hidden Units	100
	Dropout Rate	0.8
MAPE	12.78 %	



รูปที่ 55 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SVR - ANN

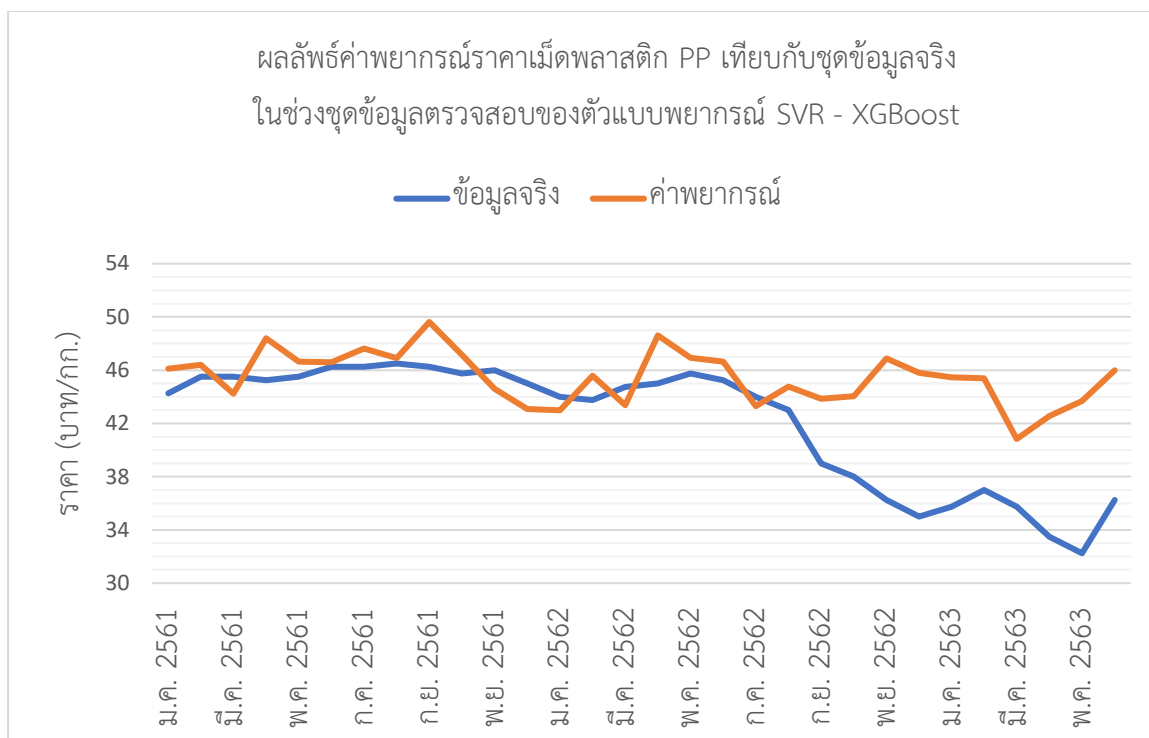
4.10.12 ผลลัพธ์การพยากรณ์ของตัวแบบพยากรณ์ SVR - XGBoost

ตัวแบบพยากรณ์ SVR - XGBoost สำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ประกอบไปด้วยตัวแบบพยากรณ์หลัก คือ ตัวแบบพยากรณ์ SVR และตัวแบบพยากรณ์รองคือ ตัวแบบพยากรณ์ XGBoost ในส่วนของตัวแบบพยากรณ์หลักจะทำการใช้ผลลัพธ์ที่ได้จากหัวข้อ 4.4 เช่นค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์ ค่าพยากรณ์ราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ และนำไปหาค่าความคลาดเคลื่อนในช่วงชุดข้อมูลฝึกสอนเพื่อนำไปใช้ในการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบโดยใช้ตัวแบบพยากรณ์รอง จากนั้นทำการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบพยากรณ์รองให้มีความเหมาะสมกับชุดข้อมูลฝึกสอนด้วยการใช้ Gridsearch CV ซึ่งจะได้ผลลัพธ์ค่าพารามิเตอร์ของตัวแบบพยากรณ์รองดังตารางที่ 30 โดยการพยากรณ์ของตัวแบบพยากรณ์ SVR - XGBoost ใช้เวลาทั้งสิ้น 63.4 นาที และผลพยากรณ์ราคาเม็ดพลาสติก PP ของตัวแบบผสมเกิดจากการนำค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก มารวมเข้ากับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองในช่วงชุดข้อมูลตรวจสอบ และเมื่อนำมาเทียบกับชุดข้อมูลตรวจสอบราคาเม็ดพลาสติก PP จะได้ค่า MAPE อยู่ที่ 10.38 % ซึ่งกราฟ

แสดงผลพีธค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบ (ตั้งแต่เดือนมกราคม พ.ศ. 2561 ถึงเดือนมิถุนายน พ.ศ. 2563) ของตัวแบบพยากรณ์ SVR - XGBoost แสดงในรูปที่ 56

ตารางที่ 30 ค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของตัวแบบพยากรณ์ SVR - XGBoost และค่า MAPE

ตัวแบบพยากรณ์	ไฮเปอร์พารามิเตอร์ / พารามิเตอร์	ค่าไฮเปอร์พารามิเตอร์ / พารามิเตอร์
ตัวแบบพยากรณ์หลัก	Kernel	'rbf'
	C	10
	Gamma	0.01
	Degree	1
ตัวแบบพยากรณ์รอง	N_estimator	400
	Max Depth	5
	Learning Rate	0.3
	Subsample	0.8
	Gamma	0.8
MAPE	10.38 %	



รูปที่ 56 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SVR – XGBoost

4.11 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์

จากผลลัพธ์ของตัวแบบพยากรณ์เดี่ยวในหัวข้อที่ 4.1 จนถึง 4.8 และตัวแบบพยากรณ์ผสมในหัวข้อ 4.10 ของราคาเม็ดพลาสติก PP ในประเทศไทย สามารถสรุปค่า MAPE รวมถึงเวลาที่ใช้ในการประมวลผลของแต่ละตัวแบบพยากรณ์ได้ดังตารางที่ 31 และ 32 โดยพบว่าตัวแบบพยากรณ์เดี่ยวที่มีความแม่นยำมากที่สุดคือตัวแบบพยากรณ์ ANN ที่มีค่า MAPE อยู่ที่ 7.71 % และในตัวแบบพยากรณ์ผสมที่มีความแม่นยำสูงที่สุดคือ ตัวแบบพยากรณ์ SARIMA – ANN ที่มีค่า MAPE อยู่ที่ 5.54 % ลดลงมาจากตัวแบบพยากรณ์เดี่ยว 2.17 % เป็นการบ่งชี้ว่าตัวแบบพยากรณ์ผสมมีความแม่นยำที่สูงกว่าตัวแบบพยากรณ์เดี่ยวในการพยากรณ์ราคาเม็ดพลาสติก PP ในประเทศไทย และเนื่องจากตัวแบบพยากรณ์ SARIMA – ANN มีระยะเวลาที่ใช้และขั้นตอนในการสร้างตัวแบบพยากรณ์ที่ไม่ซับซ้อนมาก ทำให้เหมาะสมกับการนำไปใช้งานจริงในการพยากรณ์ราคาเม็ดพลาสติก PP ในประเทศไทย

ตารางที่ 31 สรุปค่า MAPE ของแต่ละตัวแบบพยากรณ์ของราคาเม็ดพลาสติก PP ในประเทศไทย

ประเภทตัวแบบพยากรณ์	ตัวแบบพยากรณ์	เวลา (นาที)	MAPE
ตัวแบบพยากรณ์ ข้อมูลอนุกรมเวลา	ตัวแบบพยากรณ์ SARIMA	0.50	8.09 %
	ตัวแบบพยากรณ์ Holt-Winters	1.00	14.77 %
ตัวแบบพยากรณ์ข้อมูล เชิงสาเหตุ	ตัวแบบการพยากรณ์ถดถอยเชิงเส้น พหุคูณ	0.50	13.00 %
ตัวแบบพยากรณ์ด้วย เทคนิคการเรียนรู้ของ เครื่อง	ตัวแบบพยากรณ์ SVR	13.48	9.56 %
	ตัวแบบพยากรณ์ Decision Tree	1.55	16.94 %
	ตัวแบบพยากรณ์ Random Forest	35.30	14.10 %
	ตัวแบบพยากรณ์ XGBoost	106.50	8.16 %
	ตัวแบบพยากรณ์ ANN	580.45	7.71 %
ตัวแบบพยากรณ์ผสม	ตัวแบบพยากรณ์ SARIMA - SVR	5.30	7.18 %
	ตัวแบบพยากรณ์ SARIMA - XG	11.10	5.77 %
	ตัวแบบพยากรณ์ SARIMA - ANN	765.85	5.54 %
	ตัวแบบพยากรณ์ SVR - SARIMA	0.32	9.48 %
	ตัวแบบพยากรณ์ XG - SARIMA	0.20	8.16 %
	ตัวแบบพยากรณ์ ANN - SARIMA	0.50	7.38 %
	ตัวแบบพยากรณ์ ANN - XG	10.30	6.49 %
	ตัวแบบพยากรณ์ ANN - SVR	4.30	6.60 %
	ตัวแบบพยากรณ์ XG - ANN	530.20	7.87 %
	ตัวแบบพยากรณ์ XG - SVR	6.25	8.16 %
	ตัวแบบพยากรณ์ SVR - XG	65.40	10.38 %
	ตัวแบบพยากรณ์ SVR - ANN	650.00	12.38 %

ตารางที่ 32 เปรียบเทียบค่า MAPE ระหว่างตัวแบบพยากรณ์เดี่ยวและตัวแบบพยากรณ์ผสม

		ตัวแบบพยากรณ์รอง				
		เดี่ยว	SARIMA	SVR	XGBoost	ANN
ตัวแบบ พยากรณ์ หลัก	SARIMA	8.09%		7.18%	5.77%	5.54%
	SVR	9.56%	9.48%		10.38%	12.38%
	XGBoost	8.16%	8.16%	8.16%		7.87%
	ANN	7.71%	7.38%	6.60%	6.49%	

4.11.1 วิเคราะห์ผลลัพธ์การพยากรณ์ของกลุ่มตัวแบบพยากรณ์หลัก SARIMA

พิจารณาตารางที่ 33 พบว่าตัวแบบพยากรณ์ผสมที่มี SARIMA เป็นตัวแบบพยากรณ์หลักค่า MAPE ลดลงแบบมีนัยสำคัญเมื่อเทียบกับตัวแบบพยากรณ์เดี่ยว เกิดจากการที่ข้อมูลราคาเม็ดพลาสติก PP มีลักษณะแนวโน้มและฤดูกาลซึ่งทำให้ใช้งานได้ดีกับตัวแบบพยากรณ์ SARIMA และตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่องสามารถพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบที่ใกล้เคียงกับค่าความคลาดเคลื่อนที่ได้จากตัวแบบพยากรณ์หลัก ทำให้เมื่อนำค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองมารวมกับค่าพยากรณ์ราคาเม็ดพลาสติก PP จากตัวแบบพยากรณ์หลัก ทำให้ได้ผลลัพธ์ค่าพยากรณ์ของตัวแบบพยากรณ์ผสมที่มีความแม่นยำมากขึ้น

ตารางที่ 33 เปรียบเทียบผลลัพธ์การพยากรณ์ระหว่างกลุ่มตัวแบบพยากรณ์หลัก SARIMA และตัวแบบพยากรณ์เดี่ยว

		ตัวแบบพยากรณ์รอง			
		เดี่ยว	SVR	XGBoost	ANN
ตัวแบบพยากรณ์หลัก	SARIMA	8.09%	7.18%	5.77%	5.54%

4.11.2 วิเคราะห์ผลลัพธ์การพยากรณ์ของกลุ่มตัวแบบพยากรณ์รอง SARIMA

พิจารณาตารางที่ 34 พบว่าตัวแบบพยากรณ์ผสมที่มี SARIMA เป็นตัวแบบพยากรณ์รองค่า MAPE ไม่ได้ลดลงแบบมีนัยสำคัญเมื่อเทียบกับตัวแบบพยากรณ์เดี่ยวเนื่องจากข้อมูลค่าความคลาดเคลื่อนหรือ Residual ที่ตัวแบบพยากรณ์ SARIMA พยากรณ์ไม่มีรูปแบบแนวโน้มและฤดูกาล แต่ตัวแบบพยากรณ์ SARIMA เหมาะกับการใช้ในข้อมูลที่มีแนวโน้มและฤดูกาล

ตารางที่ 34 เปรียบเทียบผลลัพธ์การพยากรณ์ระหว่างกลุ่มตัวแบบพยากรณ์รอง SARIMA และตัวแบบพยากรณ์เดี่ยว

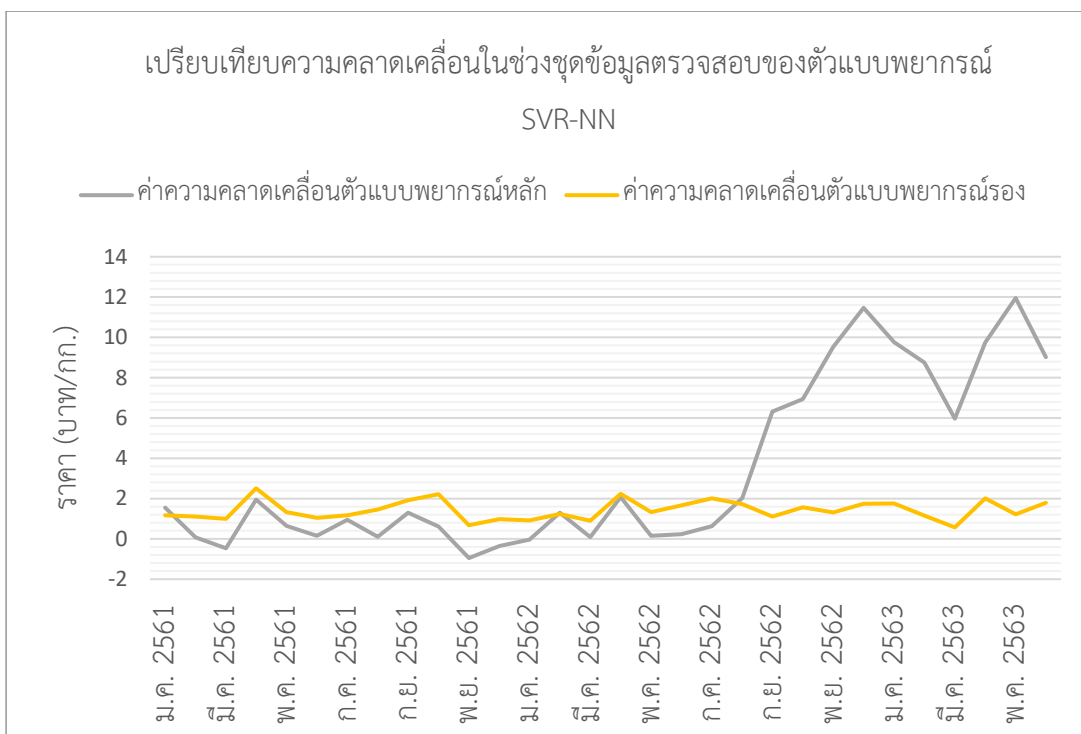
		ตัวแบบพยากรณ์รอง	
		เดี่ยว	SARIMA
ตัวแบบพยากรณ์หลัก	SVR	9.56%	9.48%
	XGBoost	8.16%	8.16%
	ANN	7.71%	7.38%

4.11.3 วิเคราะห์ผลลัพธ์การพยากรณ์ของกลุ่มตัวแบบพยากรณ์หลัก และตัวแบบพยากรณ์รองจากตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่อง

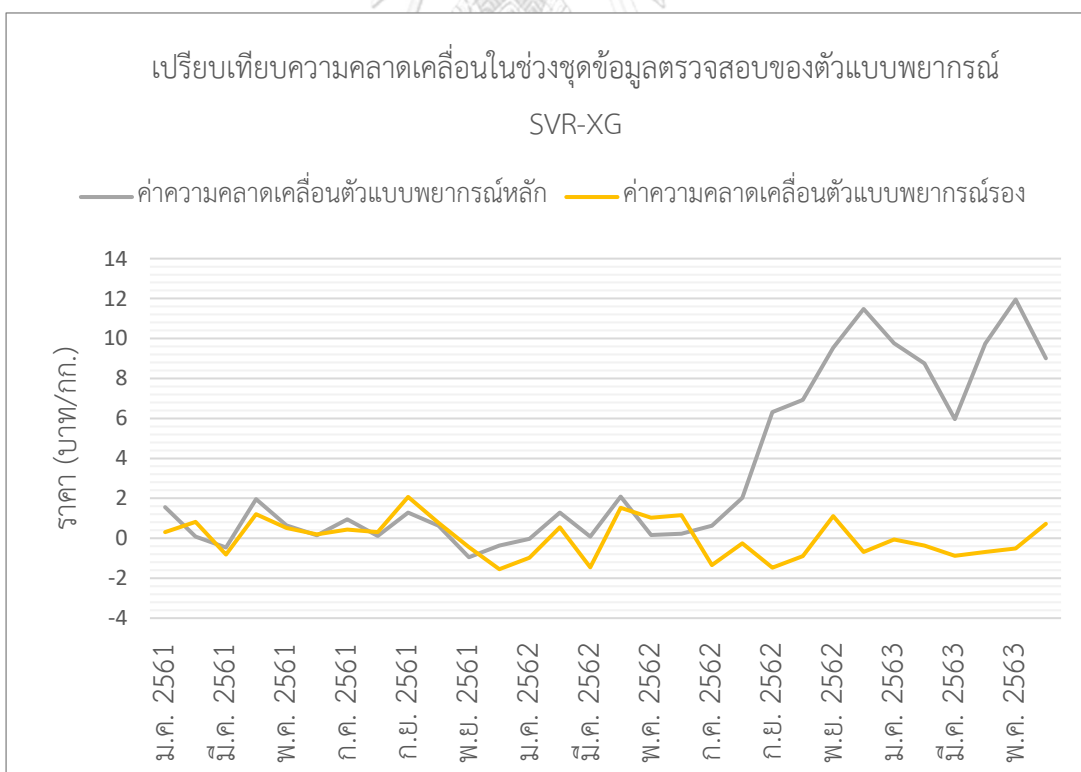
พิจารณารางที่ 35 พบว่าตัวแบบพยากรณ์ผสมที่มีตัวแบบพยากรณ์หลักคือ SVR จะมีค่า MAPE สูงกว่าตัวแบบพยากรณ์เดี่ยว SVR ตรงกันข้ามกับตัวแบบพยากรณ์ผสมที่มีตัวแบบพยากรณ์หลักเป็น XGBoost และ ANN ที่จะมีค่า MAPE ต่ำกว่าตัวแบบพยากรณ์เดี่ยว ในกรณีดังกล่าวมีสาเหตุมาจากการที่ตัวแบบพยากรณ์รอง XGBoost และ ANN ทำการพยากรณ์ค่าความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบและค่าพยากรณ์ความคลาดเคลื่อนที่ได้เป็นการเสริมค่าความคลาดเคลื่อนจากตัวแบบพยากรณ์ ดังรูปที่ 57 และ 58 ที่ค่าความคลาดเคลื่อนของตัวแบบพยากรณ์รองมีค่ามากกว่า 0 (หากมีค่าลบจะเป็นการหักล้างกับค่าความคลาดเคลื่อนจากตัวแบบพยากรณ์หลักซึ่งจะทำให้ผลพยากรณ์ตัวแบบพยากรณ์ผสมมีความแม่นยำมากขึ้น) ทำให้เมื่อนำค่าพยากรณ์ราคาเม็ดพลาสติกจากตัวแบบพยากรณ์หลัก มารวมกับค่าพยากรณ์ความคลาดเคลื่อนจากตัวแบบพยากรณ์รองเป็นค่าพยากรณ์ราคาเม็ดพลาสติกของตัวแบบผสมมีค่าห่างจากข้อมูลราคาเม็ดพลาสติก PP จริงในช่วงชุดข้อมูลตรวจสอบมากขึ้น

ตารางที่ 35 เปรียบเทียบผลลัพธ์การพยากรณ์ของกลุ่มตัวแบบพยากรณ์หลัก และตัวแบบพยากรณ์รองจากตัวแบบพยากรณ์การเรียนรู้ด้วยเครื่อง

		ตัวแบบพยากรณ์รอง			
		เดี่ยว	SVR	XGBoost	ANN
ตัวแบบพยากรณ์หลัก	SVR	9.56%		10.38%	12.38%
	XGBoost	8.16%	8.16%		7.87%
	ANN	7.71%	6.60%	6.49%	



รูปที่ 57 เปรียบเทียบความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SVR-NN



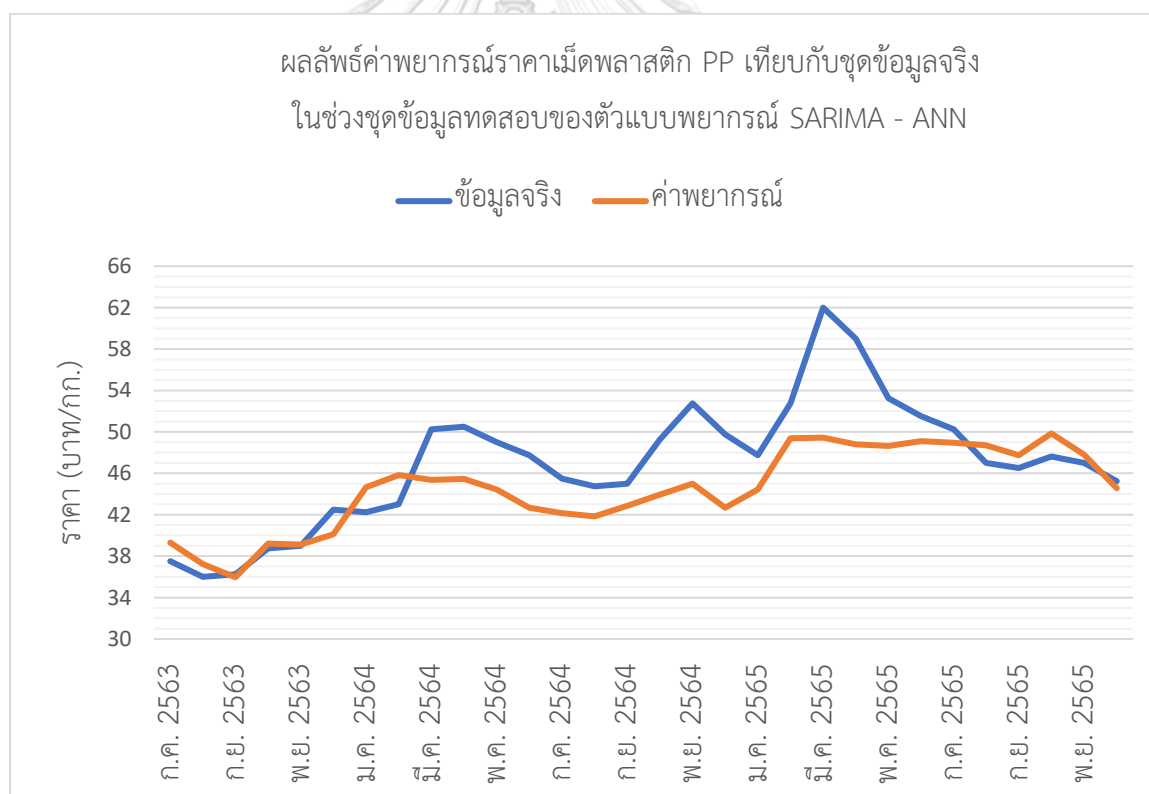
รูปที่ 58 เปรียบเทียบความคลาดเคลื่อนในช่วงชุดข้อมูลตรวจสอบของตัวแบบพยากรณ์ SVR-XG

4.12 ผลลัพธ์การพยากรณ์ในชุดข้อมูลทดสอบ (Testing Set)

จากหัวข้อที่ 4.11 จะเป็นการคัดเลือกตัวแบบพยากรณ์ที่มีความเหมาะสมที่สุดเพื่อนำมาวัดความแม่นยำกับชุดข้อมูลทดสอบ (Testing Set) โดยตัวแบบพยากรณ์ดังกล่าวคือ ตัวแบบพยากรณ์ SARIMA – ANN เมื่อนำมาวัดความแม่นยำกับชุดข้อมูลทดสอบจะได้ผลดังในตารางที่ 36 ซึ่งกราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลทดสอบ (ตั้งแต่เดือนกรกฎาคม พ.ศ. 2563 ถึงเดือนธันวาคม พ.ศ. 2565) ของตัวแบบพยากรณ์ SARIMA - ANN แสดงในรูปที่ 59

ตารางที่ 36 ผลลัพธ์ค่าความแม่นยำของตัวแบบพยากรณ์ SARIMA – ANN

ตัวแบบพยากรณ์	MAPE ในช่วงชุดข้อมูลเปรียบเทียบ	MAPE ในช่วงชุดข้อมูลทดสอบ
SARIMA - ANN	5.54 %	6.92 %



รูปที่ 59 กราฟแสดงผลลัพธ์ค่าพยากรณ์ราคาเม็ดพลาสติก PP เทียบกับชุดข้อมูลจริงในช่วงชุดข้อมูลทดสอบของตัวแบบพยากรณ์ SARIMA - ANN

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยฉบับนี้ได้ทำการศึกษาตัวแบบพยากรณ์ที่เหมาะสมสำหรับการพยากรณ์ราคาเม็ดพลาสติก PP ในประเทศไทย ด้วยกลุ่มตัวแบบพยากรณ์อนุกรมเวลา 2 ตัวแบบพยากรณ์ ได้แก่ ตัวแบบพยากรณ์ SARIMA ตัวแบบพยากรณ์ Holt-Winters กลุ่มตัวแบบพยากรณ์เชิงสาเหตุ 1 ตัวแบบพยากรณ์ คือ ตัวแบบพยากรณ์การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ กลุ่มตัวแบบพยากรณ์ด้วยเทคนิคการเรียนรู้ของเครื่อง 5 ตัวแบบพยากรณ์ ได้แก่ ตัวแบบพยากรณ์ SVR ตัวแบบพยากรณ์ Decision Tree ตัวแบบพยากรณ์ Random Forest ตัวแบบพยากรณ์ XGBoost ตัวแบบพยากรณ์ ANN และตัวแบบพยากรณ์ผสมจากการคัดเลือกตัวแบบพยากรณ์เดี่ยวที่มีความแม่นยำสูงสุดจากตัวแบบพยากรณ์อนุกรมเวลา ตัวแบบพยากรณ์เชิงสาเหตุ จับคู่กับตัวแบบพยากรณ์เดี่ยวที่มีความแม่นยำสูงสุด 3 อันดับแรกในกลุ่มตัวแบบพยากรณ์ด้วยเทคนิคการเรียนรู้ ทำให้ได้ตัวแบบพยากรณ์ผสมทั้งสิ้น 12 ตัวแบบพยากรณ์ รวมทั้งสิ้น 20 ตัวแบบพยากรณ์ จากนั้นจะนำมาเปรียบเทียบความแม่นยำของแต่ละตัวแบบพยากรณ์ด้วยการใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE)

5.2 สรุปผลการวิจัยตัวแบบพยากรณ์อนุกรมเวลา

ตัวแบบพยากรณ์อนุกรมเวลาเมื่อนำมาพยากรณ์ราคาเม็ดพลาสติก PP ในประเทศไทยพบว่าตัวแบบพยากรณ์ SARIMA มีความแม่นยำสูงสุด โดยมีค่า MAPE อยู่ที่ 8.09% รองลงมาคือตัวแบบพยากรณ์ Holt-Winters มีค่า MAPE อยู่ที่ 14.77% เมื่อพิจารณาที่กราฟแสดงการเปรียบเทียบระหว่างค่าพยากรณ์ราคาเม็ดพลาสติกกับชุดข้อมูลตรวจสอบของทั้ง 2 ตัวแบบพยากรณ์พบว่าตัวแบบพยากรณ์ Holt-Winters ค่าพยากรณ์ส่วนมากมีค่าต่ำกว่าค่าจริงโดยเฉพาะตั้งแต่ช่วงแรกจนถึงช่วงกลางของกราฟตามรูปที่ 4-1 แสดงให้เห็นว่าตัวแบบพยากรณ์ไม่สามารถตรวจจับลักษณะของข้อมูลได้ดีทำให้ผลค่าพยากรณ์ที่ได้ไม่แม่นยำ แต่ในทางกลับกันตัวแบบพยากรณ์ SARIMA เมื่อพิจารณาจากกราฟในรูปที่ 4-5 พบว่าตัวแบบพยากรณ์สามารถตรวจจับลักษณะของข้อมูลได้ดีกว่าส่งผลให้ค่าพยากรณ์ที่ได้มีความแม่นยำสูงกว่า แม้ว่าในช่วงท้ายของกราฟค่าพยากรณ์ที่ได้เริ่มห่างจากค่าข้อมูลจริง สาเหตุเพราะเป็นช่วงที่เกิดวิกฤตโรคระบาดโควิด 19 ส่งผลให้ราคาเม็ดพลาสติกมีการปรับตัวลงผิดปกติจากอดีต

5.3 สรุปผลการวิจัยตัวแบบพยากรณ์เชิงสาเหตุ

ตัวแบบพยากรณ์เชิงสาเหตุในงานวิจัยฉบับนี้คือ ตัวแบบพยากรณ์การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยได้ทำการรวบรวมปัจจัยที่ส่งผลต่อราคาเม็ดพลาสติกจากการอ้างอิงงานวิจัยที่เกี่ยวข้องได้แก่ ดัชนีการผลิตเม็ดพลาสติก PP สถิติการส่งออกและการนำเข้าของโพลิโพรพิลีนชั้นปฐมภูมิ ราคาน้ำมันดิบ Brent และ WTI ดัชนีราคาผู้บริโภค (CPI) ดัชนีราคาผู้ผลิต (PPI) และ สถิติการผลิตรถยนต์และรถจักรยานยนต์ รวมเป็น 9 ปัจจัย เมื่อทำการคัดเลือกปัจจัยด้วยวิธีการถดถอยแบบขั้นตอนที่ระดับความเชื่อมั่น 95 % จะเหลือปัจจัยที่มีความสัมพันธ์ในทิศทางเดียวกันกับราคาเม็ดพลาสติก PP ได้แก่ ค่าดัชนีราคาผู้บริโภค (CPI) ราคาน้ำมันดิบ Brent และปัจจัยที่มีความสัมพันธ์ในทิศทางตรงข้ามกับราคาเม็ดพลาสติก PP ได้แก่ ค่าดัชนีราคาผู้ผลิต (PPI) จำนวนรถยนต์ที่ถูกผลิตเมื่อนำมาสร้างสมการถดถอยพหุคูณเชิงเส้น แล้วทำการพยากรณ์เพื่อไปทำการเปรียบเทียบความแม่นยำจากข้อมูลราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ จะได้ผลลัพธ์ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเท่ากับ 20.40%

5.4 สรุปผลการวิจัยตัวแบบพยากรณ์เทคนิคการเรียนรู้ด้วยเครื่อง

ตัวแบบพยากรณ์เทคนิคการเรียนรู้ของเครื่องที่ใช้ในการพยากรณ์ราคาเม็ดพลาสติก PP ในประเทศไทยของงานวิจัยฉบับนี้ ได้แก่ ตัวแบบพยากรณ์ SVR ตัวแบบพยากรณ์ Decision Tree ตัวแบบพยากรณ์ Random Forest ตัวแบบพยากรณ์ XGBoost ตัวแบบพยากรณ์ ANN โดยเมื่อทำการพยากรณ์เพื่อไปทำการเปรียบเทียบความแม่นยำจากข้อมูลราคาเม็ดพลาสติก PP ในช่วงชุดข้อมูลตรวจสอบ จะได้ผลลัพธ์ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเท่ากับ 9.56% 16.94% 14.10% 8.16% 7.71% ตามลำดับ และสามารถสรุปได้ว่าตัวแบบพยากรณ์ ANN มีความแม่นยำสูงที่สุดในกลุ่มตัวแบบพยากรณ์เดี่ยว

5.5 สรุปผลการวิจัยตัวแบบพยากรณ์แบบผสม

ตัวแบบพยากรณ์แบบผสมที่ใช้ในการพยากรณ์ราคาเม็ดพลาสติก PP ในประเทศไทยของงานวิจัยฉบับนี้ เกิดจากการนำตัวแบบเดี่ยวที่ได้จากการคัดเลือกตัวแบบพยากรณ์เทคนิคการเรียนรู้ด้วยเครื่องที่มีความแม่นยำสูงที่สุด 3 อันดับแรก จับคู่กับตัวแบบพยากรณ์ SARIMA หรือจับคู่กับตัวแบบพยากรณ์เทคนิคการเรียนรู้ด้วยเครื่อง ทำให้ได้ตัวแบบพยากรณ์ทั้งสิ้น 12 ตัวแบบพยากรณ์ ผลลัพธ์พบว่าตัวแบบพยากรณ์ SARIMA – ANN ที่เกิดจากตัวแบบพยากรณ์หลัก SARIMA จับคู่กับตัวแบบพยากรณ์รอง ANN มีความแม่นยำสูงที่สุดโดยมีค่า MAPE อยู่ที่ 5.54% และจัดว่าเป็นตัวแบบพยากรณ์ที่มีความแม่นยำสูงที่สุดเมื่อทำการเปรียบเทียบกับข้อมูลชุดตรวจสอบในงานวิจัยฉบับนี้ และ

เมื่อนำตัวแบบพยากรณ์ SARIMA – ANN ไปทำการพยากรณ์ในช่วงชุดข้อมูลทดสอบ และทำการเปรียบเทียบระหว่างค่าพยากรณ์และค่าจริงในช่วงชุดข้อมูลทดสอบพบว่าตัวแบบพยากรณ์ให้ผลลัพธ์ค่าความผิดพลาดทั่วไป (Generalized Error) เท่ากับ 6.92%

5.6 ข้อเสนอแนะเพิ่มเติม

ปัจจัยที่นำมาพิจารณาเพื่อสร้างตัวแบบในการพยากรณ์ราคาเม็ดพลาสติก PP มีอยู่ 9 ตัวแต่ยังมีบางปัจจัยที่ต้องนำมาพิจารณาด้วยแต่เนื่องจากไม่สามารถหาข้อมูลปัจจัยนั้นได้เพราะทางหน่วยงานไม่ได้เก็บข้อมูลเป็นรายเดือน ซึ่งปัจจัยดังกล่าวได้แก่ ปริมาณการผลิตเม็ดพลาสติก และปริมาณการใช้งานเม็ดพลาสติก

ช่วงข้อมูลเปรียบเทียบเนื่องจากการเกิดโรคระบาดโควิด 19 ทำให้ข้อมูลราคาเม็ดพลาสติก PP ในประเทศไทยเกิดการผันผวนมีทิศทางปรับตัวลงผิดไปจากปกติ จากการที่เกิดจากปัญหาด้านเศรษฐกิจ การส่งออกสินค้า การสั่งซื้อสินค้าที่ลดลง หรือราคาน้ำมันที่ผันผวนเช่นกัน ส่งผลให้ค่าพยากรณ์จากตัวแบบพยากรณ์มีความแม่นยำลดลง ถ้าหากตัดข้อมูลชุดเปรียบเทียบในช่วงที่เกิดวิกฤตโรคระบาดโควิด 19 ออกไป จะได้ความแม่นยำที่เพิ่มขึ้น 2 เท่าในบางตัวแบบพยากรณ์

รูปแบบการพยากรณ์สามารถทำในรูปแบบรายไตรมาสเพื่อตรวจสอบว่าสามารถทำให้ผลลัพธ์การพยากรณ์มีความแม่นยำมากขึ้นหรือไม่ เพราะมีหลายข้อมูลปัจจัยที่ส่งผลต่อราคาเม็ดพลาสติก PP แต่ถูกเก็บบันทึกในรูปแบบรายไตรมาสในหน่วยงานรัฐ เช่น ข้อมูล GDP ของประเทศไทย เป็นต้น

การนำตัวแบบพยากรณ์ราคาเม็ดพลาสติก PP ของวิทยานิพนธ์ฉบับนี้ไปทำการใช้งานจริง เนื่องจากการพยากรณ์รายเดือนจึงไม่จำเป็นต้องดูผลพยากรณ์ตลอดเวลา ทำให้สามารถใช้วิธีพยากรณ์ที่ใช้เวลานานอย่าง ANN ได้

วิทยานิพนธ์ฉบับนี้มีข้อจำกัดในการพยากรณ์ค่าราคาเม็ดพลาสติก PP โดยต้องรู้ค่าข้อมูลปัจจัยมาก่อน ดังนั้นก่อนที่จะทำการพยากรณ์ราคาเม็ดพลาสติก PP ต้องทราบค่าพยากรณ์ของแต่ละปัจจัยก่อน ซึ่งอาจทำให้ผลลัพธ์การพยากรณ์ราคาเม็ดพลาสติก PP จะมีความแม่นยำลดลง

บรรณานุกรม



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

- Ahmed, R. A., & Shabri, A. B. (2014). Daily crude oil price forecasting model using arima, generalized autoregressive conditional heteroscedastic and support vector machines. *American Journal of Applied Sciences*, 11(3), 425.
- Amazon Web Services. (n.d.). XGBoost: How It Works. AWS Documentation. Retrieved June 17, 2023, from <https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost-HowItWorks.html>
- BigML. (2017, March 14). Introduction to Boosted Trees. BigML Blog. Retrieved from <https://blog.bigml.com/2017/03/14/introduction-to-boosted-trees>
- Chai, J., Xing, L. M., Zhou, X. Y., Zhang, Z. G., & Li, J. X. (2018). Forecasting the WTI crude oil price by a hybrid-refined method. *Energy Economics*, 71, 114-127.
- Chintakowit, N. (2007). Improvement of demand forecasting system: case study in PVC leather and plastic company (Doctoral dissertation, Chulalongkorn University).
- Géron, A. (2019). *Hands-on Machine Learning with Scikit-Learn Keras and TensorFlow* (2nd ed.). Sebastopol, CA: O'Reilly Media, Inc.
- Hanke, J. E., & Wichern, D. W. (2009). *Business forecasting* (9th ed.). Upper Saddle River, N J: Pearson Education.
- Haque, M. I., & Shaik, A. R. (2021). Predicting Crude Oil Prices During a Pandemic: A Comparison of Arima and Garch Models. *Montenegrin Journal of Economics*, 17(1), 197-207.
- Hurwitz, J., & Kirsch, D. (2018). *Machine learning for dummies*. IBM Limited Edition, 75.
- Hyndman, R. J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice* (2nd ed.). Melbourne, Australia: OTexts. [OTexts.com/fpp2](https://otexts.com/fpp2). Accessed on 2021, Oct 10.

- IBM. (n.d.). Random Forest - Overview. IBM. Retrieved June 17, 2023, from <https://www.ibm.com/topics/random-forest>
- IBM. (n.d.). Neural Networks. IBM. Retrieved June 17, 2023, from <https://www.ibm.com/topics/neural-networks>
- Jong, L. J., Ismail, S., Mustapha, A., Abd Wahab, M. H., & Idrus, S. Z. S. (2020, September). The combination of autoregressive integrated moving average (arima) and support vector machines (svm) for daily rubber price forecasting. In IOP Conference Series: Materials Science and Engineering (Vol. 917, No. 1, p. 012044). IOP Publishing.
- Kalinggo, B. A., & Zulkarnain. (2020, June). Time series forecasting for non-stationary data: A case study of petrochemical product price. In Proceedings of the 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering (pp. 86-92).
- Keerthan, J. S., Nagasai, Y., & Shaik, S. (2019). Machine Learning Algorithms for Oil Price Prediction. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(8), 958-963.
- Khan, P. W., Byun, Y. C., Lee, S. J., Kang, D. H., Kang, J. Y., & Park, H. S. (2020). Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. Energies, 13(18), 4870.
- Li, J. (2020). Using Hybrid Method Based on Machine Learning for Energy Consumption Prediction of Oil and Gas Production. Recent Developments in Intelligent Computing, Communication and Devices: Proceedings of ICCD 2019, 1185, 234.
- Manickavasagam, J., Visalakshmi, S., & Apergis, N. (2020). A novel hybrid approach to forecast crude oil futures using intraday data. Technological Forecasting and Social Change, 158, 120126.

- Nowneow, A., & Rungreunganun, V. (2018). Poly Vinyl Chloride Pellet Price Forecasting Using Arima Model. *International Journal of Mechanical Engineering and Technology (IJMET)*, 8(13), 224-232.
- Shiva, J. K., Nagasai, Y., & Subhani, S. (2019). Machine Learning Algorithms for Oil Price Prediction. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(8), 958-963.
- Suppalakpanya, K., Nikhom, R., Booranawong, T., & Booranawong, A. (2019). Forecasting oil palm and crude palm oil data in Thailand using exponential time-series methods. *Engineering and Applied Science Research*, 46(1), 44-55.
- VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media.
- Zahari, F. Z., Khalid, K., Roslan, R., Sufahani, S., Mohamad, M., Rusiman, M. S., & Ali, M. (2018, April). Forecasting natural rubber price In Malaysia using ARIMA. In *Journal of Physics: Conference Series* (Vol. 995, No. 1, p. 012013). IOP Publishing.
- ปฐุมิพพร อุดม. การ พยการณั ยอด ขาย สำนหรั บ ผู้ แทน จำหน่ย เม็ด พลาสติค (Doctoral dissertation, จุฬาลงกรณั มหาวิทยาลัย).
- วรรณิภา ลูฐิตะสัถธาวารกุล, พรพิชยา กุวลั ยรัถนั. การศึ กษารู ปแบบการ พยการณั ราคาเม็ดพลาสติคพอลิเอทิลีน.; 2009.
- ศิริเทพ จันทรบุญญแก้ว. การ พยการณั ปริมาณสายโทรเข้าสำหรั บศูนย์บริการลูกค้าธนาคารพานิษั (Doctoral dissertation, จุฬาลงกรณั มหาวิทยาลัย).
- อัจฉรา จันวดี. การปรึ บปรุงการ พยการณั ความต้งการวัตุดิบและนโยบายการเติมเต็มวัตุดิบคงคลังสำหรั บโรงงานผลิตขึ้นรูปลเหล็กหล่อแบบออกแบบตามคำสั่งซื้อ (Doctoral dissertation, จุฬาลงกรณั มหาวิทยาลัย).

ประวัติผู้เขียน

ชื่อ-สกุล	ภัทรพล ภัทรอัมพรชัย
วัน เดือน ปี เกิด	29 สิงหาคม 2538
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	วิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมเครื่องกล มหาวิทยาลัยเกษตรศาสตร์ ปัจจุบันกำลังศึกษาต่อระดับปริญญาโท คณะวิศวกรรมศาสตร์ สาขา วิศวกรรมอุตสาหกรรม จุฬาลงกรณ์มหาวิทยาลัย
ที่อยู่ปัจจุบัน	68/173 ซอยรามคำแหง 164 ถนนรามคำแหง หมู่บ้านเพอร์เฟคเพลส เขต มีนบุรี แขวงมีนบุรี กทม 10510



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY