

การพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและ
การทดสอบประสิทธิภาพของเซอร์วิส เพื่อเป็นแนวทางในการ
วางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ

Forecasting Calling the Applications through API and
the Performance of Services to Guide Server Cost
Management



สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตาม
หลักสูตรปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Forecasting Calling the Applications through API and the Performance of Services to Guide Server Cost Management

Lalitpat Jinaprom

A COOPARATIVE EDUCATION SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS)
DEPARTMENT OF STATISTICS, SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2022

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา การพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ
Forecasting Calling the Applications through API and the Performance of Services to Guide Server Cost Management

ชื่อนักศึกษา นางสาวลลิตภัทร จินะพรหม รหัสนักศึกษา 62050825




ปริญญา วิทยาศาสตร์บัณฑิต (สถิติประยุกต์)

ภาควิชา สถิติ

ปีการศึกษา 2565

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.พรพิมล ชัยวุฒิศักดิ์

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตร์บัณฑิต (สถิติประยุกต์) ประจำปีการศึกษา 2565

คณะกรรมการสอบ	ลายมือชื่อ
ดร.สุกัญญา ศรีอินมัย ประธานกรรมการ	
คุณสุพิชฌาย์ แสงชาติ กรรมการ	
ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์ กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	การพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ Forecasting Calling the Applications through API and the Performance of Services to Guide Server Cost Management
ชื่อนักศึกษา	นางสาวลลิตภัทร จินะพรม รหัสนักศึกษา 62050825
ปริญญา	วิทยาศาสตร์บัณฑิต (สถิติประยุกต์)
ภาควิชา	สถิติ
คณะ	วิทยาศาสตร์
มหาวิทยาลัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา	2565
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.พรพิมล ชัยวุฒิศักดิ์

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพและเปรียบเทียบความถูกต้องของวิธีการพยากรณ์ โดยวิธีการพยากรณ์ที่ทำการศึกษามี 4 วิธี ได้แก่ 1) วิธีบ็อกซ์-เจนกินส์ได้แก่ตัวแบบ SARIMA 2) วิธีบ็อกซ์-เจนกินส์ได้แก่ตัวแบบ SARIMAX 3) วิธีการเรียนรู้ของเครื่อง ได้แก่ตัวแบบ Random Forest และ 4) วิธีการเรียนรู้เชิงลึก ได้แก่ตัวแบบ LSTM ข้อมูลที่นำมาใช้ในการศึกษาเป็นข้อมูลจำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันรายวัน ของธนาคารพาณิชย์แห่งหนึ่ง ซึ่งแบ่งข้อมูลเป็น 2 ชุด คือ ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2564 ถึง เดือนธันวาคม พ.ศ. 2565 เพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุด และ ข้อมูลตั้งแต่เดือนมกราคม พ.ศ.2566 ถึง เดือนมีนาคม พ.ศ.2566 เพื่อเปรียบเทียบค่าความคลาดเคลื่อนของตัวแบบพยากรณ์ที่ได้ โดยเกณฑ์ที่ใช้ในการเปรียบเทียบความคลาดเคลื่อนของการพยากรณ์คือค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) ผลการศึกษาพบว่า ตัวแบบที่มีค่า MAPE ต่ำสุดคือตัวแบบ Random Forest ซึ่งมีค่า MAPE เท่ากับ 4.8174% ซึ่งน้อยกว่าตัวแบบ SARIMAX ซึ่งมีค่า MAPE เท่ากับ 4.9051%, ตัวแบบ LSTM ซึ่งมีค่า MAPE เท่ากับ 5.0942% และตัวแบบ SARIMA ซึ่งมีค่า MAPE เท่ากับ 5.3796% โดยตัวแบบ แบบ Random Forest ที่ได้ถูกนำไปพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ต่อไป

คำสำคัญ : การวิเคราะห์อนุกรมเวลา, วิธีบ็อกซ์-เจนกินส์, วิธีการเรียนรู้ของเครื่อง, วิธีการเรียนรู้เชิงลึก, จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชัน, การทดสอบประสิทธิภาพของเซิร์ฟเวอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Forecasting Calling the Applications through API and the Performance of Services to Guide Server Cost Management
Students	Miss. Lalitpat Jinaprom Student ID 62050825
Degree	Bachelor of Science (Applied Statistics)
Department	Statistics
School	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2022
Advisor	Assistant Professor Dr.Pornpimol Chaiwuttisak

Abstract

This research aimed to forecast the number of calling applications through API and the performance tests of service to support efficient server cost management and compare the performance of different forecasting methods. Four forecasting methods were considered in the study: SARIMA, SARIMAX, Random Forest, and LSTM. The data used in this study was the number of daily transaction service calls of the commercial bank. It was split into 2 datasets. The first dataset containing from January 2021 to December 2022 was used for selecting the most suitable model, and the second dataset containing from January 2023 to March 2023 was used to test the performance of the predicted model. The criterion used to compare the error of forecast models was the Mean Absolute Percentage Error (MAPE). Consequently, the lowest MAPE achieved by Random Forest was 4.8174%, while MAPEs achieved by SARIMAX, LSTM, and SARIMA were 4.9051%, 5.0942%, and 5.3796% respectively. It indicates that the Random Forest model was the most accurate to forecast the number of application service calls and service performance tests.

Keywords: Time Series Analysis, Box-Jenkins Method, Machine Learning, Deep Learning, the Number of Calling the Applications through API, Performance Tests of Services

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

สหกิจศึกษาฉบับนี้สำเร็จได้ เนื่องจากได้รับความเมตตา ความอนุเคราะห์จากอาจารย์ที่
ปรึกษาสหกิจศึกษา ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์ ที่เมตตาให้คำปรึกษา สละเวลาอันมีค่าให้คำแนะนำ
 อีกทั้งยังช่วยแก้ไขข้อบกพร่อง ต่าง ๆ ด้วยความเอาใจใส่ จนกระทั่งสหกิจศึกษาสำเร็จลุล่วงไปด้วยดี
 ผู้วิจัยทราบซึ่งในความเมตตาจากท่านอาจารย์และกราบขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้

ผู้วิจัยขอขอบพระคุณ ดร.สกุณา ศรีอินมัย ที่ให้เกียรติเป็นประธานกรรมการสอบสหกิจที่
เมตตาให้คำปรึกษา และสละเวลามาตรวจทานแก้ไขข้อบกพร่องต่าง ๆ ตลอดการทำสหกิจศึกษาฉบับ
นี้ อีกทั้งคุณอาจารย์ประจำภาคสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร
ลาดกระบัง (สจล.) ที่ให้คำแนะนำที่เป็นประโยชน์แก่ผู้วิจัย มอบวิชาความรู้ และประสบการณ์การ
เรียนการสอนกับผู้วิจัย

ขอขอบพระคุณ คุณสุพิชฌาย์ แสงชาติ ที่ให้เกียรติเป็นกรรมการสอบสหกิจในครั้งนี้ และ
ขอขอบคุณบุคลากรของธนาคารพาณิชย์แห่งหนึ่งที่ทำให้การดูแล ให้คำปรึกษารวมถึงให้ความช่วยเหลือ
ในเรื่องต่างๆ เสมอมา รวมถึงขอขอบคุณธนาคารพาณิชย์แห่งหนึ่งที่ได้อนุญาตให้นำข้อมูลมาใช้งานใน
การทำสหกิจศึกษาฉบับนี้

สุดท้ายนี้ผู้วิจัยขอขอบพระคุณบิดา มารดา ที่สนับสนุนให้กำลังใจแก่ผู้วิจัยในการศึกษาระดับ
ปริญญาตรี ตลอดจนการทำสหกิจศึกษาครั้งนี้สำเร็จลุล่วงไปด้วยดี ผู้วิจัยหวังเป็นอย่างยิ่งว่าสหกิจ
ศึกษานี้จะเป็นประโยชน์และจะเป็นแนวทางในการพัฒนาต่อไป

ลลิตภัทร จินะพรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ก
บทคัดย่อภาษาอังกฤษ.....	ค
กิตติกรรมประกาศ.....	ง
สารบัญ.....	จ
สารบัญตาราง.....	ช
สารบัญรูป.....	ฉ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	2
1.3 ขอบเขตของงานวิจัย.....	3
1.3.1 ขอบเขตด้านข้อมูล.....	3
1.3.2 ขอบเขตด้านเวลา.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.5 นิยามศัพท์เฉพาะ.....	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	5
2.1 ทฤษฎีที่เกี่ยวข้อง.....	5
2.1.1 ความหมายของการพยากรณ์.....	5
2.1.2 เทคนิควิธีการพยากรณ์.....	5
2.1.3 ขั้นตอนในการพยากรณ์.....	6
2.1.4 การพยากรณ์วิธีบ็อกซ์-เจนกินส์ (Box and Jenkins Method).....	7
2.1.4.1 การทดสอบความเป็นสแตชันนารีของข้อมูล (Stationary).....	9
2.1.4.2 การทดสอบแนวโน้มและฤดูกาล.....	11
2.1.4.3 การแปลงอนุกรมเวลาให้เป็นสแตชันนารีด้วยวิธีหาผลต่าง (Differencing)	13
2.1.4.4 ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง บางส่วน.....	14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.1.4.5 ตัวแบบพยากรณ์ด้วยวิธีBox-Jenkins สำหรับอนุกรมเวลา.....	15
2.1.4.6 ขั้นตอนการพยากรณ์โดยวิธีของ Box-Jenkins.....	17
2.1.5 ตัวแบบการเรียนรู้ของเครื่อง (Machine Learning).....	19
2.1.5.1 ประเภทของการเรียนรู้ของเครื่อง.....	20
2.1.5.2 การพยากรณ์โดยใช้เทคนิคแรนดอมฟอเรสต์ (Random Forest).....	21
2.1.5.3 ตัวแบบหน่วยความจำระยะสั้นระยะยาว สำหรับการพยากรณ์อนุกรมเวลา.....	24
2.1.6 เกณฑ์ที่ใช้สำหรับการวัดความแม่นยำของตัวแบบพยากรณ์.....	27
2.1.7 การแทนค่าข้อมูลสูญหาย.....	28
2.1.7.1 วิธีค่าเฉลี่ยเคลื่อนที่อย่างง่าย (Simple Moving Average : SMA)....	28
2.1.7.2 วิธีค่าเฉลี่ยเคลื่อนที่เอกซ์โพเนนเชียล (Exponential Moving Average : EMA).....	29
2.1.7.3 วิธีค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก (Weight Moving Average : WMA).....	30
2.1.8 การกำหนดค่าไฮเปอร์พารามิเตอร์.....	30
2.2 งานวิจัยที่เกี่ยวข้อง.....	31
บทที่ 3 วิธีการดำเนินงานวิจัย.....	33
3.1 ขั้นตอนการเตรียมข้อมูล.....	33
3.1.1 ลักษณะและแหล่งที่มาของข้อมูล.....	33
3.1.2 การจัดการข้อมูล.....	33
3.1.2.1 การแทนค่าข้อมูลสูญหาย (Data Imputation).....	34
3.1.2.2 การแบ่งข้อมูล (Dividing Data).....	34
3.1.2.3 การเตรียมข้อมูลบน Power BI.....	34
3.2 เครื่องมือที่ใช้ในการศึกษาวิจัย.....	35
3.3 เครื่องมือที่ใช้ในการทำพอร์ต (Report).....	35
3.4 ขั้นตอนการดำเนินงานวิจัย.....	35
3.5 การวิเคราะห์ข้อมูล.....	36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.5.1 การวิเคราะห์ลักษณะข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสรายวัน.....	36
3.5.2 การทดสอบความคงที่ของข้อมูล (Stationary).....	39
3.6 ตัวแบบพยากรณ์ที่ใช้ในงานวิจัย	39
3.6.1 ตัวแบบพยากรณ์วิธีบ็อก-เจนกินส์ด้วยตัวแบบ SARIMA.....	39
3.6.2 ตัวแบบพยากรณ์วิธีบ็อก-เจนกินส์ด้วยตัวแบบ SARIMAX.....	40
3.6.3 ตัวแบบพยากรณ์วิธีการเรียนรู้ของเครื่อง วิธีแรนดอมฟอเรส (Random Forest)	41
3.6.4 ตัวแบบพยากรณ์วิธีการเรียนรู้เชิงลึก วิธีหน่วยความจำระยะสั้นระยะยาว..	42
3.7 การพยากรณ์รายวันล่วงหน้าเป็นระยะเวลาหนึ่งเดือน หรือ 30 วัน	43
บทที่ 4 ผลการวิจัยและการอภิปรายผล.....	44
4.1 การวิเคราะห์ข้อมูลและแสดงข้อมูลในรูปของแผนภาพหรือกราฟ (Visualization)	44
4.2 การพยากรณ์วิธีการบ็อกซ์-เจนกินส์.....	57
4.2.1 ผลการทดลองวิธีบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMA.....	57
4.2.2 ผลการทดลองวิธีบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMAX.....	60
4.3 ผลการทดลองวิธี Random Forest.....	63
4.4 ผลการทดลองวิธีหน่วยความจำระยะสั้นระยะยาว (LSTM)	64
4.5 การเปรียบเทียบผลประสิทธิภาพของตัวแบบพยากรณ์.....	65
4.6 ผลการทดลองจากตัวแบบที่ดีที่สุดในการพยากรณ์รายวันใน 30 วันข้างหน้า	66
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	68
5.1 สรุปผลการวิจัย	68
5.2 อภิปรายผล	70
5.3 ข้อเสนอแนะ	70
เอกสารอ้างอิง.....	71
ภาคผนวก	75
ภาคผนวก ก	76
ภาคผนวก ข	88

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่		หน้า
ตารางที่ 2.1	ลักษณะของ ACF และ PACF ของตัวแบบ SARIMA	14
ตารางที่ 2.2	ความสัมพันธ์ระหว่างค่า MAPE กับระดับของความแม่นยำของลิวอิส (Lewis).....	28
ตารางที่ 4.1	ค่า AIC ในแต่ละอันดับของตัวแบบ ARIMA (p, d, q) SARIMA (P, D, Q)s	57
ตารางที่ 4.2	การทดสอบค่าพารามิเตอร์ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30}$	58
ตารางที่ 4.3	การทดสอบความเป็นอิสระของค่าความคลาดเคลื่อนของตัวแบบ ด้วย Box-Pierce and Ljung-Box Tests.....	59
ตารางที่ 4.4	ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดของตัวแบบ SARIMA.....	60
ตารางที่ 4.5	การทดสอบค่าพารามิเตอร์ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30} + \alpha_1 X_1$	61
ตารางที่ 4.6	การทดสอบความเป็นอิสระของค่าความคลาดเคลื่อนของตัวแบบ ด้วย Box-Pierce and Ljung-Box Test.....	62
ตารางที่ 4.7	ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดของตัวแบบ SARIMA.....	63
ตารางที่ 4.8	การกำหนดค่าไฮเปอร์พารามิเตอร์หรือกระบวนการในการสร้างตัวแบบ	63
ตารางที่ 4.9	ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดของตัวแบบแรนดอมฟอเรส (Random Forest).....	64
ตารางที่ 4.10	การกำหนดค่าไฮเปอร์พารามิเตอร์หรือกระบวนการในการสร้างตัวแบบ	64
ตารางที่ 4.11	ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดของตัวแบบการเรียนรู้เชิงลึก วิธีหน่วยความจำระยะสั้น ระยะยาว.....	65
ตารางที่ 4.12	ค่า MAPE ของแต่ละตัวแบบ.....	66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
รูปที่ 2.1 ขั้นตอนกระบวนการพยากรณ์.....	7
รูปที่ 2.2 อนุกรมเวลาที่เป็นสเตชันนารี (Stationary Series) และอนุกรมเวลาที่ไม่เป็นสเตชันนารี (Nonstationary Series).....	9
รูปที่ 2.3 อัลกอริทึมต้นไม้ตัดสินใจ.....	22
รูปที่ 2.4 หลักการทำ Random Forest จากการวิเคราะห์ข้อมูล ด้วยเทคนิค Data Mining เบื้องต้น.....	23
รูปที่ 2.5 ตัวอย่างการทำงานในเซลล์ของอัลกอริทึม โครงข่ายประสาทแบบ LSTM	24
รูปที่ 3.1 การเตรียมข้อมูลบน Power BI.....	34
รูปที่ 3.2 ข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของ เซอร์วิส ในช่วง 1/01/2021 – 31/12/2022.....	36
รูปที่ 3.3 ค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชัน และการทดสอบประสิทธิภาพของเซอร์วิส.....	37
รูปที่ 3.4 ค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของจำนวนการเรียกใช้เซอร์วิสของ แอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส.....	38
รูปที่ 3.5 ค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของผลต่างฤดูกาลครั้งที่ 1.....	38
รูปที่ 3.6 ค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของผลต่างฤดูกาลครั้งที่ 1.....	39
รูปที่ 3.7 สถาปัตยกรรมของแรนดอมฟอเรสสำหรับอนุกรมเวลา	41
รูปที่ 3.8 สถาปัตยกรรมของหน่วยความจำระยะสั้นระยะยาวสำหรับอนุกรมเวลา.....	42
รูปที่ 4.1 แสดงภาพรวมปริมาณงานที่เข้ามาทั้งปีและปริมาณทรัพยากรที่มีอยู่.....	45
รูปที่ 4.2 แสดงปริมาณงานในแต่ละเดือนของ Man – day ที่ใช้จริง.....	46
รูปที่ 4.3 แสดงปริมาณภาพรวมปริมาณ Man – day ที่ประมาณการณ์ล่วงหน้า.....	47
รูปที่ 4.4 แสดงภาพรวมปริมาณการแจ้งปัญหาการใช้งานของแอปพลิเคชัน	48
รูปที่ 4.5 แสดงปริมาณ Man – day ที่ใช้ Support incident จริงในแต่ละเดือน.....	49
รูปที่ 4.6 แสดงภาพรวมของปริมาณงานทั้งหมดที่ปิดงานในระบบ Service Now.....	50
รูปที่ 4.7 แสดงปริมาณการเรียกใช้เซอร์วิสของเกตเวย์.....	51
รูปที่ 4.8 แสดงภาพรวมเซอร์วิสทั้งหมดที่มีการเรียกใช้มากที่สุด	52
รูปที่ 4.9 แสดงปริมาณการเรียกใช้เซอร์วิสของเกตเวย์.....	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่	หน้า
รูปที่ 4.10 แสดงภาพรวม Transaction of Gateway.....	54
รูปที่ 4.11 แสดงอันดับ 1 – 10 ของเกตเวย์ที่มีการตอบกลับมากที่สุด.....	54
รูปที่ 4.12 แสดงปริมาณการเรียกใช้เซอร์วิสของไมโครเซอร์วิส.....	55
รูปที่ 4.13 แสดงภาพรวม 7 วันย้อนหลัง โดยแสดงตามชุดข้อมูลไมโครเซอร์วิส.....	56
รูปที่ 4.14 แสดงค่าเฉลี่ยระยะเวลาที่ดีที่สุด 10 อันดับแรกของไมโครเซอร์วิส.....	56
รูปที่ 4.15 ผลการพยากรณ์รายวันในอีก 30 วันข้างหน้า.....	66



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันธนาคารในประเทศไทยได้นำดิจิทัลแบงก์กิ้งเข้ามาใช้ในการพัฒนาผลิตภัณฑ์และบริการทั้งในด้านไอที ด้านของการลงทุนระบบโครงสร้างพื้นฐาน เพื่อให้สอดคล้องกับความต้องการของผู้บริโภคและพฤติกรรมของผู้บริโภคที่เปลี่ยนแปลงไป ภาคครัวเรือนไทยเข้าถึงและใช้บริการทางการเงินในภาพรวมเพิ่มขึ้นเป็นร้อยละ 89.8 ในปี 2561 จากร้อยละ 86.4 ในปี 2559 (ธปท, 2561) เนื่องจากเทคโนโลยีในปัจจุบันมีการเปลี่ยนแปลงไปอย่างรวดเร็ว ส่งผลให้โครงสร้างต้นทุนของอุตสาหกรรมลดลง เช่น ช่องทางบริการของธนาคารแบบเดิมที่เคยเห็นเป็นรูปธรรม เช่น สาขาของธนาคารและตู้เอทีเอ็ม ก็ถูกแทนที่ด้วยสมาร์ตโฟน ซึ่งเป็นการช่วยอำนวยความสะดวกให้แก่ลูกค้า เพราะลูกค้าสามารถทำธุรกรรมได้ทุกที่ ทุกเวลา สะดวก รวดเร็วและมีค่าธรรมเนียมที่ถูกลงกว่าเดิม

หน่วยงาน Enterprise Application Programming Interface (EAPI) ของธนาคารพาณิชย์แห่งหนึ่งทำหน้าที่เป็นตัวกลาง (Middleware) ในการเชื่อมต่อแอปพลิเคชันหน้าบ้าน (Frontend) กับหลังบ้าน (Backend) เกี่ยวกับเซิร์ฟเวอร์ต่าง ๆ ที่แอปพลิเคชันธนาคาร เช่น แอปพลิเคชันของทางธนาคารอยากทราบข้อมูลเกี่ยวกับการสร้างคิวอาร์โค้ด (QR Code) หรือรายการฝากเงิน ทางหน่วยงานจะทำหน้าที่ในการส่งข้อมูลเซิร์ฟเวอร์ที่ได้รับมาไปยังหน้าบ้าน (Frontend) เพื่อแสดงให้กับทางลูกค้า โดยทางหน่วยงานนั้นมีข้อมูลเกี่ยวกับการดำเนินการ (Transaction) เป็นจำนวนมากและมีความซับซ้อน ทำให้ยากต่อการเรียกค้น อีกทั้งทางหน่วยงานยังไม่มีการนำแดชบอร์ด (Dashboard) เข้ามาช่วยในการควบคุม จึงทำให้เกิดปัญหาในการตรวจสอบปัญหาหรือข้อผิดพลาดต่าง ๆ ทางหน่วยงานจึงมีความประสงค์ที่จะทราบถึงปัญหาหรือข้อผิดพลาดต่าง ๆ ในภาพรวม เพื่อใช้เป็นข้อมูลในการวางแผนจัดสรร รวมถึงเป็นเครื่องมือที่จะนำไปใช้ในการวางแผนการปฏิบัติงานของพนักงานในหน่วยงานในอนาคตให้เพิ่มประสิทธิภาพอีกด้วย

จากการที่ผู้วิจัยได้นำแดชบอร์ด (Dashboard) เข้ามาช่วยในการควบคุม ทำให้พบถึงปัญหาของทางหน่วยงาน ในการที่จะทราบได้ว่าระยะเวลาอีกนานเท่าไรที่ระบบของทางหน่วยงานจะล่ม เนื่องจากยังขาดการพยากรณ์ ผู้วิจัยจึงได้นำข้อมูลเพียงบางส่วนจากทางองค์กรมาทำการศึกษา โดยข้อมูลที่นำมาศึกษาคือข้อมูลสำหรับการส่งข้อมูลหรือข้อความบางอย่างไปยังซอฟต์แวร์ (Message Queue) ตามหัวข้อและลำดับที่กำหนด โดยทำการวิเคราะห์ลักษณะของข้อมูล พร้อมทั้งจัดเตรียมข้อมูลให้ได้ข้อมูลที่มีความเหมาะสมที่สุดในการทำ Visualization และการทำรีพอร์ต (Report) ด้วย Microsoft Power BI

เอกสารนี้เป็นเอกสารที่ผู้วิจัยได้ศึกษาเกี่ยวกับจำนวนเซิร์ฟเวอร์ที่เรียกผ่านแอปพลิเคชันของธนาคารในแต่ละวัน, ไม่ว่าความผิดปกติของเซิร์ฟเวอร์ ปริมาณความสามารถ (Capacity) ที่ใช้สนับสนุนระบบ ปริมาณงานที่เข้า

มาในแต่ละเดือนว่ามีกำลังคนที่เพียงพอหรือไม่ เพื่อให้องค์กรนำรีพอร์ต (Report) ไปใช้ประกอบตัดสินใจในการวางแผนการดำเนินการต่าง ๆ ที่ส่งผลต่อการเพิ่มประสิทธิภาพและลดระยะเวลาในการดำเนินการ

นอกจากนี้การศึกษาวิจัยครั้งนี้ได้ทำการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ เนื่องจากการพยากรณ์รายวันจะเป็นประโยชน์ต่อการตัดสินใจการวางแผนและการเตรียมตัวจากการคาดการณ์ในอนาคต ผู้วิจัยได้นำข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันรายวัน จากฐานข้อมูลของบริษัทมาทำการศึกษา เพื่อสร้างตัวแบบที่จะใช้สำหรับพยากรณ์รายวันจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส โดยทำการวิเคราะห์ลักษณะของข้อมูล พร้อมทั้งจัดการกับข้อมูล เพื่อให้ข้อมูลมีความเหมาะสมที่จะนำไปใช้เป็นตัวแบบพยากรณ์ทางสถิติ นอกจากนี้ยังได้นำตัวแบบที่ดีที่สุดนี้ไปพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันรายวัน ในอนาคตเป็นระยะเวลา 30 วัน โดยทำการนับจากวันสุดท้ายของข้อมูลที่นำตัวแบบนี้ไปพยากรณ์

1.2 วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้จัดทำขึ้นมา โดยมีวัตถุประสงค์ดังนี้

- 1) เพื่อนำเสนอข้อมูลการใช้งานและค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ โดยการวิเคราะห์ข้อมูลและแสดงข้อมูลในรูปของแผนภาพหรือกราฟ Visualization รวมทั้งจัดทำรีพอร์ต (Report)
- 2) สร้างตัวแบบพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันรายวัน โดยอาศัยวิธีบ็อกซ์-เจนกินส์ วิธีการเรียนรู้ของเครื่อง และวิธีการเรียนรู้เชิงลึก
- 3) เปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ โดยพิจารณาจากร้อยละค่าสัมบูรณ์ของค่าความคลาดเคลื่อนโดยเฉลี่ย

1.3 ขอบเขตของงานวิจัย

1.3.1 ขอบเขตด้านข้อมูล

งานวิจัยนี้ใช้ข้อมูลจากธนาคารพาณิชย์แห่งหนึ่ง ซึ่งเป็นข้อมูลทุติยภูมิ โดยข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลอนุกรมเวลารายวันจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชัน

1.3.2 ขอบเขตด้านเวลา

- 1) ข้อมูลตั้งแต่วันที่ 1 มกราคม 2564 ถึง 31 ธันวาคม 2565 รวมทั้งสิ้น 707 วัน สำหรับใช้ในการสร้างตัวแบบสมการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) ข้อมูลตั้งแต่วันที่ 1 มกราคม 2566 ถึง 31 มีนาคม 2566 รวมทั้งสิ้น 90 วัน
 สำหรับใช้ในการทดสอบความถูกต้องของตัวแบบสมการพยากรณ์ โดยพิจารณาจากค่าความคลาดเคลื่อนระหว่างค่าพยากรณ์และค่าที่เกิดขึ้นจริง

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) มองเห็นภาพรวมของข้อมูลเกี่ยวกับจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชัน ประสิทธิภาพของเซอร์วิส และทรัพยากรของเซิร์ฟเวอร์ ทำให้ทราบถึงสถานการณ์และปัญหาความเพียงพอของทรัพยากรที่มีในปัจจุบัน
- 2) สามารถนำผลการพยากรณ์ไปใช้ประกอบการวางแผนและจัดสรรทรัพยากรของเซิร์ฟเวอร์ให้รองรับกับจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชัน ในขณะเดียวกัน ค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์มีค่าต่ำสุด
- 3) นำตัวแบบพยากรณ์ไปเป็นต้นแบบในการพัฒนาสำหรับชุดข้อมูลอื่น ๆ

1.5 นิยามศัพท์เฉพาะ

Application Programming Interface (API) คือ การเชื่อมต่อจากระบบหนึ่งไปสู่อีก ระบบหนึ่ง เพื่อให้ซอฟต์แวร์ภายนอกสามารถเข้าถึงและอัปเดตข้อมูลได้ และเป็นตัวกลางในการรับคำสั่งต่างๆ ระหว่างแอปพลิเคชัน ประมวลผลและส่งกลับคืนไปโดยอัตโนมัติ

Service Level Agreement (SLA) คือ วันปิดงาน ซึ่งไม่นับรวมวันเสาร์-อาทิตย์

Man - days คือ วันทำงานต่อคน เป็นหน่วยที่ใช้วัดการทำงานของคนงานในสถานประกอบการ โดยคิดเป็นหนึ่งวันต่อหนึ่งคน

Capacity คือ ชัดความสามารถของคนงาน เครื่องจักร หรือองค์กรในการผลิตผลผลิตต่อหน่วยเวลา (เป็นปริมาณของงานที่สามารถทำได้ในช่วงเวลาใดเวลาหนึ่งที่กำหนดไว้) ซึ่งกำลังการผลิตเป็นอัตราการทำงานไม่ใช่ปริมาณของงานที่ทำได้

Transactions Per Second (TPS) คือ จำนวนการทำรายการต่อวินาที

อนุกรมเวลา คือ ชุดของข้อมูลเชิงปริมาณที่จัดเก็บช่วงเวลาใดช่วงเวลาหนึ่ง โดยจัดเก็บตามเวลาที่บันทึกข้อมูลนั้นโดยมีระยะห่างเท่าๆ กัน เช่น รายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือจะเป็นรายชั่วโมง

การวิเคราะห์อนุกรมเวลาด้วยตัวแบบ คือ การวิเคราะห์อนุกรมเวลาโดยอาศัยการสร้างแบบจำลองพยากรณ์ (Forecasting Model) ซึ่งจะใช้ข้อมูลในอดีตเพื่อศึกษาการเปลี่ยนแปลงของข้อมูลในช่วงเวลาที่สนใจ

Project Planning คือ โครงการทั้งหมดที่อยู่ในแผนงาน

Business as usual (BAU) คือ ปริมาณงานทั้งหมดที่ทีมรับมา

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น ยกเว้นกรณีเห็นเหตุจำเป็นและต้องขออนุญาตจากอธิการบดีของมหาวิทยาลัยทุกครั้งที่มีการนำไปใช้

EAPI คือ งาน Internal Improvement ของทีม

Support Business คือ สนับสนุนธุรกิจ

Design Engineer (Design) คือ วิจัยพัฒนา ทดสอบ ออกแบบตัวอย่าง

Developer (Dev) คือ ผู้พัฒนาเขียนโปรแกรม เขียนโค้ด

System Integration Test (SIT) คือ แผนกตรวจสอบความถูกต้องของโปรแกรมก่อนส่งให้ลูกค้า

User Acceptance Test (UAT) คือ การทดสอบใช้โปรแกรมกับ user ผู้ใช้งาน

Close_IC_Done คือ งานที่ปิดใบงานผ่านระบบเรียบร้อยแล้ว

To_Be_Closed_IC คือ งานที่ยังไม่ปิดงานหรือรอปิดงาน

Pending คือ งานที่รอดำเนินการแก้ไขหรือหาสาเหตุของปัญหาที่แจ้งเข้ามา

Transaction of microservice คือ การทำรายการเรียกใช้เซอร์วิสของไมโครเซอร์วิส

Endpoint คือ url path ของการเรียกใช้ไมโครเซอร์วิส

Service คือ เส้น microservice ที่เปิดใช้ให้แอปพลิเคชันเรียกใช้งาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ ผู้วิจัยจะกล่าวถึงทฤษฎีต่าง ๆ และงานวิจัยที่เกี่ยวข้องเพื่อเป็นพื้นฐานในการศึกษาวิเคราะห์ และเป็นแนวทางในการพยากรณ์ตัวแบบพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ โดยมีรายละเอียดดังต่อไปนี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ความหมายของการพยากรณ์

กัลยา (2544) ได้ให้คำนิยาม การพยากรณ์คือ การประมาณหรือการคาดคะเนว่าอะไรจะเกิดขึ้นในอนาคต การพยากรณ์แต่ละปัญหามีความแตกต่างกัน ในการพยากรณ์แต่ละปัญหาจึงควรต้องมีการพิจารณาถึงปัจจัยที่สำคัญบางประการ เช่น ระยะเวลาในอนาคตที่ต้องการพยากรณ์ เพื่อนำผลที่ได้มาช่วยสนับสนุนการตัดสินใจของผู้บริหาร

ทรงศิริ (2549) ได้ให้คำนิยาม การพยากรณ์ (Forecasting) คือ การคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคต โดยอาศัยการวิเคราะห์ข้อมูลในอดีตที่เก็บรวบรวมอย่างมีระบบโดยอาศัยเทคนิคการพยากรณ์ต่าง ๆ เพื่อเพิ่มประสิทธิภาพในการดำเนินงานทางธุรกิจสามารถนำไปใช้เพื่อให้ทราบถึงแนวโน้มการเปลี่ยนแปลงของสถานการณ์หรือสภาพแวดล้อมที่มีผลในอนาคตและทำให้สามารถที่จะวางแผนหรือกำหนดนโยบายเพื่อให้บรรลุวัตถุประสงค์

นิภา (2551) ได้ให้คำนิยาม การพยากรณ์ (Forecasting) หมายถึง การคาดคะเนหรือทำนายการเกิดเหตุการณ์หรือสภาพการณ์ต่าง ๆ ในอนาคต โดยการพยากรณ์จะทำการศึกษานโยบายและรูปแบบการเกิดเหตุการณ์จากข้อมูลในอดีตหรือใช้ความรู้ ความสามารถ ประสบการณ์ และดุลยพินิจของผู้พยากรณ์

2.1.2 เทคนิควิธีการพยากรณ์

วิธีการพยากรณ์ที่ใช้กันอยู่ทั่วไปสามารถแบ่งได้เป็นสองประเภท คือ การพยากรณ์เชิงคุณภาพ (Qualitative Forecast) และการพยากรณ์เชิงปริมาณ (Quantitative Forecast)

1) วิธีการพยากรณ์เชิงคุณภาพ

มุกดา (2549) ได้กล่าวถึงการพยากรณ์โดยใช้ความคิดเห็นของผู้เชี่ยวชาญเป็นวิธีการพยากรณ์เชิงคุณภาพ โดยทั่วไปการพยากรณ์เชิงคุณภาพ ทำซ้ำให้เหมือนเดิมกับผู้พยากรณ์คนใหม่ไม่ได้ ซึ่งอาจเนื่องมาจากสารสนเทศที่ไม่ชัดเจน อย่างไรก็ตามการพยากรณ์เชิงคุณภาพอาจเหมาะสมที่สุด หรืออาจเป็นวิธีการพยากรณ์วิธีเดียวที่ใช้ได้ในบางสถานการณ์ วิธีการพยากรณ์เชิงคุณภาพมีหลายวิธี เช่น Delphi หรือวิธีการพยากรณ์โดยใช้ Subjective Curve เป็นต้น

2) วิธีการพยากรณ์เชิงปริมาณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

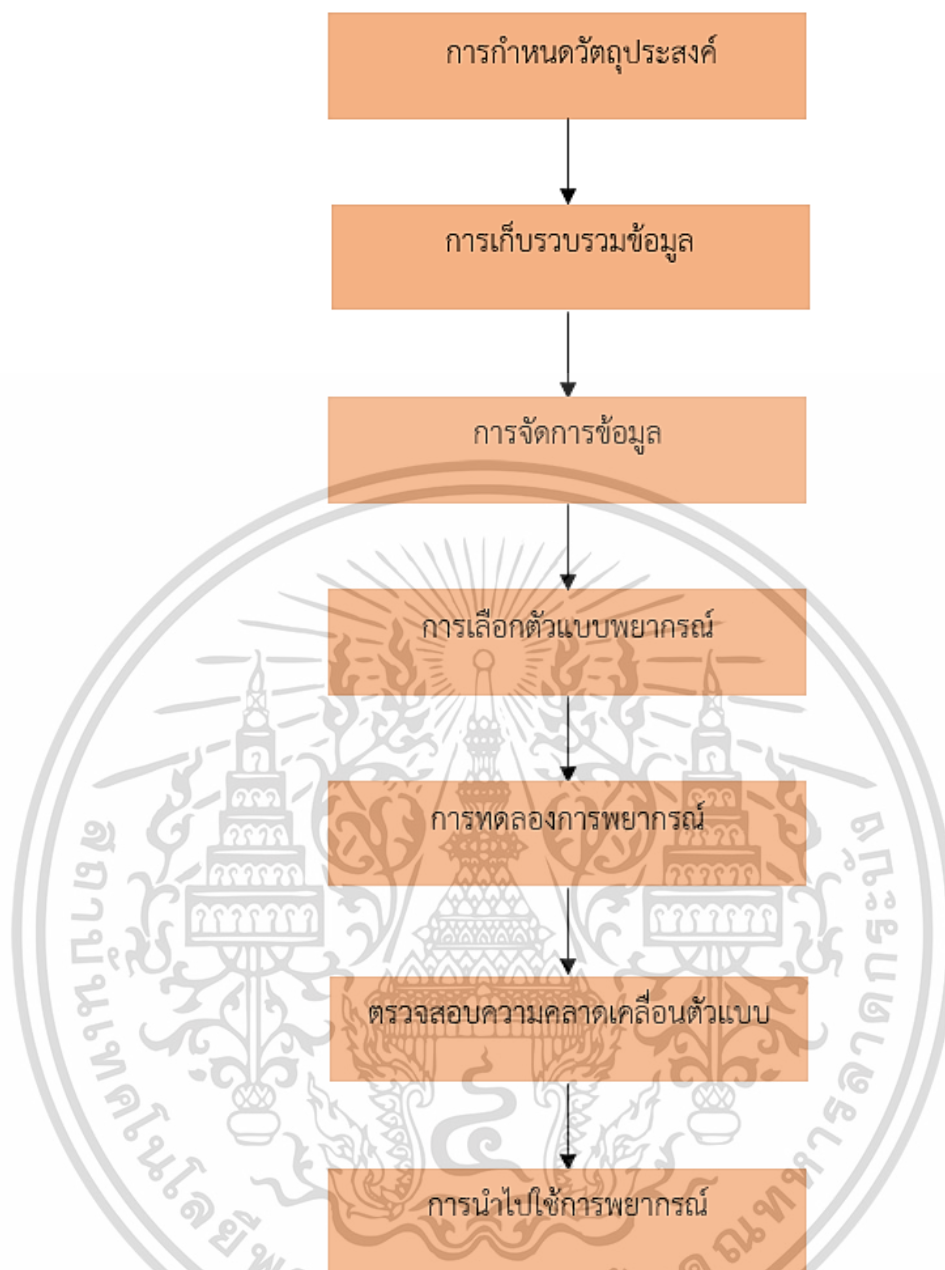
มุกดา (2549) ได้ให้คำนิยาม การพยากรณ์โดยพิจารณาว่าข้อมูลในอดีตและปัจจุบัน เป็นค่าของตัวแปรที่สามารถอธิบายได้ด้วยตัวแบบทางคณิตศาสตร์หรือสถิติ และใช้ตัวแบบที่ได้พยากรณ์ค่าของตัวแปรนั้นต่อไปในอนาคตเรียกว่าการพยากรณ์เชิงปริมาณ ตัวอย่างเช่น การพยากรณ์โดยใช้ตัวแบบเศรษฐมิติ (Econometrics Models) หรือตัวแบบการถดถอยพหุคูณ (Multiple Regression Models) เป็นต้น

2.1.3 ขั้นตอนในการพยากรณ์

กฤษฎดา (2556) ได้เสนอกระบวนการพยากรณ์สามารถแบ่งออกเป็น 7 ขั้นตอน ดังรูปที่ 2.1 แสดงกระบวนการพยากรณ์ทั้ง 7 ขั้นตอน ซึ่งแต่ละขั้นตอนต้องตรวจสอบอย่างละเอียดเพื่อให้ได้ค่าพยากรณ์ที่ถูกต้องมากที่สุดดังต่อไปนี้

- 1) การกำหนดวัตถุประสงค์ เป็นการกำหนดวัตถุประสงค์ให้ชัดเจนว่าการพยากรณ์จะไปใช้ในการตัดสินใจอย่างไร เช่น ใช้เพื่อวางแผนกลยุทธ์
- 2) การเก็บรวบรวมข้อมูล เป็นการเก็บรวบรวมข้อมูลที่เกี่ยวข้องกับการพยากรณ์อย่างเหมาะสม และเป็นข้อมูลที่ต้องการ
- 3) การจัดการข้อมูล บางครั้งข้อมูลที่เก็บรวบรวมมีมากเกินไปและทำให้การพยากรณ์ถูกต้องน้อยลง จึงจำเป็นต้องจัดการข้อมูลบางตัวที่อาจไม่เกี่ยวข้องกับการพยากรณ์ลง
- 4) การเลือกตัวแบบพยากรณ์ การเลือกตัวแบบพยากรณ์ที่เหมาะสมกับข้อมูลจะช่วยลดความผิดพลาดในการพยากรณ์
- 5) การทดลองการพยากรณ์ เป็นการพยากรณ์เหตุการณ์ที่ผ่านไปโดยใช้ข้อมูลจริงที่มีอยู่แล้วประเมินว่าวิธีใดเหมาะสม โดยการวัดค่าความคลาดเคลื่อนที่เกิดขึ้น
- 6) การตรวจสอบความแม่นยำของตัวแบบ ประเมินว่าวิธีใดเหมาะสม โดยการวัดค่าความคลาดเคลื่อนที่เกิดขึ้น
- 7) การนำไปใช้พยากรณ์ การติดตามผลอย่างต่อเนื่องว่าผลการพยากรณ์ ที่เคยพยากรณ์ได้ดีที่สุดมีความถูกต้องน้อยลงหรือไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 ขั้นตอนกระบวนการพยากรณ์ (กฤษดา, 2556)

2.1.4 การพยากรณ์วิธีบ็อกซ์ - เจนกินส์ (Box and Jenkins Method)

Box and Jenkins (1970) ได้นำเสนอรูปแบบ Autoregressive Integrated Moving Average (ARIMA) หรือรู้จักกันในชื่อของรูปแบบของ Box และ Jenkins ไว้ในหนังสือชื่อ Time Series Analysis: Forecasting and Control ที่เริ่มพิมพ์เผยแพร่ในปี ค.ศ.1970 และปรับปรุงในปี ค.ศ.1994 ซึ่งได้รับความนิยมมากจนถึงปัจจุบัน โดยกำหนดรูปแบบ ARIMA ซึ่งเป็นรูปแบบที่อธิบายลักษณะการเคลื่อนไหวของอนุกรมเวลาที่ค่าสังเกตมีสหสัมพันธ์แบบอัตโนมัติและเป็นสเตชันนารี แม้ว่าอนุกรมเวลาที่เก็บรวบรวมโดยทั่วไปจะมีการเคลื่อนไหวเนื่องจากแนวโน้มหรือฤดูกาล ซึ่งเป็นลักษณะของอนุกรมเวลาที่ไม่เป็นสเตชันนารี แต่สามารถใช้รูปแบบ ARIMA กับอนุกรมเวลาที่ไม่เป็นสเตชันนารีได้ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สเตชันนารี โดยกำหนดรูปแบบ ARIMA ให้กับอนุกรมเวลาใหม่ ซึ่งอนุกรมเวลาใหม่ที่ได้จะเป็นอนุกรมเวลาที่ เป็นสเตชันนารี

อนุกรมเวลาสามารถแบ่งออกเป็น 2 ประเภท ได้ดังนี้

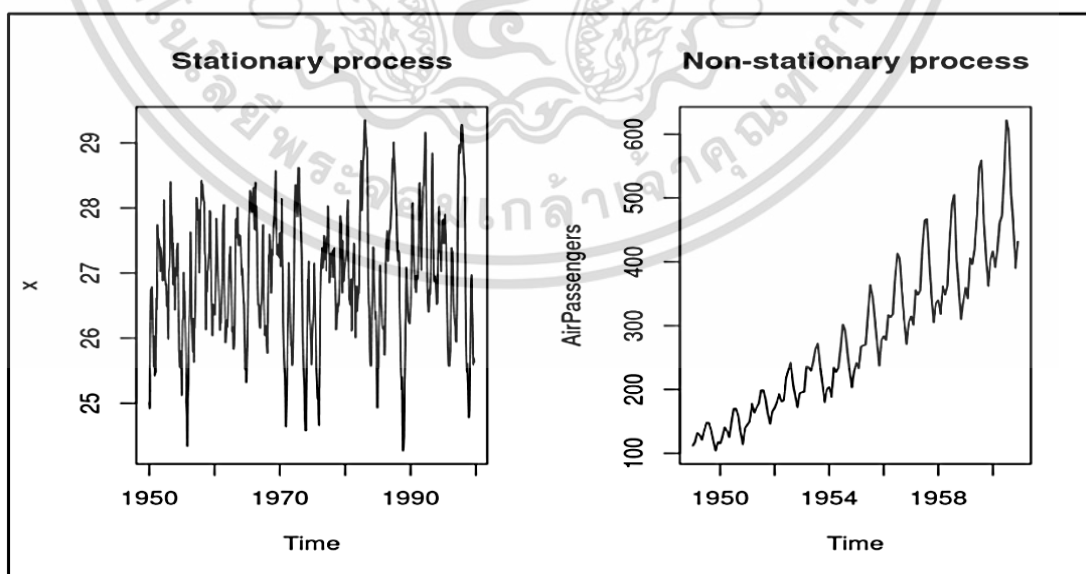
1) อนุกรมเวลาที่เป็นสเตชันนารี (Stationary Series)

อนุกรมเวลาที่เป็นสเตชันนารี เป็นอนุกรมเวลา $\{y_t\}$ ที่มีค่าเฉลี่ยและค่าความแปรปรวนของ $\{y_t\}$ คงที่ นอกจากนั้นค่าสัมประสิทธิ์สหสัมพันธ์แบบอัตโนมัติในช่วงเวลาห่าง k ขึ้นกับค่า k อย่างเดียว นั่นคือ ไม่ขึ้นกับเวลา t ตัวอย่างของอนุกรมเวลาที่เป็นสเตชันนารีเช่น อนุกรมเวลาที่มีลักษณะการเกิดคองที่ ได้แก่ จำนวนผู้ป่วยนอกรายเดือน จำนวนผู้เข้ามาปรึกษาจิตกับสภานายความ อนุกรมเวลาที่มีการปรับการเจริญเติบโตแล้ว อนุกรมเวลาของค่าความคลาดเคลื่อน อนุกรมเวลาที่ได้จากการแปลงค่าสังเกต

2) อนุกรมเวลาที่ไม่เป็นสเตชันนารี (Nonstationary Series)

อนุกรมเวลาที่ไม่เป็นสเตชันนารีเป็นอนุกรมเวลาที่ไม่มีความสมบัติเป็นสเตชันนารี นั่นคือค่าเฉลี่ยและหรือค่าความแปรปรวนของ $\{y_t\}$ ไม่คงที่ที่เวลา t และค่าสัมประสิทธิ์สหสัมพันธ์แบบอัตโนมัติในช่วงเวลาห่าง k ไม่ขึ้นอยู่กับเฉพาะค่า k แต่ขึ้นอยู่กับเวลา t ด้วย สำหรับอนุกรมที่มีแนวโน้มและ/หรือฤดูกาลเป็นลักษณะของอนุกรมเวลาที่มีค่าเฉลี่ยหรือค่าคาดหวังของ $\{y_t\}$ หรือ μ_t ไม่คงที่ส่วนอนุกรมเวลาที่ความผันแปรของ $\{y_t\}$ ที่เวลา t ไม่คงที่จะเป็นลักษณะของอนุกรมเวลาที่ค่าความแปรปรวนของ $\{y_t\}$ หรือ σ_t^2 ที่เวลา t ไม่คงที่ อนุกรมเวลาที่มีค่าเฉลี่ยหรือค่าความแปรปรวนไม่คงที่ดังกล่าวจะไม่มีคุณสมบัติเป็นสเตชันนารี

ตัวอย่างอนุกรมเวลาที่เป็นสเตชันนารี (Stationary Series) และ อนุกรมเวลาที่ไม่เป็นสเตชันนารี (Nonstationary Series) แสดงดังรูปที่ 2.2



รูปที่ 2.2 อนุกรมเวลาที่เป็นสเตชันนารี (Stationary Series) และอนุกรมเวลาที่ไม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการศึกษาเท่านั้น ไม่ควรเผยแพร่ไปโดยไม่ได้รับอนุญาตจากเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เป็นสเตชันนารี (Nonstationary Series) (Döring, 2018)

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.4.1 การทดสอบความเป็นสแตชันนารีของข้อมูล (Stationary)

การทดสอบอนุกรมเวลา เป็นสิ่งที่ควรกระทำก่อนที่จะนำข้อมูลอนุกรมเวลามาใช้ในการวิเคราะห์ โดยเฉพาะเงื่อนไขความนิ่งของอนุกรมเวลา ซึ่งเป็นเงื่อนไขที่สำคัญในการนำข้อมูลอนุกรมเวลามาใช้ ดังนั้นถ้าหากอนุกรมเวลาที่ไม่คงที่จะต้องทำให้อนุกรมเวลาดังกล่าวคงที่ก่อน โดยการหาผลต่างของอนุกรมเวลา การคงที่ของอนุกรมเวลา หมายถึงอนุกรมเวลาที่อยู่ในสภาวะสมดุลเชิงสถิติ (Statistical Equilibrium) ซึ่งก็คือการที่คุณสมบัติทางสถิติของอนุกรมเวลาไม่เปลี่ยนแปลงตามกาลเวลา

สถิติ ADF (Augmented Dickey-Fuller Test) ใช้เพื่อตรวจสอบข้อมูลที่ได้ทำการแปลงแล้วนั้นเป็นข้อมูลที่อยู่ในสถานะคงที่หรือไม่ การทดสอบ ADF พัฒนามาจากวิธี Dickey – Fuller Test เพื่อแก้ปัญหา Serial Correlation หรือแก้ปัญหา Autocorrelation ซึ่งค่าสถิติที่ได้จะขาดความถูกต้อง ดังนั้นจึงได้มีการนำเสนอให้ปรับสมการของวิธีการ Dickey-Fuller Test ใหม่โดยใส่ตัวแปรล่าช้า (Lag) ของ ข้อมูลในลำดับที่สูงขึ้น เรียกวิธีการนี้ว่า Augmented Dickey-Fuller Test ในการตรวจสอบว่าข้อมูลนั้นมีลักษณะนิ่งหรือไม่ โดยสามารถทำการทดสอบตัวแปรในกรณีที่เป็น Serial correlation ในค่าความคลาดเคลื่อน Error Term มีลักษณะความสัมพันธ์กันเองในระดับสูง โดยสามารถแสดงดังสมการที่ 2.1

$$\Delta y_t + \alpha_1 + \alpha_2 T + \gamma y_{t-1} + \sum_{i=2}^{\infty} \beta_i \Delta y_{t-i} + \varepsilon_t \quad (2.1)$$

โดยกำหนดให้

Δy_t แทนค่าความแตกต่างครั้งที่ 1 ของตัวแปรที่ทำการศึกษา

α_1 แทน ค่าคงที่

α_2 แทน ค่าสัมประสิทธิ์แนวโน้ม

T แทน แนวโน้ม

γ แทน ค่าสัมประสิทธิ์ความล่าช้า (Lag)

β แทน ค่าสัมประสิทธิ์ของความล่าช้า (Lag) ที่ i

ε_t แทน ค่าความคลาดเคลื่อน

ขั้นตอนการทดสอบสมมติฐาน เป็นดังนี้

1) กำหนดสมมติฐานเพื่อการทดสอบ

$$H_0 : \gamma = 0 \text{ (ข้อมูลมีลักษณะไม่คงที่)}$$

$$H_1 : \gamma \neq 0 \text{ (ข้อมูลมีลักษณะคงที่)}$$

2) ระดับนัยสำคัญ $\alpha = 0.05$

3) สถิติทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานที่ $t = \frac{\gamma}{Se_\gamma}$ มีการแจกแจงแบบที (t) ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยกำหนด γ แทน ค่าประมาณของพารามิเตอร์
 Se_γ แทน ค่าความคลาดเคลื่อนมาตรฐานของ γ

4) การตัดสินใจ

ปฏิเสธ H_0 ถ้า $|t| > DF_{Critical}$

ยอมรับ H_0 ถ้า $|t| \leq DF_{Critical}$

หรือ

ปฏิเสธ H_0 ถ้า $p\text{-value} < \alpha$

ยอมรับ H_0 ถ้า $p\text{-value} \geq \alpha$

ถ้ายอมรับ H_0 แสดงว่าข้อมูลมีลักษณะไม่คงที่ (Nonstationary) แต่ถ้าปฏิเสธ H_0 แสดงว่าข้อมูลมีลักษณะคงที่ โดยข้อมูลมีลักษณะคงที่ หมายถึง ข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) เท่ากันตลอดระยะเวลาที่ศึกษา ส่วนข้อมูลมีลักษณะไม่คงที่ (Non-Stationary) หมายถึง ข้อมูลอนุกรมเวลาที่มีค่าเฉลี่ย (Mean) และความแปรปรวน (Variance) ไม่เท่ากันระยะเวลาที่ศึกษา

2.1.4.2 การทดสอบแนวโน้มและฤดูกาล

ทรวงศิริ (2559) กล่าวว่า ก่อนการวิเคราะห์อนุกรมเวลาเพื่อพยากรณ์ค่าในอนาคต ผู้พยากรณ์ควรทราบว่าอนุกรมเวลามีการเคลื่อนไหวเนื่องจากส่วนประกอบใดบ้าง เพื่อกำหนดรูปแบบอนุกรมเวลาจากส่วนประกอบที่เป็นไปได้ ได้แก่ แนวโน้ม ฤดูกาล วัฏจักรและเหตุการณ์ที่ผิดปกติ การพิจารณาการเคลื่อนไหว ขั้นต้นเริ่มจากการพล็อตอนุกรมเวลา อย่างไรก็ตามบางครั้งไม่สามารถสรุปจากพล็อตว่ามีส่วนประกอบใดบ้างจึงต้องใช้การทดสอบสมมติฐานเพื่อหารูปแบบอนุกรมเวลาที่เหมาะสมโดยมีวิธีการต่าง ๆ ดังนี้

1. การทดสอบแนวโน้ม

การทดสอบแนวโน้มในกรณีไม่ใช้พารามิเตอร์เป็นการทดสอบที่ไม่มีมีความจำเป็นต้องทราบการแจกแจงของประชากร และสามารถใช้ทดสอบสมมติฐานที่มีตัวอย่างขนาดเล็ก ซึ่งวิธีการทดสอบมักใช้กับข้อมูลที่อยู่ในรูปแสดงถึงลำดับที่ หรือข้อมูลที่อยู่ในรูปความถี่ โดยใช้การทดสอบแบบวิ่ง (Run Test) เป็นการทดสอบที่ตัวทดสอบสถิติได้จากการพิจารณาแต่ละค่าสังเกตในอนุกรมเวลาเทียบกับค่ามัธยฐานของอนุกรมเวลา โดยพิจารณาเครื่องหมายของผลต่างระหว่างค่าสังเกตและค่ามัธยฐาน เครื่องหมายจะเป็นบวกถ้าค่าสังเกตสูงกว่าค่ามัธยฐานและเป็นลบถ้าค่าสังเกตต่ำกว่าค่ามัธยฐาน และนับจำนวนครั้งที่เครื่องหมายเดียวกันอยู่ติดกัน

ขั้นตอนของการทดสอบแบบวิ่ง มีดังนี้

1) กำหนดสมมติฐาน

H_0 : อนุกรมเวลาไม่มีแนวโน้ม

H_1 : อนุกรมเวลามีแนวโน้ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่สามารถนำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 2) ระดับนัยสำคัญ $\alpha = 0.05$
- 3) สถิติทดสอบ
กำหนด $R =$ จำนวนครั้งที่เครื่องหมายต่างกันตามเวลา
- 4) การตัดสินใจ

$$m = \begin{cases} \frac{n}{2} & \text{เมื่อ } n \text{ เป็นขนาดอนุกรมเวลาที่เป็นเลขคู่} \\ \frac{n-1}{2} & \text{เมื่อ } n \text{ เป็นขนาดอนุกรมเวลาที่เป็นเลขคี่} \end{cases}$$

4.1 สำหรับอนุกรมเวลาขนาดเล็กหรือ $m \leq 20$

บริเวณวิกฤติ คือ $R \geq R_{U,m}$ หรือ $R \leq R_{L,m}$

4.2 สำหรับอนุกรมเวลาขนาดใหญ่หรือ $m > 20$ ใช้ตัวสถิติ

$$Z = \frac{R - \mu_R}{\sigma_R} \quad (2.2)$$

$$\text{เมื่อ } \mu_R = m + 1 \text{ และ } \sigma_R = \sqrt{\frac{m(m-1)}{2m-1}}$$

บริเวณวิกฤติ คือ $|Z| \geq Z_{\frac{\alpha}{2}}$

2. การทดสอบฤดูกาล

การทดสอบความแตกต่างระหว่างกลุ่มของ Kruskal และ Wallis Kruskal and Wallis (1952) กล่าวว่า เป็นการทดสอบความแตกต่างระหว่างกลุ่มของข้อมูลมาจากประชากรเดียวกันหรือไม่ สำหรับการทดสอบฤดูกาลด้วยการทดสอบความแตกต่างระหว่างกลุ่มของ Kruskal และ Wallis กลุ่มจะหมายถึงฤดูกาล มีจำนวนกลุ่มที่ต้องการทดสอบเท่ากับจำนวนฤดูกาลต่อปี

ขั้นตอนของการทดสอบความแตกต่างระหว่างกลุ่มของ Kruskal และ Wallis มีดังนี้

- 1) กำหนดสมมติฐาน
 H_0 : อนุกรมเวลาไม่มีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง
 H_1 : อนุกรมเวลามีอิทธิพลฤดูกาลเข้ามาเกี่ยวข้อง
- 2) ระดับนัยสำคัญ $\alpha = 0.05$
- 3) สถิติทดสอบ

$$H = \sum_{i=1}^L n_i (\bar{R}_i - \bar{R})^2 = \frac{12}{n(n+1)} \left(\sum \frac{R_i^2}{n_i} \right) - 3(n+1)$$

(2.3)

เมื่อ $n_i =$ จำนวนค่าสังเกต Y_i ในฤดูกาลที่ i ซึ่ง $n = \sum n_i$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่สามารถนำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\bar{R}_i = \text{ค่าเฉลี่ยของลำดับของฤดูกาลที่ } i \text{ ซึ่ง } \bar{R} = \frac{R_i}{n_i}$$

$$\bar{R} = \text{ค่าเฉลี่ยของลำดับซึ่ง } \bar{R} = \frac{\sum R_i}{n_i} = \frac{\sum \bar{R}_i}{L}$$

4) การตัดสินใจ

$$\text{บริเวณวิกฤติ คือ } H \geq \chi_{\alpha, L-1}^2$$

2.1.4.3 การแปลงอนุกรมเวลาให้เป็นสเตชันนารีด้วยวิธีหาผลต่าง (Differencing)

การแปลงอนุกรมเวลาเพื่อทำให้อนุกรมเวลาอนุกรมเวลาเดิมที่ไม่สเตชันนารีเป็นอนุกรมเวลาใหม่ที่เป็นสเตชันนารีมีดังนี้

1) หาอนุกรมเวลาผลต่าง ถ้าอนุกรมเวลา $\{y_t\}$ มีการเคลื่อนไหวเนื่องจากแนวโน้ม จะแปลงอนุกรมเวลาเดิม $\{y_t\}$ เป็นอนุกรมเวลาใหม่ $\{z_t\}$ ที่ไม่มีแนวโน้ม ซึ่ง $Z_t = \nabla^d Y_t$ โดย d เป็นจำนวนครั้งของการหาผลต่าง เช่น

$$\text{เมื่อ } d = 1 \text{ จะมี } Z_t = \nabla Y_t = Y_t - Y_{t-1}$$

$$\begin{aligned} \text{เมื่อ } d = 2 \text{ จะมี } Z_t &= \nabla^2 Y_t \\ &= \nabla^2 (Y_t - Y_{t-1}) \\ &= \nabla Y_t - \nabla Y_{t-1} \\ &= Y_t - 2Y_{t-1} + Y_{t-2} \end{aligned}$$

จำนวนครั้งที่หาผลต่างขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็นสเตชันนารีแล้วหรือไม่ ถ้ายังไม่เป็นสเตชันนารีต้องหาผลต่างต่อไป

2) หาอนุกรมเวลาผลต่างฤดูกาล ถ้าอนุกรมเวลา $\{y_t\}$ มีการเคลื่อนไหวเนื่องจากฤดูกาลจะแปลงอนุกรมเดิม $\{y_t\}$ เป็นอนุกรมเวลาใหม่ $\{z_t\}$ ที่ไม่มีฤดูกาล ซึ่ง $Z_t = \nabla_L^D Y_t$ โดย D เป็นจำนวนครั้งของการหาผลต่างฤดูกาลและ L เป็นจำนวนฤดูกาลต่อปี เช่น สำหรับอนุกรมเวลารายไตรมาส ($L = 4$)

$$\text{เมื่อ } D = 1 \text{ จะมี } Z_t = \nabla_4 Y_t = Y_t - Y_{t-4}$$

$$\begin{aligned} \text{เมื่อ } D = 2 \text{ จะมี } Z_t &= \nabla_4^2 Y_t \\ &= \nabla_4 (Y_t - Y_{t-4}) \\ &= Y_t - 2Y_{t-4} + Y_{t-8} \end{aligned}$$

จำนวนครั้งที่หาผลต่างฤดูกาลขึ้นอยู่กับว่าเมื่อหาผลต่างฤดูกาลแล้วอนุกรมเวลาใหม่เป็นสเตชันนารีแล้วหรือไม่ ถ้ายังไม่เป็นสเตชันนารีต้องหาผลต่างฤดูกาลต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) **หาค่าอนุกรมเวลาผลต่างและผลต่างฤดูกาล** กรณีที่อนุกรมเวลา $\{y_t\}$ มีการเคลื่อนไหวเนื่องจากแนวโน้มและฤดูกาล จะแปลงอนุกรมเวลาเดิม $\{y_t\}$ เป็นอนุกรมเวลาใหม่ $\{z_t\}$ ที่ไม่มีแนวโน้มและฤดูกาลโดยหาผลต่างและผลต่างฤดูกาล ซึ่ง $Z_t = \nabla^d \nabla_L^D Y_t$ เช่น สำหรับอนุกรมเวลารายไตรมาสที่มีการเคลื่อนไหวเนื่องจากแนวโน้มและฤดูกาล

$$\begin{aligned} \text{เมื่อ } d = 1 \text{ และ } D = 1 \text{ จะมี } Z_t &= \nabla \nabla_4 Y_t \\ &= \nabla(Y_t - Y_{t-4}) \\ &= Y_t - Y_{t-1} - Y_{t-4} + Y_{t-5} \end{aligned}$$

จำนวนครั้งที่หาผลต่างและผลต่างฤดูกาลที่กำหนดด้วย d และ D ตามลำดับ ขึ้นกับว่าอนุกรมเวลาใหม่ที่สร้างขึ้นเป็นสเตชันนารีแล้วหรือไม่

2.1.4.4 ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองและค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วน

ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองหรือ Autocorrelation Function (ACF) คือ ความสัมพันธ์ของข้อมูลระหว่างค่า Y_t กับค่า Y_t ที่ห่างกัน k ค่า หรือความสัมพันธ์ระหว่างค่า Y_t กับ Y_{t-k} โดย ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองของตัวอย่าง แทนด้วย r_k และค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองของประชากร แทนด้วย ρ_k

ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนหรือ Partial Autocorrelation Function (PACF) เป็นค่าที่ใช้วัดความสัมพันธ์ระหว่าง Y_t กับค่า Y_{t+k} โดยไม่มีอิทธิพลของ $Y_{t+1} \dots Y_{t+k-1}$ เข้ามาเกี่ยวข้อง ดังตารางที่ 2.1 แสดงลักษณะของ ACF และ PACF ของตัวแบบ SARIMA

โดย ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนของตัวอย่าง แทนด้วย r_{kk} และค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วนของประชากร แทนด้วย ρ_{kk}

ตารางที่ 2.1 ลักษณะของ ACF และ PACF ของตัวแบบ SARIMA

รูปแบบ	ลักษณะของ ACF	ลักษณะของ PACF
$SAR(1)_{30}$ * กรณีฤดูกาลเกิดขึ้น 30 วัน $S=30$	$\rho_{30}, \rho_{60}, \dots$ ลดลงเร็วใกล้ 0	ρ_{30}, ρ_{30} โดยที่ $\rho_{kk} = 0$ สำหรับ $k = 60, 90, \dots$
$SAR(2)_{30}$	$\rho_{30}, \rho_{60}, \dots$ ลดลงเร็วใกล้ 0	ρ_{30}, ρ_{30} และ $\rho_{60}, \rho_{60} \neq 0$ โดยที่ $\rho_{kk} = 0$ สำหรับ $k = 90, 120, \dots$
$SMA(1)_{30}$	$\rho_{30} \neq 0$ โดยที่ $\rho_k = 0$ สำหรับ $k = 60, 90, \dots$	$\rho_{30}, \rho_{30}, \rho_{60}, \rho_{60}$ ลดลงเร็ว ใกล้ 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานภายในเท่านั้น ไม่ควรเผยแพร่ภายนอกโดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปแบบ	ลักษณะของ ACF	ลักษณะของ PACF
$SMA(2)_{30}$	ρ_{30} และ $\rho_{60} \neq 0$ โดยที่ $\rho_k = 0$ สำหรับ $k = 90, 120, \dots$	$\rho_{30}, \rho_{30}, \rho_{60}, \rho_{60}, \dots$ ลดลงเร็วใกล้ 0

2.1.4.5 ตัวแบบพยากรณ์ด้วยวิธี Box - Jenkins สำหรับอนุกรมเวลา

1. ตัวแบบ Autoregressive Moving Average (ARMA)

รูปแบบ ARMA เป็นรูปแบบที่นำรูปแบบ AR (p) ผสานกับรูปแบบ MA (q) สร้างรูปแบบ ARMA (p, q) เมื่อกำหนดอันดับ p และ q สามารถกำหนดสมการตามรูปแบบ ARMA (p, q) โดยสามารถแสดงดังสมการที่ 2.4

$$Y_t = \theta_0 + \phi_1 y_{t-1} + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.4)$$

โดยทั่วไป รูปแบบ ARMA(p, q) จะมีอันดับ $p + q \leq 2$ เช่น รูปแบบ ARMA (1, 1) ที่มีรูปแบบ โดยสามารถแสดงดังสมการที่ 2.5

$$Y_t = \theta_0 + \phi_1 y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad \text{เมื่อ } p = 1, q = 1 \quad (2.5)$$

เมื่อกำหนดตัวดำเนินการย้อนกลับ B จะสามารถเขียนสมการที่ 2.4 ในรูปแบบของตัวดำเนินการย้อนกลับ โดยสามารถแสดงดังสมการที่ 2.6

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = \theta_0 + (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t \quad (2.6)$$

2. ตัวแบบ Autoregressive Integrated Moving Average Model (ARIMA)

Box et al. (2018) กล่าวว่า รูปแบบ ARIMA เป็นรูปแบบที่ใช้กับอนุกรมเวลาที่ไม่เป็นสเตชันนารี เนื่องจากแนวโน้มต้องแปลงอนุกรมเวลาเดิม $\{Y_t\}$ ที่ไม่เป็นสเตชันนารี ให้เป็นอนุกรมเวลาใหม่ $\{z_t\}$ ที่เป็นสเตชันนารี ด้วยการหาผลต่างระหว่างค่าสังเกตในอนุกรมเวลาเดิมนั้นคือ $Z_t = \nabla^d Y_t$ ซึ่งกำหนดได้ด้วยการใช้รูปแบบ ARIMA (p, d, q) โดยสามารถแสดงดังสมการที่ 2.7

$$\phi(B) Y_t = \phi(B) \nabla^d Y_t = \theta_0 \theta(B) \varepsilon_t \quad (2.7)$$

$$\text{เมื่อ } \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad ; \theta_0 = \text{ค่าคงที่}$$

โดยกำหนดให้

D แทน ระยะห่างของคาบเวลาที่น่ามาหาผลต่างระหว่างค่าสังเกต

P แทน อันดับของรูปแบบ AR

q แทน อันดับของรูปแบบ MA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้จากสมการที่ 2.7 และเทอมของสมการมีความหมายดังนี้ในการดำเนินการคำนวณว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) $\phi(B)$ เรียกว่า ตัวดำเนินการออโตรีเกรสซีฟ (Autoregressive operator) เมื่ออนุกรมเวลาเป็นสเตชันนารี นั่นคือ รากของสมการ $\phi(B) = 0$

2) $\phi(B) = \phi(B)\nabla^d$ เรียกว่า ตัวดำเนินการเจนเนอร์ไลซ์ออโตรีเกรสซีฟ (Generalized autoregressive operator) เป็นตัวดำเนินการไม่เป็นสเตชันนารี เมื่อ d คือ ระยะห่างของคาบเวลาที่นำมาหาผลต่างระหว่างค่าสังเกต ที่ทำให้รากของสมการ $\phi(B) = 0$

3) $\theta(B)$ เรียกว่า ตัวดำเนินการมูฟวิงเอเวอร์เรจ (Moving average operator) เมื่ออนุกรมเวลาเป็นอินเวอร์ติเบิล นั่นคือ รากของสมการ $\theta(B) = 0$

3. ตัวแบบ Seasonal Autoregressive Moving Average (SARMA)

รูปแบบ SARMA เป็นรูปแบบที่ใช้สำหรับอนุกรมเวลาที่มีการเคลื่อนไหวเนื่องจากฤดูกาลจะกำหนดรูปแบบ SARMA ด้วยรูปแบบ $SARMA(P, Q)$, โดยมีส่วนประกอบ 2 ส่วน ได้แก่ ส่วนที่ 1 คือ รูปแบบ $AR(P)$, ที่ค่าสังเกต Y_t ขึ้นอยู่กับค่าสังเกตก่อนหน้าที่ล่าช้าไป $P \cdot S$ คาบเวลา ส่วนที่ 2 คือ รูปแบบ $MA(Q)$, ที่ค่าสังเกต Y_t ขึ้นอยู่กับค่าคลาดเคลื่อนก่อนหน้าที่ล่าช้าไป $Q \cdot S$ คาบเวลา ดังนั้นสามารถกำหนดรูปแบบ $SARMA(P, Q)$, โดยสามารถแสดงดังสมการที่ 2.8

$$Y_t = \theta_0 \phi_s Y_{t-s} + \theta_s \varepsilon_{t-s} \quad (2.8)$$

และสามารถกำหนดตัวแบบในรูปของตัวดำเนินการย้อนกลับ ได้ดังสมการที่ 2.9

$$\phi_p(B^s) Y_t = \theta_0(B^s) \varepsilon_t \quad (2.9)$$

$$\text{เมื่อ } \phi_p(B^s) = 1 - \phi_s B^s - \phi_{2s} B^{2s} - \dots - \phi_{ps} B^{ps}$$

$$\theta_p(B^s) = 1 - \theta_s B^s - \theta_{2s} B^{2s} - \dots - \theta_{ps} B^{ps} \quad ; \theta_0 = \text{ค่าคงที่}$$

โดยที่ S หมายถึงจำนวนคาบเวลาใน 1 ฤดูกาล

P หมายถึง จำนวนฤดูกาลล่าช้า

Q หมายถึง จำนวนฤดูกาลล่าช้า

4. ตัวแบบ Seasonal Autoregressive Integrated Moving Average (SARIMA)

รูปแบบ SARIMA เป็นรูปแบบที่ใช้สำหรับอนุกรมเวลาที่ไม่เป็นสเตชันนารีเนื่องจากฤดูกาล โดยต้องแปลงอนุกรมเวลาเดิม $\{Y_t\}$ ที่ไม่เป็นสเตชันนารี ให้เป็นอนุกรม

เวลาใหม่ $\{z_t\}$ ที่เป็นสเตชันนารี ด้วยการหาผลต่างฤดูกาล ซึ่ง $Z_t = \nabla^D Y_t = (1 - B^D)$ โดยสามารถกำหนดรูปแบบ $SARIMA(P, D, Q)$, ด้วยรูปแบบในรูปแบบเต็มแสดงดังสมการที่ 2.10

$$Y_t = \theta_0 + (1 + \phi_s)Y_{t-s} - \phi_s Y_{t-2s} + \varepsilon_t - \theta_s \varepsilon_{t-12} \quad (2.10)$$

และสามารถกำหนดตัวแบบในรูปของตัวดำเนินการย้อนกลับ ได้ดังสมการที่ 2.11

$$\phi_p(B^{DS})(1 - B^S)Y_t = \theta_0 + \theta_Q(B^S)\varepsilon_t \quad (2.11)$$

$$\text{เมื่อ } \phi_p(B^S) = 1 - \phi_s B^S - \phi_{2s} B^{2s} - \dots - \phi_{PS} B^{PS}$$

$$\theta_p(B^S) = 1 - \theta_s B^S - \theta_{2s} B^{2s} - \dots - \theta_{PS} B^{PS} \quad ; \theta_0 = \text{ค่าคงที่}$$

โดยที่ S หมายถึง จำนวนคาบเวลาใน 1 ฤดูกาล

D หมายถึง จำนวนครั้งของการหาผลต่างฤดูกาล

2.1.4.6 ขั้นตอนการพยากรณ์โดยวิธีของ Box - Jenkins

ขั้นตอนการพยากรณ์โดยวิธีของ Box - Jenkins มี 4 ขั้นตอน ดังนี้

1. การตรวจสอบข้อมูล

ตรวจสอบลักษณะข้อมูล เช่น มีส่วนประกอบของแนวโน้มหรือมีความแปรผันของฤดูกาลหรือไม่ โดยพิจารณาจากฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF)

2. การกำหนดตัวแบบ

Akaike (1978) กล่าวว่า เป็นการหารูปแบบที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลา โดยพิจารณาจากเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Akaike's Information Criterion: AIC) เป็นเกณฑ์ที่พิจารณาจากการประมาณความคลาดเคลื่อนรวมเข้ากับข้อสนเทศ (Information) ของค่าสังเกตและใช้แนวคิดค่าต่ำสุดของคูลแบ็ค - โลทท์เบอร์ (Kullback - Lieber) เพื่อนำมาใช้ในการปรับค่าประมาณของการพยากรณ์ให้มีความแม่นยำมากขึ้น โดยตัวแบบที่ให้ค่า AIC ต่ำสุด จะเป็นตัวแบบที่ดีที่สุด สมการ AIC สามารถเขียนได้ดังสมการที่ 2.12

$$AIC = n \log \sigma^2 + 2(p+1) \quad (2.12)$$

โดยที่ n คือ ขนาดตัวอย่าง

P คือ จำนวนพารามิเตอร์

σ^2 คือ $\frac{SSE}{n}$

$$\text{และ } SSE = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. การประมาณค่าพารามิเตอร์

จินตนาส (2554) กล่าวว่า เป็นการหาความสัมพันธ์ของฟังก์ชันสหสัมพันธ์กับพารามิเตอร์ซึ่งเป็นความสัมพันธ์สำหรับ อนุกรมเวลาแต่ละรูปแบบการประมาณค่าพารามิเตอร์โดยใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Square : OLS) ซึ่งสามารถใช้โปรแกรมสำเร็จรูปคำนวณในการกำหนดค่าเบื้องต้น จนกว่าจะได้ตัวประมาณที่ได้ความคลาดเคลื่อนน้อยที่สุด

4. การตรวจสอบความเหมาะสมของตัวแบบ

พิจารณาว่า $\rho_k(e_t) = 0$ สำหรับ $k = 1, \dots, m$ หรือไม่ นั่นคือพิจารณาว่าค่าคลาดเคลื่อนของการพยากรณ์ที่อยู่ห่างกัน $1, \dots, m$ ช่วงเวลาเป็นอิสระกันหรือไม่ การทดสอบสมมติฐานจะทำโดยใช้การทดสอบของ Box และ Pierce หรือการทดสอบของ Box และ Ljung ซึ่งต่างก็เป็นการทดสอบสมมติฐานที่กำหนดสมมติฐานหลักและสมมติฐานทางเลือกเหมือนกัน มีช่วงวิกฤติเดียวกัน แต่ตัวสถิติทดสอบต่างกัน

1) สมมติฐานในการทดสอบ

$$H_0 : \rho_1(e_t) = \dots = \rho_m(e_t) = 0$$

$$H_1 : \rho_k(e_t) \text{ อย่างน้อยหนึ่งค่าไม่เท่ากับ } 0 \text{ สำหรับ } k = 1, \dots, m$$

2) ระดับนัยสำคัญ $\alpha = 0.05$

3) สถิติทดสอบ

Box – Pierce

$$Q = n \sum_{m=k=1}^m r_k^2(e_t) \quad (2.13)$$

Box – Ljung

$$Q_m = n(n+2) \sum_{k=1}^m \frac{r_k^2(e_t)}{(n-k)} \quad (2.14)$$

โดยกำหนดให้

n คือ ขนาดของอนุกรมเวลา $\{e_t\}$

m คือ ช่วงเวลาห่างสูงสุดของ e_t ในอนุกรมเวลา $\{e_t\}$ ที่นำมาพิจารณา

Q คือ จำนวนพารามิเตอร์ทั้งหมดในรูปแบบ ซึ่งรวมทั้ง θ_0 ด้วย

r_k คือ ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ตัวอย่างอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลง4) เมื่อการตัดสินใจอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปฏิเสธ H_0 ถ้า Q หรือ $Q_m \geq \chi_{a,m-a}^2$

ยอมรับ H_0 ถ้า Q หรือ $Q_m < \chi_{a,m-a}^2$

ถ้าปฏิเสธ H_0 จะสรุปได้ว่ามี $\rho_k(e_i)$ อย่างน้อยหนึ่งค่าไม่เท่ากับ 0 สำหรับ $k = 1, \dots, m$ นั่นคือมีสหสัมพันธ์ในตนเองระหว่างค่าความคลาดเคลื่อนของการพยากรณ์ที่อยู่ห่างกัน k ค่า หรือสรุปได้ว่ารูปแบบที่กำหนดให้กับอนุกรมเวลายังไม่เหมาะสม

2.1.5 ตัวแบบการเรียนรู้ของเครื่อง (Machine Learning)

Witten et al. (2005) กล่าวว่า การเรียนรู้ของเครื่อง (Machine Learning) จัดเป็นสาขาหนึ่งของปัญญาประดิษฐ์ที่พัฒนามาจากการศึกษาการรู้จำรูปแบบ เกี่ยวข้องกับการทำนายข้อมูล อัลกอริทึมจะทำงานโดยอาศัยโมเดลที่ศึกษาเรียนรู้และสร้างมาจากชุดข้อมูลตัวอย่างที่มีอยู่จำนวนมาก เพื่อสร้างโมเดลการทำนายหรือตัดสินใจในภายหลัง โดยการเรียนรู้ของเครื่องสามารถแบ่งได้เป็น 3 ประเภท ตามประเภทของข้อมูลฝึกสอนได้ดังนี้ คือ

2.1.5.1 ประเภทของการเรียนรู้ของเครื่อง

1) การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือ มีข้อมูลตัวอย่างและผลลัพธ์ที่ผู้สอนต้องการถูกป้อนเข้าสู่คอมพิวเตอร์ เป้าหมายคือการสร้างกฎทั่วไปที่สามารถเชื่อมโยงข้อมูลขาเข้ากับขาออกได้ เป็นเทคนิคการเรียนรู้ของเครื่องซึ่งสร้างฟังก์ชันจากข้อมูลสอน (Training Data) ข้อมูลสอนประกอบด้วย วัตถุนำเข้า และผลที่ต้องการ ผลจากการเรียนรู้จะเป็นฟังก์ชันที่อาจจะให้ค่าต่อเนื่องหรืออาจจะเรียกวิธีการว่า การถดถอย (Regression) หรือใช้ทำนายประเภทของวัตถุ อีกอย่างเรียกว่า การแบ่งประเภท (Classification) ภารกิจของเครื่องเรียนรู้แบบมีผู้สอน คือการทำนายค่าของฟังก์ชันจากวัตถุนำเข้าที่ถูกต้อง โดยใช้ตัวอย่างในการสอนโดยใช้ตัวอย่างในการสอนโดยใช้ข้อมูลนำเข้าจำนวนมาก (Training Set) และผลที่เป็นเป้าหมาย

2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่อง โดยการสร้างโมเดลที่เหมาะสมกับข้อมูล การเรียนรู้แบบนี้แตกต่างจากการเรียนรู้แบบมีผู้สอน จัดเป็นเทคนิคการเรียนรู้อีกรูปแบบหนึ่งของการเรียนรู้ของเครื่อง โดยการใช้ชุดข้อมูลที่มีทั้งหมดในการให้เครื่องทำการเรียนรู้ โดยไม่จำเป็นต้องระบุผลลัพธ์ที่ต้องการ หรือชนิดประเภทข้อมูลไว้ก่อนที่จะให้เครื่องทำการเรียนรู้ การเรียนรู้แบบนี้จะทำหน้าที่แยกผลที่ต้องการหรือประเภทให้เองจากชุดข้อมูลที่เราใช้ในการเรียนรู้ การเรียนรู้แบบนี้ จะพิจารณาวัตถุเป็นเซตของตัวเองแล้วจึงสร้างโมเดลความหนาแน่นร่วมของชุดข้อมูลเพื่อพิจารณา

3) การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นเทคนิคหนึ่งที่เรียนรู้จากสิ่งแวดล้อมรอบตัวเอง คือเรียนรู้จากธรรมชาติรอบตัวที่มีอยู่ในชีวิตจริงนำมาดัดแปลงให้คอมพิวเตอร์ ตัวอย่างเช่น การเล่นเกมหมากรุก เราจะต้องมีการทำนายล่วงหน้าว่าจะมีโอกาสสามารถเกิดอะไรขึ้นได้บ้าง ซึ่งการเดินแต่ละครั้งอาจจะไม่เป็นผลดีต่อครั้งนั้นแต่อาจมีผลดีในครั้งถัดจากนั้นก็ได้อีกหรือกล่าวอีกอีกนัยหนึ่งคือการเรียนรู้แบบเสริมกำลังจะพิจารณาว่า เอเยนต์ควรจะมี

การกระทำใดในสิ่งแวดล้อมใดเพื่อที่จะได้รางวัลสูงสุด อัลกอริทึมของการเรียนรู้แบบเสริมกำลังนี้พยายามหานโยบายที่เชื่อมโยงสถานะของโลกเข้ากับการกระทำที่เอเจนต์ควรจะทำในสถานะนั้นๆ การเรียนรู้แบบเสริมกำลังนี้มีความแตกต่างไปจากการเรียนรู้แบบมีผู้สอนตรงที่ว่า คอมพิวเตอร์จะไม่รู้เลยว่าอะไรถูกอะไรผิด กล่าวคือ ไม่มีการบอกอย่างชัดเจนว่าการกระทำใดยังไม่ดี แต่อาศัยการเรียนรู้สะสมองค์ความรู้เสริมทบทวนไปเรื่อยๆ

2.1.5.2 การพยากรณ์โดยใช้เทคนิคแรนดอมฟอเรสต์ (Random Forest)

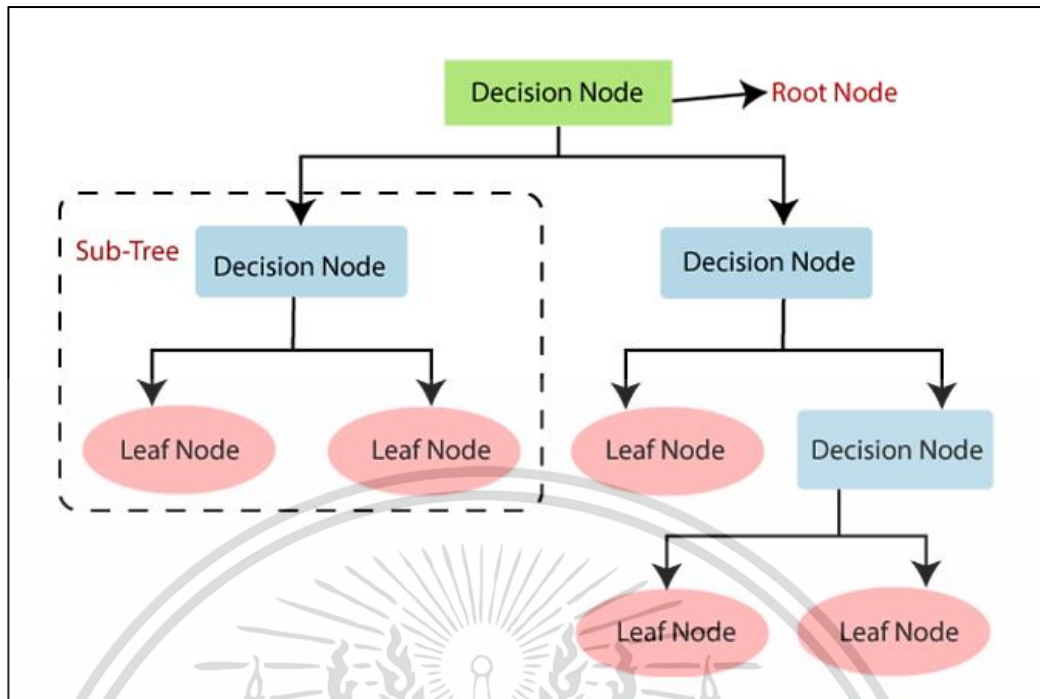
อัลกอริทึม Random Forest เป็นอัลกอริทึมประเภทหนึ่งของการเรียนรู้แบบ Supervised learning โดย Random Forest ซึ่งมีฐานความคิดจากอัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) ที่เน้นการสร้างโมเดลด้วยวิธีการ Decision Tree ขึ้นมาหลายๆโมเดลที่ไม่ซ้ำกัน โดยวิธีการสุ่มตัวแปรแล้วนำผลที่ได้แต่ละโมเดลมารวมกันพร้อมนับจำนวนผลที่มีจำนวนซ้ำกันมากที่สุด สกัดออกมาเป็นผลลัพธ์สุดท้าย ซึ่งข้อดีของวิธีการนี้คือให้ผลการพยากรณ์ที่แม่นยำและเกิดปัญหา Overfitting หรือการที่ตัวแบบตอบสนองต่อการรบกวน (noise) จำนวนมาก จนเริ่มเรียนจากการรบกวนและรายละเอียดของข้อมูลที่ไม่ถูกต้อง ทำให้ตัวแบบไม่เหมาะสมสำหรับการพยากรณ์ข้อมูลนั้นน้อยลง

เบื้องต้นสามารถสรุปพื้นฐานที่สำคัญ อย่างอัลกอริทึมต้นไม้ตัดสินใจ ได้ดังนี้ ต้นไม้ตัดสินใจ เป็นหนึ่งใน การเรียนรู้ของเครื่อง (Machine Learning) อัลกอริทึมที่ใช้กันอย่างแพร่หลาย เรียกว่าแผนภูมิต้นไม้ (tree-based model) ตัวแบบเริ่มต้นที่ใช้กันชื่อว่า CART ย่อมาจาก “Classification and Regression Trees” หรือเรียกสั้นๆว่า “Decision Tree” ซึ่งถือว่าเป็นโมเดลที่สร้างไม่ยาก อธิบายง่าย (interpret ability) แต่ความแม่นยำ (Accuracy) ไม่สูงนักเมื่อเทียบกับโมเดลรูปแบบอื่นๆ โดยจะเรียก DT ว่า Weak Classifier

Lee et al. (2006) กล่าวว่า รูปแบบของลักษณะโครงสร้างต้นไม้ประกอบไปด้วย โหนด (node) ซึ่งแต่ละโหนดจะมีเงื่อนไขของคุณลักษณะ (Feature) เป็นตัวทดสอบกิ่งของต้นไม้ (branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือกทดสอบและใบ (leaf) เป็นสิ่งที่อยู่ล่างสุดของต้นไม้แสดงถึงกลุ่มของข้อมูล (class) ก็คือผลลัพธ์ที่ได้จากการพยากรณ์ ส่วนเกณฑ์การเลือกคุณลักษณะเพื่อนำมาเป็นโหนดของต้นไม้ นั้นมาจากการคำนวณค่าเกนสารสนเทศ (Information Gain) โดยพิจารณาคุณลักษณะที่มีค่าเกนสารสนเทศหรือค่าเอนโทรปี (Entropy) ต่ำ หมายความว่า คุณลักษณะนั้นมีความสามารถในการจำแนกหมวดหมู่สูง

Qiang et al. (2002) กล่าวว่า อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) คือ การนำเข้าข้อมูลมาสร้างแบบจำลองการพยากรณ์ที่มีลักษณะคล้ายกับต้นไม้ จะมีการสร้างกฎต่าง ๆ ขึ้นเพื่อใช้ในการตัดสินใจ ซึ่งต้นไม้ตัดสินใจนั้นมีการทำงานแบบ Supervised Learning คือ สามารถสร้างแบบจำลองการจัดหมวดหมู่จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ก่อนล่วงหน้า เรียกว่า Training Set ได้อัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้แสดงอัลกอริทึมต้นไม้ตัดสินใจได้ดังรูปที่ 2.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ พังสั่น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 อัลกอริทึมต้นไม้ตัดสินใจ

Breima (2001) กล่าวว่า อัลกอริทึม Random Forest เป็นอัลกอริทึมประเภทหนึ่งของต้นไม้ตัดสินใจที่มีลักษณะแบบไม่ตัดแต่งกิ่ง (Unpruned) หรือต้นไม้ถดถอย (Regression Trees) ซึ่งถูกสร้างจากการนำข้อมูลฝึกสอนไปสุ่มเลือกตัวอย่างข้อมูล และคุณลักษณะข้อมูลแล้วนำมาสร้างเป็นต้นไม้ตัดสินใจ ซึ่งมีตัวอย่างส่วนหนึ่งที่ไม่ถูกเลือกจะถูกนำมาใช้ในการทดสอบต้นไม้ตัดสินใจ เรียกข้อมูลส่วนนี้ว่า Out-of-Bag (OOB) ซึ่งวิธีการนี้เรียกว่า Bagging ผลลัพธ์ที่ได้จะอิสระจากต้นไม้ตัดสินใจแต่ละต้นถูกนำมาคิดเป็นผลการโหวตที่มากที่สุด อัลกอริทึม Random Forest ไม่จำเป็นต้องมีข้อมูลทดสอบเพื่อประมาณความผิดพลาดเพราะข้อมูล OOB นั้นถูกนำมาใช้ทดสอบต้นไม้ตัดสินใจแล้ว

หลักการทำงานของอัลกอริทึม Random Forest

การทำงานของ Random Forest จะใช้เทคนิควิธีที่เรียกว่า เทคนิค Ensemble ซึ่งเป็นการนำหลาย ๆ แบบจำลองมาช่วยกันหาคำตอบ โดยเริ่มต้นจากการนำชุดข้อมูลที่ใช้เรียนรู้ (Training Data) มาสร้างโมเดลต่าง ๆ โดยข้อมูลเหล่านั้นจะเป็นข้อมูลที่แตกต่างกัน โดยหลังจากได้แบบจำลองมาชุดหนึ่งแล้วจึงนำไปทำนายข้อมูลที่ยังไม่ทราบคำตอบ สำหรับการทำนายด้วยเทคนิค Ensemble ซึ่งมีหลายๆแบบจำลองนี้ แต่ละโมเดลก็จะให้คำตอบออกมา ในขั้นตอนสุดท้าย เราจะต้องนำคำตอบเหล่านั้นมารวมกันเพื่อดูว่าคำตอบไหนเหมาะสมที่สุด โดยอาจจะใช้วิธีการโหวต (Vote) เลือกคำตอบที่ตอบตรงกันมากที่สุด

Random Forest จะมีการสุ่มเฉพาะข้อมูลตัวอย่าง อีกทั้งเพิ่มการเอกสารนี้สร้างความหลากหลายของโมเดลด้วยการสุ่มแอตทริบิวต์ (หรือตัวแปรต้นในทางสถิติ) ด้วยผ่านเทคนิคนี้ ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

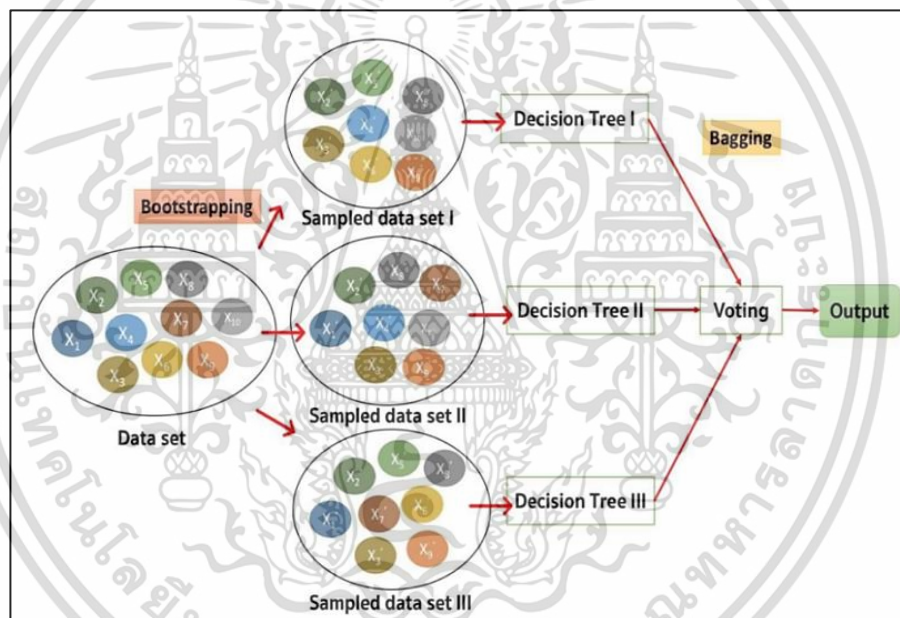
ที่ใช้ในการสร้างโมเดล Decision Tree อย่างเดียว ซึ่งเมื่อข้อมูลและแอตทริบิวต์ที่ใช้ในการสร้างโมเดลต่างกัน ทำให้โมเดลที่สร้างขึ้นมามีลักษณะที่ต่างกันด้วย

หลักการทำ Random Forest แสดงดังรูปที่ 2.4 มีรายละเอียดดังนี้

1. การสุ่มข้อมูล โดยการสุ่มข้อมูลตัวอย่าง (Bootstrapping) หรือการสร้างต้นไม้หลายๆต้นไม่ให้ซ้ำกัน จากข้อมูลทั้งหมดให้ได้ข้อมูลออกมา n ชุด ที่ไม่เหมือนกัน ตามจำนวน Decision Tree ใน Random Forest เช่น ชุดข้อมูลตั้งต้นมีอยู่ 10 feature ($X_1, X_2, X_3, X_4, \dots, X_{10}$) แต่ละ Decision Tree จะได้ feature ไปไม่เหมือนกันและจะได้ข้อมูลไม่ครบทุก row ด้วยจาก data set ทั้งหมดด้วย ($X_1 > X'_1, X_2 > X'_2, \dots$)

2. สร้างตัวแบบ Decision Tree สำหรับแต่ละชุดข้อมูล

3. ทำการรวมผลลัพธ์จากแต่ละตัวแบบ (bagging) เช่น Voting ในกรณี Classification หรือหาค่าเฉลี่ยในกรณี Regression



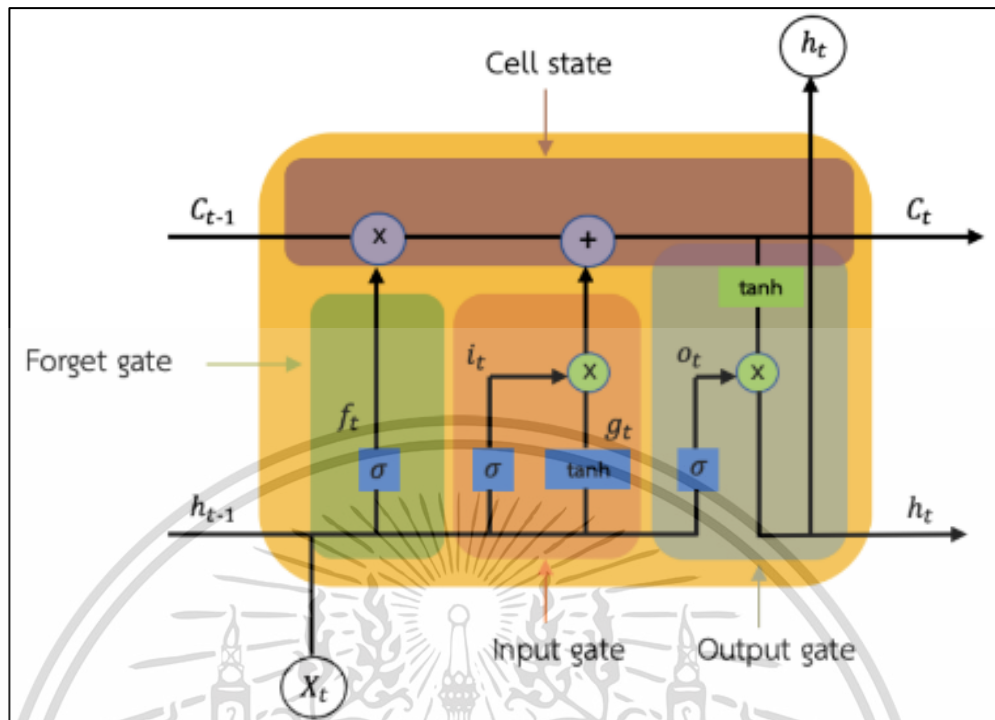
รูปที่ 2.4 หลักการทำ Random Forest จากการวิเคราะห์ข้อมูลด้วยเทคนิค Data Mining เบื้องต้น เอกสิทธิ์ (2557)

2.1.5.3 ตัวแบบหน่วยความจำระยะสั้นระยะยาว สำหรับการพยากรณ์อนุกรมเวลา

Hochreiter et al. (1997) กล่าวว่า หน่วยความจำระยะสั้น-ระยะยาว (Long Short-Term Memory: LSTM) เป็นอัลกอริทึมที่ถูกพัฒนาต่อยอดมาจาก RNN โดยทำการแก้ปัญหาในส่วนของ Gradient Vanishing ด้วยการออกแบบการทำงานในส่วนของ Cell ใหม่ ให้สามารถเก็บสถานะของการคำนวณได้ โดยใน Cell ของ LSTM นั้นมีหน่วยคำนวณย่อย ซึ่ง

ประกอบด้วย Input Gate, Forget Gate, Memory Cell State Gate และ Output Gate ดังแสดงในรูปที่ 2.5

ไม่ว่ากรณีนี้ ทั้งสั้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.5 ตัวอย่างการทำงานในเซลล์ของอัลกอริทึมโครงข่ายประสาทแบบ LSTM

Quang Hung Do et al. (2017)

Premjith et al. (2018) กล่าวว่า Input Gate เป็นหน่วยย่อยในการกำหนดข้อมูลที่จะนำเข้ามาวิเคราะห์ใน Cell โดยการรับข้อมูลเข้ามาเพื่อทำการเขียนค่าลงไปในแต่ละ Cell แสดงดังสมการที่ 2.15

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2.15)$$

เมื่อ i_t แทน ผลลัพธ์ที่ได้จาก Input

σ แทน ฟังก์ชัน Sigmoid

W_{xi} แทน ค่าน้ำหนักสำหรับคำนวณ Input ใน Input Gate

x_t แทน ค่า Input ที่นำเข้ามาคำนวณ

W_{hi} แทน ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Input Gate

h_{t-1} แทน ค่า Hidden State ที่ได้มาจากการคำนวณในหน่วยเวลาก่อนหน้า

W_{ci} แทน ค่าน้ำหนักสำหรับคำนวณ Memory Cell State ใน Input Gate

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยามให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

c_{t-1} แทน ค่า Memory Cell State จากการคำนวณใน
หน่วยเวลาก่อนหน้า

b_i แทน ค่า Bias ที่ใช้ในการคำนวณใน Input Gate

Premjith et al. (2018) กล่าวว่า Forget Gate เป็นหน่วยย่อยที่ใช้ในการกำหนดข้อมูลที่จะนำเข้ามาวิเคราะห์ใน Cell โดยทำการกำหนดว่าข้อมูลนั้นควรที่จะถูกบันทึกหรือถูกลืม โดยสามารถกำหนดได้ดังสมการที่ 2.16

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2.16)$$

เมื่อ f_t แทน ผลลัพธ์ที่ได้จาก Forget Gate

σ แทน ฟังก์ชัน Sigmoid

W_{xf} แทน ค่าน้ำหนักสำหรับคำนวณ Input ใน Forget Gate

x_t แทน ค่า Input ที่นำเข้ามาคำนวณ

W_{hf} แทน ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Forget Gate

h_{t-1} แทน ค่า Hidden State ที่ได้มาจากการคำนวณในหน่วย
เวลาก่อนหน้า

W_{cf} แทน ค่าน้ำหนักสำหรับคำนวณ Memory Cell State ใน
Forget Gate

c_{t-1} แทน ค่า Memory Cell State ที่ได้จากการคำนวณใน
หน่วยเวลาก่อนหน้า

b_f แทน ค่า Bias ที่ใช้ในการคำนวณใน Forget Gate

Memory Cell State Gate เป็นหน่วยย่อยในการกำหนดข้อมูลที่จะนำเข้ามาวิเคราะห์ใน Cell และทำการคำนวณค่าสถานะ เพื่อใช้ในการคำนวณในครั้งถัดไป โดยสามารถแสดงดังสมการที่ 2.17

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (2.17)$$

เมื่อ c_t แทน ค่า Memory Cell State ในช่วงหน่วยเวลา

f_t แทน ผลลัพธ์ที่ได้จาก Forget Gate

c_{t-1} แทน ค่า Memory Cell State จากหน่วยเวลาก่อนหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงชื่อของเอกสารทุกครั้งที่มีการนำไปใช้

tanh	แทน ฟังก์ชัน Hyperbolic Tangent
W_{xc}	แทน ค่าน้ำหนักสำหรับคำนวณค่า Input จาก Memory Cell State Gate
x_t	แทน ค่า Input ที่นำเข้ามาคำนวณ
W_{hc}	แทน ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Memory Cell State Gate
h_{t-1}	แทน ค่า Hidden State ที่ได้จากการคำนวณในหน่วยเวลาก่อนหน้า
b_c	แทน ค่า Bias ที่ใช้ในการคำนวณใน Forget Gate

Premjith et al. (2018) กล่าวว่า Output Gate เป็นหน่วยย่อยสำหรับการคำนวณ Output ของ Cell ซึ่งผลลัพธ์ที่ได้จาก Cell นี้มีอยู่ 2 อย่าง ได้แก่ Output และ Hidden State สำหรับใช้ในการคำนวณครั้งถัดไป โดยมีสมการดังสมการที่ 2.18 และ 2.19

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \quad (2.18)$$

$$h_t = o_t \tanh(c_t) \quad (2.19)$$

เมื่อ o_t แทน ผลลัพธ์ที่ได้จาก Output Gate

σ แทน ฟังก์ชัน Sigmoid

W_{xo} แทน ค่าน้ำหนักสำหรับคำนวณ Input ใน Output Gate

x_t แทน ค่า Input ที่นำเข้ามาคำนวณ

W_{ho} แทน ค่าน้ำหนักสำหรับคำนวณ Hidden State ใน Output Gate

h_{t-1} แทน ค่า Hidden State ที่ได้จากการคำนวณในหน่วยเวลาก่อนหน้า

W_{co} แทน ค่าน้ำหนักสำหรับคำนวณ Memory Cell State ใน Output Gate

c_{t-1} แทน ค่า Memory Cell State ที่ได้จากการคำนวณในหน่วยเวลาก่อนหน้า

b_o แทน ค่า Bias ที่ใช้ในการคำนวณใน Output Gate

h_t แทน ค่า Hidden State จากการคำนวณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.6 เกณฑ์ที่ใช้สำหรับการวัดค่าความคลาดเคลื่อนของตัวแบบพยากรณ์

การเลือกวิธีพยากรณ์ที่ดีที่สุดและมีความเหมาะสมที่สุดจะพิจารณาจากค่าความคลาดเคลื่อนของพยากรณ์ (Forecast Errors) หรือ ค่าส่วนเหลือ (Residuals) โดยค่าความคลาดเคลื่อนจะต้องมีค่าน้อยที่สุด ตามทฤษฎีแล้ว หากเลือกตัวแบบที่เหมาะสมในการพยากรณ์แล้ว ค่าความคลาดเคลื่อนจากการพยากรณ์จะต้องมีลักษณะเป็นแบบสุ่ม (Random) และมีค่าเฉลี่ยเท่ากับศูนย์ ดังนั้นผลรวมของค่าความคลาดเคลื่อนจากการพยากรณ์ควรมีค่าเท่ากับศูนย์หรือมีค่าต่ำที่สุด เพราะค่าความคลาดเคลื่อนจากการพยากรณ์ที่ค่าบวกและลบก็จะทำการหักลบกันเอง ในทางกลับกัน ตัวแบบที่ไม่เหมาะสมกับการพยากรณ์ก็จะมีค่าความคลาดเคลื่อนที่สูง เกณฑ์ที่ใช้ในการคัดเลือกตัวแบบในการพยากรณ์ที่ดีที่สุดคือ ค่าความคลาดเคลื่อนจะต้องมีค่าน้อยที่สุด

Rajakovich and Vladimirov (2009) กล่าวว่า ร้อยละของค่าสัมบูรณ์ของค่าความคลาดเคลื่อนโดยเฉลี่ย (Mean Absolute Percentage Error : MAPE) เป็นค่าที่ใช้วัดความถูกต้องของการพยากรณ์ที่วัดจากขนาดของความคลาดเคลื่อนของการพยากรณ์เทียบกับค่าจริงเนื่องจากหน่วยค่าความคลาดเคลื่อนของตัววัดนี้วัดอยู่ในรูปร้อยละหรือเปอร์เซ็นต์ สามารถคำนวณได้จากสมการที่ 2.20

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| 100 \frac{Y_t - Y_t}{Y_t} \right| \quad (2.20)$$

โดยกำหนดให้ n แทน จำนวนข้อมูลที่ทำการทดสอบ

Y_t แทน ข้อมูลจริงที่เวลา t

\hat{Y}_t แทน ค่าพยากรณ์ที่เวลา t

Lawrence et al. (2009) กล่าวว่า สำหรับจุดเด่นของ MAPE ก็คือไม่ผูกพันกับค่าของตัวแปรที่ต้องการจะพยากรณ์แต่มีการแสดงออกมาในรูปของค่าเปอร์เซ็นต์ซึ่งง่ายต่อการเข้าใจของบุคคล ตัวอย่างเช่น หากค่า MAPE มีค่าเท่ากับ 5 เปอร์เซ็นต์ ผู้อ่านก็จะเข้าใจได้ทันทีว่าค่าความคลาดเคลื่อนจากการพยากรณ์นั้นมีค่าเท่ากับเท่าไร

Lewis-Beck Rice (1982) กล่าวว่า เกณฑ์ที่ใช้ในการตีความค่าของ MAPE ไว้ในหนังสือชื่อ Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting พิมพ์เผยแพร่ใน ปี.ศ.1982 ได้มีการพัฒนาระดับในการวัดเพื่อใช้ในการตีความค่าของ MAPE ดังต่อไปนี้

ตารางที่ 2.2 ความสัมพันธ์ระหว่างค่า MAPE กับระดับของความแม่นยำของลิวอิส (Lewis)

MAPE	ระดับของความแม่นยำ
น้อยกว่า 10%	สูง
11-20%	ดี
21-50%	สมเหตุสมผล
มากกว่า 50%	ไม่มีความแม่นยำ

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้ทางเพื่อการศึกษาเท่านั้น ไม่สามารถนำไปใช้ประโยชน์ทางการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งยังมีให้ตีแบบสงวนสิทธิ์ และต้องยังอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.7 การแทนค่าข้อมูลสูญหาย

ปัจจุบันการแทนที่ค่าข้อมูลสูญหายจะมีงานวิจัยหลายเทคนิควิธีที่ถูกนำไปใช้จึงจำเป็นต้องพิจารณาถึงแนวทางที่เหมาะสมสำหรับใช้จัดการกับข้อมูลสูญหายในทุกๆ ครั้งที่พบปัญหานี้ ซึ่งวิธีการที่ใช้สำหรับการจัดการกับข้อมูลสูญหายที่เหมาะสมสำหรับข้อมูลอนุกรมเวลา มี 3 วิธีดังนี้

2.1.7.1 วิธีค่าเฉลี่ยเคลื่อนที่อย่างง่าย (Simple Moving Average : SMA)

Colby (2003) กล่าวว่า วิธีนี้จะถ่วงน้ำหนักให้ค่าที่นำมาคำนวณมีความสำคัญเท่ากันหมด โดยการอาศัยเอาข้อมูลในช่วงเวลาหนึ่งมาหาค่าเฉลี่ย มีวิธีการคำนวณแสดงดังสมการที่ 2.21

$$SMA_t = \frac{(x_t + x_{t-1} + x_{t-2} + \dots + x_{t-n+1})}{n} \quad (2.21)$$

โดย SMA_t คือ ค่าเฉลี่ยเคลื่อนที่ ณ เวลาปัจจุบัน

n คือ จำนวนข้อมูล

x_t คือ ข้อมูลที่เลือกใช้ในการคำนวณ ณ เวลาปัจจุบัน

x_{t-n} คือ ข้อมูลที่เลือกใช้ในการคำนวณย้อนหลังไป n คาบเวลา

2.1.7.2 วิธีค่าเฉลี่ยเคลื่อนที่เอกซ์โพเนนเชียล (Exponential Moving

Average : EMA)

Colby (2003) กล่าวว่า เป็นการให้ความสำคัญกับข้อมูลล่าสุดมากที่สุด และจะเอาข้อมูลทุกข้อมูลในอดีตมาหาค่าเฉลี่ยวิธีนี้ไม่ให้ความสำคัญกับเวลาในการวิเคราะห์ แต่ค่า EMA จะให้ความสำคัญกับ Smoothing Factor (SF) โดยที่ $SF = 2/(n+1)$ ซึ่งวิธีการสร้าง EMA มีสูตรการคำนวณดังสมการที่ 2.22

$$EMA_t = EMA_{t-1} + SF(x_t - EMA_{t-1}) \quad (2.22)$$

โดยที่ EMA_t คือ ค่าของ Exponential Moving Average ณ เวลาปัจจุบัน

EMA_{t-1} คือ ค่าของ Exponential Moving Average ณ คาบเวลาก่อนหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งาน SF คือ ค่าของ Smoothing Factor
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลง x_t คือ ข้อมูลปัจจุบัน เอกสารทุกครั้งที่มีการนำไปใช้

n คือ จำนวนข้อมูล

2.1.7.3 วิธีค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก (Weight Moving Average :

WMA)

Colby (2003) กล่าวว่า เป็นการวิเคราะห์ที่ให้ความสำคัญกับข้อมูลที่ใช้คำนวณในคาบเวลาสุดท้ายมากที่สุด โดยข้อมูลถัดไปจะถูกลดความสำคัญลงไปเรื่อยๆ วิธีการคำนวณดังสมการที่ 2.23

$$WMA_t = \frac{[x_t n + x_{t-1}(n-1) + x_{t-2}(n-2) + \dots + x_{t-n+1}(1)]}{n + (n-1) + (n-2) + \dots + 2 + 1} \quad (2.23)$$

โดยที่ WMA_t คือ ค่าเฉลี่ยถ่วงน้ำหนัก ณ เวลาปัจจุบัน

x_t คือ ข้อมูลที่เลือกใช้ในการคำนวณ ณ เวลาปัจจุบัน

x_{t-n} คือ ข้อมูลที่เลือกใช้ในการคำนวณย้อนหลังไป n

คาบเวลา

n คือ จำนวนข้อมูล

2.1.8 การกำหนดค่าไฮเปอร์พารามิเตอร์

Bakhashwain et al. (2021) กล่าวว่า การกำหนดค่าไฮเปอร์พารามิเตอร์ (Hyper Parameters) เป็นขั้นตอนสำคัญในการหาค่าเหมาะสมที่สุดในการสร้างตัวแบบ จึงต้องมีการเลือกค่าไฮเปอร์พารามิเตอร์เพื่อให้ตัวแบบมีความแม่นยำที่สุด ซึ่งการปรับแต่งไฮเปอร์พารามิเตอร์ไม่ใช่เรื่องง่าย โดยเฉพาะอย่างยิ่งในอนุกรมเวลาที่ไม่มีความเป็นเชิงประจักษ์ อย่างไรก็ตามมีกลยุทธ์การคำนวณและวิธีการค้นหาที่กล่าวถึงการเพิ่มประสิทธิภาพของไฮเปอร์พารามิเตอร์ ได้แก่ วิธีการค้นหาแบบสุ่ม (Random search)

Sánchez-Il et al. (2018) กล่าวว่า วิธีการค้นหาแบบสุ่ม (Random search) เป็นการทดสอบชุดของไฮเปอร์พารามิเตอร์แบบสุ่มตัวอย่างที่มีการกำหนดจำนวนไว้ล่วงหน้าซึ่งการสุ่มตัวอย่างจะสุ่มอย่างสม่ำเสมอ

Bergstra et al. (2012) กล่าวว่า มีการแสดงให้เห็นอย่างกว้างขวางว่าการค้นหาแบบสุ่มเกี่ยวกับค่าไฮเปอร์พารามิเตอร์นั้นมีประสิทธิภาพมากกว่าในทางปฏิบัติมากกว่าการค้นหาแบบกริด (Grid Search)

2.2 งานวิจัยที่เกี่ยวข้อง

วรางคณา (2558) ศึกษาตัวแบบพยากรณ์ที่เหมาะสมกับอนุกรมเวลาปริมาณการจำหน่ายเบียร์ในประเทศไทย โดยใช้ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2543 ถึงเดือนมิถุนายน พ.ศ. 2557

จำนวน 174 ค่า สำหรับการสร้างตัวแบบพยากรณ์ด้วยวิธีบ็อกซ์-เจนกินส์ วิธีการปรับเรียบด้วยเส้นโค้งเลขชี้กำลังของวินเทอร์แบบคูณ และวิธีการพยากรณ์รวม และทำการเปรียบเทียบความแม่นยำไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุขัดแย้งและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าพยากรณ์ โดยใช้เกณฑ์เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด ผลการวิจัยพบว่า วิธีบ็อกซ์-เจนกินส์เป็นวิธีที่มีความแม่นยำมากที่สุด เนื่องจากให้ค่าพยากรณ์ที่มีความแตกต่างกับข้อมูลจริงน้อยที่สุดหรือมีค่า MAPE และ RMSE ต่ำที่สุด

Taylor (2008) ศึกษาการเปรียบเทียบการพยากรณ์อนุกรมเวลาของปริมาณสายที่โทรเข้าของศูนย์บริการลูกค้า โดยศึกษาจากกลุ่มการให้บริการ 5 กลุ่ม ของธนาคารพาณิชย์รายย่อยแห่งหนึ่งของสหราชอาณาจักร โดยทำการเปรียบเทียบ 2 วิธี คือ วิธีการของบ็อกซ์-เจนกินส์ และ วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลแบบโพลท์ - วินเทอร์ ผลการวิจัยพบว่าวิธีการของบ็อกซ์-เจนกินส์ ให้ความแม่นยำในการพยากรณ์ดีกว่าเนื่องให้ค่า MAPE ต่ำที่สุด

บุญฤทธิ์และเสาวภา (2551) ศึกษาตัวแบบพยากรณ์ที่เหมาะสมกับอนุกรมเวลาปริมาณการส่งออกมะม่วงรายเดือนของประเทศไทย โดยใช้ข้อมูลตั้งแต่เดือนมกราคม พ.ศ.2550 ถึงเดือนธันวาคม พ.ศ.2559 จำนวนทั้งสิ้น 120 ค่า สำหรับการสร้างตัวแบบพยากรณ์ด้วย วิธีแยกองค์ประกอบ วิธีโพลท์-วินเทอร์ และวิธีบ็อกซ์-เจนกินส์ และทำการเปรียบเทียบความแม่นยำของค่าพยากรณ์ โดยใช้เกณฑ์เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย และเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด ผลการวิจัยพบว่า วิธีบ็อกซ์-เจนกินส์ให้ค่า MAPE และ RMSE ต่ำที่สุด

วิชารักษ์และวัชระ (2019) ศึกษาตัวแบบจำลองทางคณิตศาสตร์สำหรับพยากรณ์ปริมาณขยะมูลฝอยในเขตพื้นที่เทศบาลเมืองบุรีรัมย์ โดยใช้ข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2544 ถึงเดือนธันวาคม พ.ศ. 2559 จำนวน 192 ค่า สำหรับศึกษาตัวแบบการพยากรณ์ โดยวิธีปรับเรียบเอ็กซ์โปเนนเชียลอย่างง่าย วิธีปรับเรียบเอ็กซ์โปเนนเชียลแบบโพลท์ วิธีปรับเรียบเอ็กซ์โปเนนเชียลแบบวินเทอร์รูปแบบการคูณและการบวกและวิธีบ็อกซ์เจนกินส์ และทำการเปรียบเทียบความแม่นยำของค่าพยากรณ์โดยใช้เกณฑ์รากที่สองของความคลาดเคลื่อนเฉลี่ยและเกณฑ์เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยผลการวิจัยพบว่าวิธีบ็อกซ์เจนกินส์ มีความแม่นยำในการพยากรณ์มากที่สุด เนื่องจากให้ค่า RMSE และ MAPE ต่ำสุด

Sima et al. (2018) ศึกษาการเปรียบเทียบระหว่าง ARIMA และ LSTM ในการพยากรณ์อนุกรมเวลา เพื่อศึกษาว่าหน่วยความจำระยะสั้นระยะยาว (LSTM) นั้นดีกว่าวิธีการพยากรณ์แบบเดิมหรือไม่ โดยใช้ข้อมูลอนุกรมเวลาทางการเงินหรือตลาดหุ้นรายเดือนจากเว็บไซต์ Yahoo ตั้งแต่เดือนมกราคม ค.ศ.1985 ถึงเดือนสิงหาคม ค.ศ.2018 จากการศึกษาพบว่าวิธี LSTM มีประสิทธิภาพดีกว่าวิธี ARIMA ซึ่งมีค่า RMSE โดยใช้ Rolling ARIMA และ Rolling LSTM คือ 11.481 และ 64.213 ตามลำดับ

Anggraeni et al. (2015) ศึกษาการพยากรณ์ปริมาณความต้องการชุดประกอบพีดีดีของเด็กของชาวมุสลิมโดยความต้องการมีแนวโน้มเพิ่มขึ้นในใกล้วันฮีดซึ่งเป็นวันเฉลิมฉลองทางศาสนาอิสลามแต่จะลดลงในเดือนหน้าซ้ำทุกปีแต่จะเกิดขึ้นในเวลาที่แตกต่างกันทุกปี โดยงานวิจัยนี้ได้เปรียบเทียบวิธีบ็อกซ์-เจนกินส์ และวิธีบ็อกซ์-เจนกินส์กับตัวแปรอิสระ X ในการเปรียบเทียบประสิทธิภาพใช้ค่าเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยามให้นำไปใช้ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

MAPE ผลจากการวิจัยพบว่า ตัวแบบของบ็อกซ์-เจนกินส์กับตัวแปรอิสระ X ให้ค่าความแม่นยำมากกว่าวิธีบ็อกซ์-เจนกินส์ แบบไม่มีตัวแปรอิสระ X

Umematsu et al. (2019) ศึกษาการปรับปรุงการพยากรณ์ความเครียดในชีวิตประจำวันของนักเรียนโดยใช้โครงข่ายประสาทเทียม LSTM โดยใช้แบบจำลองเครือข่ายประสาทเทียมหน่วยความจำระยะสั้นระยะยาว (LSTM), การถดถอยโลจิสติกส์ (LR) และวิธีซัพพอร์ตเวกเตอร์แมทซ์ (SVM) โดยใช้ข้อมูลทั้งหมด 2,276 วันโดยมีลำดับข้อมูล 8 วันที่ทับซ้อนกัน 1,231 ลำดับจากผู้เข้าร่วม 142 คน จากการศึกษาพบว่า LSTM มีประสิทธิภาพเหนือกว่า LR และ SVM ด้วยผลลัพธ์ที่ดีที่สุดถึง 83.6% โดยใช้ข้อมูลก่อนหน้า 7 วัน

Mei et al. (2014) ศึกษาการคาดการณ์ราคาแบบเรียลไทม์ในตลาดไฟฟ้านิวเจอร์กด้วยวิธี Random Forest จากการศึกษาพบว่า การคาดการณ์ราคาด้วยวิธี Random Forest ช่วยให้ผู้ใช้สามารถประมาณความเสี่ยงของกลยุทธ์การเสนอราคา นอกจากนี้ตัวแบบยังสามารถปรับให้เข้ากับเงื่อนไขการคาดการณ์ล่าสุด เช่น เงื่อนไขของตลาดตามฤดูกาลล่าสุดโดยการปรับค่าไฮเปอร์พารามิเตอร์ใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีการดำเนินงานวิจัย

การศึกษาวิจัยในครั้งนี้มุ่งเน้นการศึกษาวิธีการที่เหมาะสมสำหรับการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ ซึ่งในการวิจัยมีลำดับดังต่อไปนี้

1. การเตรียมพร้อมข้อมูล (Data Preparation)
2. การสร้างตัวแบบพยากรณ์ (Modeling)
3. การพัฒนาตัวแบบพยากรณ์ (Development)

3.1 ขั้นตอนการเตรียมข้อมูล

ขั้นตอนการจัดเตรียมข้อมูลให้พร้อมก่อนนำไปสร้างตัวแบบการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ ด้วยตัวแบบต่าง ๆ มีรายละเอียดดังต่อไปนี้

3.1.1 ลักษณะและแหล่งที่มาของข้อมูล

การศึกษานี้เป็นการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ โดยเป็นข้อมูลทุติยภูมิที่ทำการรวบรวมจากข้อมูลจำนวนการเรียกใช้เซอร์วิสรายวัน โดยใช้ข้อมูลเดือน มกราคม พ.ศ. 2564 ถึงเดือน ธันวาคม พ.ศ. 2565 เป็นข้อมูลสำหรับการสร้างตัวแบบการพยากรณ์ในวิธีการต่างๆ และใช้ข้อมูลเดือน มกราคม พ.ศ. 2566 ถึงเดือน มีนาคม พ.ศ. 2566 เป็นข้อมูลสำหรับการตรวจสอบผลของการพยากรณ์ของตัวแบบพยากรณ์ที่สร้างขึ้น

3.1.2 การจัดการข้อมูล

ผู้วิจัยต้องจัดการกับข้อมูลให้เหมาะสมก่อนที่จะนำข้อมูลไปวิเคราะห์ เพื่อใช้ในการสร้างตัวแบบการพยากรณ์ เช่น การหา outlier การสำรวจข้อมูลหรือการหาความผิดปกติของข้อมูล มีรายละเอียดดังต่อไปนี้

3.1.2.1 การแทนค่าข้อมูลสูญหาย (Data Imputation)

หลังจากทำความสะอาดข้อมูลเสร็จแล้ว ถัดไปจะเป็นขั้นตอนของการจัดการกับข้อมูลที่มีข้อมูลสูญหาย (Missing Value) เนื่องจากหากนำเข้าไปวิเคราะห์ในการสร้างตัวแบบแล้ว จะทำให้ได้ผลลัพธ์ที่ไม่ถูกต้อง จึงต้องทำการแทนที่ค่าสูญหายโดยใช้วิธีค่าเฉลี่ยเคลื่อนที่อย่างง่าย (Simple Moving Average : SMA) วิธีค่าเฉลี่ยเคลื่อนที่เอกซ์โพเนนเชียล

(Exponential Moving Average: EMA) และวิธีค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก (Weight Moving Average : WMA)

3.1.2.2 การแบ่งข้อมูล (Split Data)

ทำการแบ่งข้อมูลออกเป็น 2 ส่วน ดังนี้

- 1) ชุดข้อมูลฝึกฝน (Training Set) จำนวน 707 ข้อมูล คิดเป็นร้อยละ 88 ของข้อมูลทั้งหมด เริ่มตั้งแต่เดือน มกราคม พ.ศ.2564 ถึง ธันวาคม พ.ศ. 2565 สำหรับสร้างตัวแบบ
- 2) ชุดข้อมูลทดสอบ (Test Set) จำนวน 90 ข้อมูล คิดเป็นร้อยละ 12 ของข้อมูลทั้งหมด เริ่มตั้งแต่เดือน มกราคม พ.ศ.2566 ถึง มีนาคม พ.ศ. 2566 สำหรับประเมินตัวแบบ

3.1.2.3 การเตรียมข้อมูลบน Power BI

Power BI Desktop มีการแบ่งส่วนการทำงานออกเป็น 2 ส่วนด้วยกัน คือ

- 1) Power BI Desktop เป็นส่วนที่ใช้ในการสร้าง Visualization
- 2) Power Query เป็นส่วนที่ใช้สำหรับการบริหารจัดการเกี่ยวกับข้อมูล เช่น กระบวนการทำความสะอาดข้อมูล (Data cleansing) หรือ กระบวนการแปลงข้อมูล (Data Transformation)



รูปที่ 3.1 การเตรียมข้อมูลบน Power BI

3.2 เครื่องมือที่ใช้ในการศึกษาวิจัย

การวิเคราะห์ข้อมูลดำเนินการทดสอบด้วยการเขียนโปรแกรมภาษา Python บน Jupyter เอกสารนี้ Notebook และศึกษาไลบรารีเสริมประกอบด้วยอีกภาษาเท่านั้น ไม่นอญาติให้หน้าไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ 1) ไลบรารี pandas ใช้สำหรับจัดการข้อมูลและวิเคราะห์ข้อมูลเอกสารทุกครั้งที่มีการนำไปใช้

- 2) โลบายี่ numpy ใช้สำหรับจัดการข้อมูลและวิเคราะห์ข้อมูล
- 3) โลบายี่ statmodel ใช้สำหรับสร้างตัวแบบบ็อกซ์-เจนกินส์และทดสอบตัวแบบทางสถิติ
- 4) โลบายี่ sklearn ใช้สำหรับสร้างตัวแบบการเรียนรู้เชิงลึก วิธีหน่วยความจำระยะสั้น ระยะยาว และตัวแบบการเรียนรู้ของเครื่อง วิธี Random Forest
- 5) โลบายี่ keras ใช้สำหรับสร้างตัวแบบการเรียนรู้เชิงลึก วิธีหน่วยความจำระยะสั้น ระยะยาว
- 6) โลบายี่ matplotlib ใช้สำหรับแสดงผลในรูปแบบ Visualization

3.3 เครื่องมือที่ใช้ในการทำรีพอร์ต (Report)

- 1) 1) Microsoft Excel คือ เครื่องมือที่ใช้เป็นฐานข้อมูล
- 2) Microsoft Power BI คือ เครื่องมือที่ใช้ในการสื่อสารด้วยภาพข้อมูล (Data Visualization)
- 3) Jira Software คือ แพลตฟอร์มสำหรับติดตามปัญหาและติดตามความคืบหน้าของโครงการ

3.4 ขั้นตอนการดำเนินงานวิจัย

- 1) กำหนดปัญหาและขอบเขตของการทำวิจัย
- 2) ศึกษางานวิจัยและรวบรวมข้อมูลทั้งหมดทั้งทางทฤษฎีที่เกี่ยวข้องและงานวิจัยที่เกี่ยวข้อง
- 3) รวบรวมข้อมูล ตัวแปรต่างๆ ที่จำเป็นต่อการวิเคราะห์ข้อมูล
- 4) วิเคราะห์ข้อมูลและจัดทำ Data Visualization
- 5) ศึกษาและรวบรวมข้อมูลทางทฤษฎีและงานวิจัยที่เกี่ยวข้องเกี่ยวกับการพยากรณ์อนุกรมเวลา
- 6) รวบรวมข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสตั้งแต่วันที่ 1 มกราคม 2564 ถึง 31 ธันวาคม 2565 รวมทั้งสิ้น 707 วัน เพื่อใช้ในการสร้างตัวแบบหรือสมการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสและเก็บข้อมูลตั้งแต่วันที่ 1 มกราคม 2566 ถึง 31 มีนาคม 2566 รวมทั้งสิ้น 90 วัน เพื่อใช้ในการทดสอบค่าความคลาดเคลื่อนของการพยากรณ์แต่ละตัวแบบ
- 7) วิเคราะห์และจัดการข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส
- 8) ศึกษาและทดสอบตัวแบบต่างๆที่ใช้ในการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส
- 9) เปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ โดยพิจารณาจากค่า MAPE และเลือกตัวแบบที่เหมาะสมที่สุด
- 10) ประยุกต์ใช้ตัวแบบที่ดีที่สุดกับการพยากรณ์รายวันล่วงหน้าเป็นระยะเวลา 30 วัน
- 11) สรุปผลงานวิจัยและจัดทำรายงาน

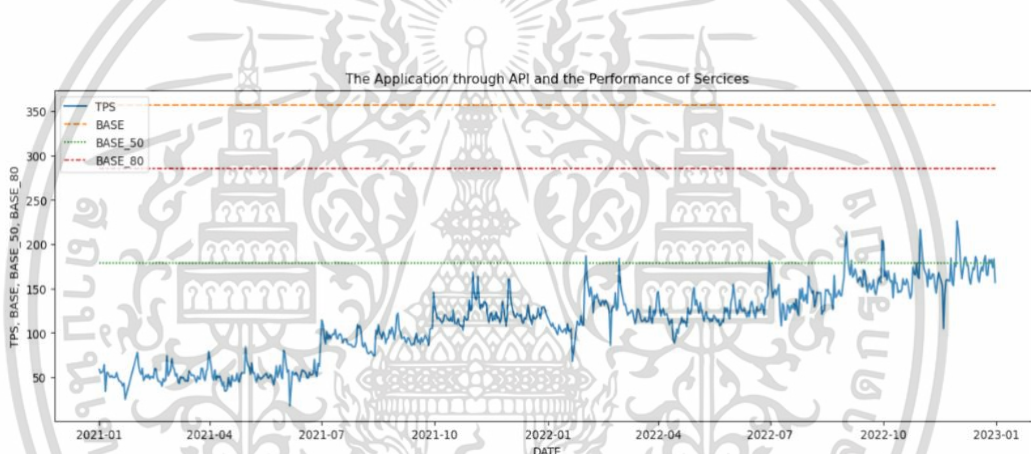
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเท่านั้น มิใช่สัญญาที่เห็นชอบโดยคณะกรรมการ
ไม่ว่ากรณีใดๆ ทั้งสิ้น ยกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.5 การวิเคราะห์ข้อมูล

เมื่อข้อมูลมีความถูกต้องจากการจัดการข้อมูล ผู้วิจัยจึงทำการวิเคราะห์ข้อมูล โดยการสร้างตัวแบบพยากรณ์รายวันนั้นจะต้องพิจารณาว่าข้อมูลที่จะพยากรณ์มีลักษณะแบบใด ดังนั้น ควรตรวจสอบลักษณะข้อมูลก่อนเป็นอันดับแรก เพื่อที่จะได้เลือกวิธีการพยากรณ์ได้ถูกต้องเหมาะสม

- 1) วิเคราะห์ข้อมูลอนุกรมเวลาว่ามีค่าเฉลี่ยคงที่และค่าความแปรปรวนคงที่หรือไม่ กล่าวคือต้องมีสมบัติเป็นสเตชันนารี
- 2) ถ้าหากข้อมูลอนุกรมเวลาไม่มีความคงที่ไม่เป็นสเตชันนารี จะต้องทำการปรับข้อมูลให้มีสมบัติเป็นสเตชันนารีก่อนนำไปใช้งาน

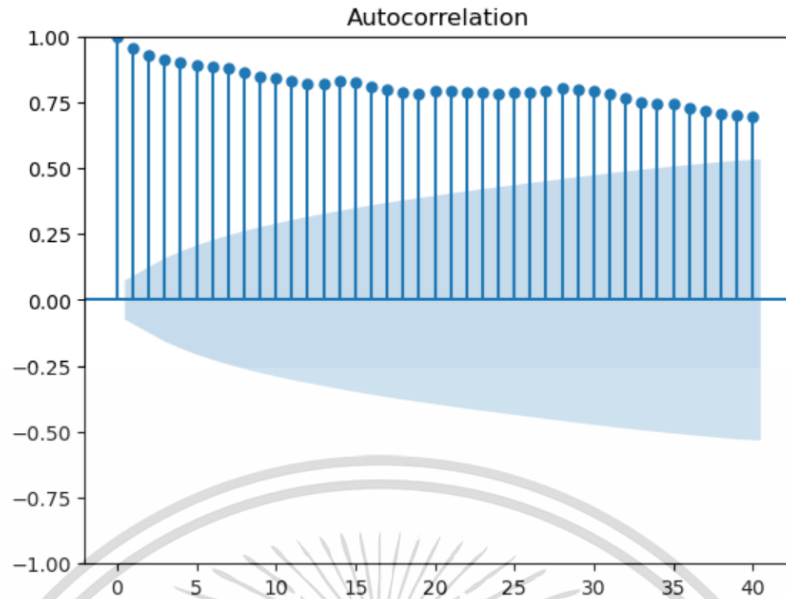
3.5.1 การวิเคราะห์ลักษณะข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสรายวัน



รูปที่ 3.2 ข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสในช่วง 1/01/2021 – 31/12/2022

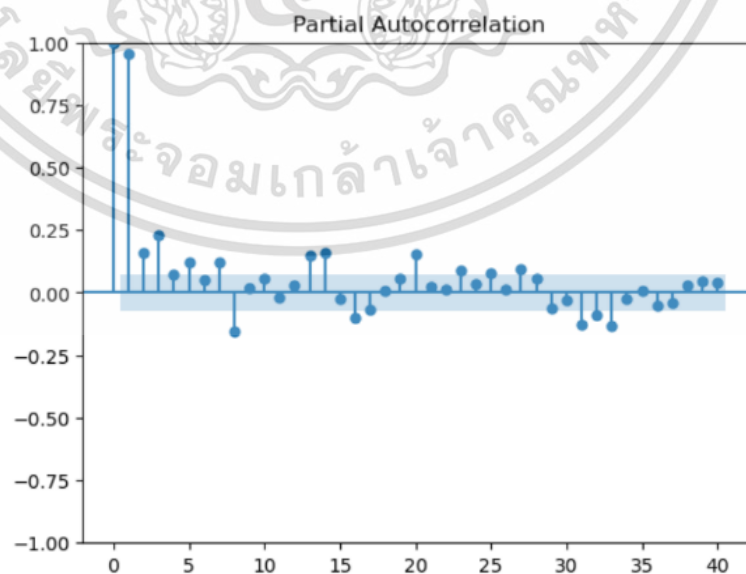
จากการพล็อตกราฟระหว่างวันที่และจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสระหว่างวันที่ 1/01/2021 ถึงวันที่ 31/12/2022 เพื่อวิเคราะห์ลักษณะข้อมูล แสดงดังรูปที่ 3.2 พบว่าข้อมูลมีลักษณะไม่คงที่ กล่าวคือ เห็นได้ว่าจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสมีความแปรผันของฤดูกาลอย่างชัดเจน และทำการทดสอบด้วยตัวทดสอบทางสถิติ โดยทดสอบแนวโน้มและทดสอบอิทธิพลของฤดูกาล พบว่าข้อมูลมีลักษณะไม่คงที่ หากอนุกรมเวลาใด ๆ ที่พิจารณามีความผันแปรทางฤดูกาลจะต้องทำการกำจัดความผันแปรทางฤดูกาลออกไปก่อน ซึ่งวิธีกำจัดความผันแปรทางฤดูกาลนั้นมีหลายวิธี ในวิจัยนี้จะใช้วิธีหาผลต่างของฤดูกาล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



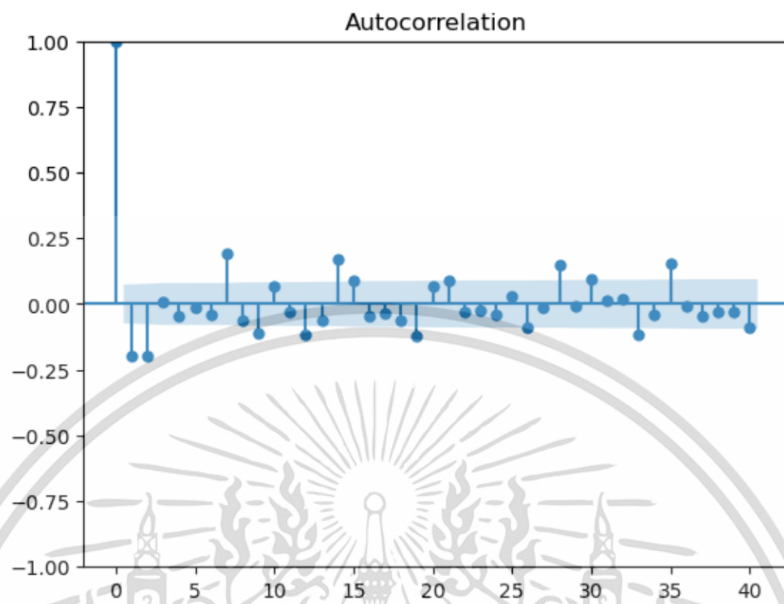
รูปที่ 3.3 ค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส

เมื่อพิจารณาจากรูปที่ 3.3 แสดงค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส พบว่าฟังก์ชันสหสัมพันธ์ในตัวเอง ลดลงเข้าสู่ศูนย์อย่างช้า ๆ และรูปที่ 3.4 พบว่ากราฟฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ลดลงเข้าสู่ศูนย์อย่างช้า ๆ แสดงว่า อนุกรมเวลาของจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสไม่คงที่ จึงทำการหาผลต่างครั้งที่ 1 ของความแปรผันเกี่ยวกับฤดูกาลของจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส

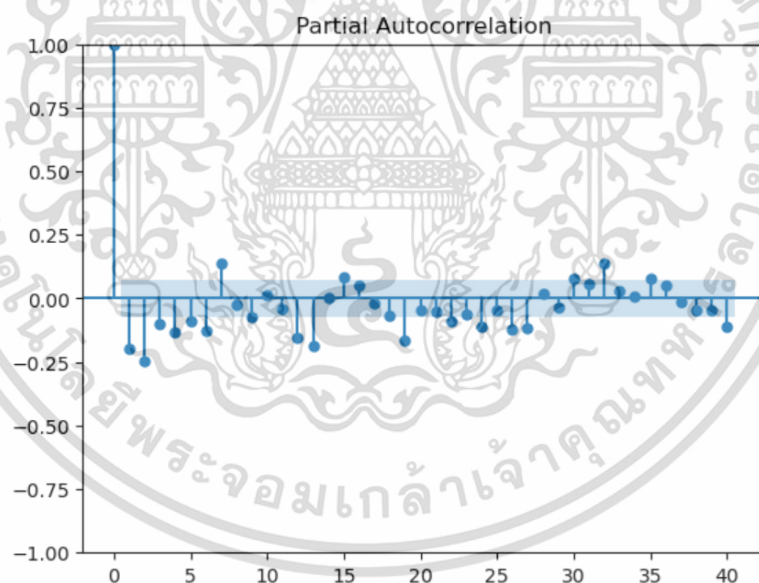


รูปที่ 3.4 ค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส

จากรูปที่ 3.5 และ 3.6 เมื่อทำการหาผลต่างครั้งที่ 1 ของความแปรผันเกี่ยวกับฤดูกาล พบว่าค่า ACF และ PACF ลดลงเข้าสู่ศูนย์อย่างรวดเร็ว แสดงว่าอนุกรมเวลาคงที่



รูปที่ 3.5 ค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของผลต่างฤดูกาลครั้งที่ 1



รูปที่ 3.6 ค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของผลต่างฤดูกาลครั้งที่ 1

3.5.2 การทดสอบความคงที่ของข้อมูล (Stationary)

เนื่องจากคุณสมบัติความคงที่ของข้อมูลเป็นสิ่งที่จำเป็นที่ข้อมูลพึงมีก่อนที่จะทำการพยากรณ์ด้วยวิธีบ็อก-เจนกินส์ ดังนั้นจึงทำการทดสอบด้วยตัวทดสอบทางสถิติ Augmented Dickey-Fuller Test (ADF) ที่ระดับนัยสำคัญ 0.05 โดยมีสมมติฐานมีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้ภายในของนักศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$H_0 : \gamma = 0 \text{ (ข้อมูลมีลักษณะไม่คงที่)}$$

$$H_1 : \gamma \neq 0 \text{ (ข้อมูลมีลักษณะคงที่)}$$

ผลที่ได้พบว่าค่า $p\text{-value} = 0.00000000000000033 < \alpha = 0.05$ ดังนั้น ข้อมูลชุดนี้มีลักษณะคงที่ ที่ระดับนัยสำคัญ 0.05

3.6 ตัวแบบพยากรณ์ที่ใช้ในงานวิจัย

3.6.1 ตัวแบบพยากรณ์วิธีบ็อก-เจนกินส์ด้วยตัวแบบ SARIMA

เนื่องจากพบว่าข้อมูลมีความแปรผันทางฤดูกาลอย่างชัดเจนทางผู้วิจัยจึงใช้การพยากรณ์วิธีบ็อก-เจนกินส์ด้วยตัวแบบ SARIMA ซึ่งเป็นตัวแบบที่เหมาะสมกับข้อมูลที่มีลักษณะดังนี้
วิธีการสร้างตัวแบบ SARIMA มีขั้นตอนดังนี้

1) การตรวจสอบข้อมูล

ตรวจสอบลักษณะข้อมูล เช่น มีส่วนประกอบของแนวโน้มและมีความแปรผันของฤดูกาลหรือไม่ โดยพิจารณาจากฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) และ ฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF)

2) การกำหนดตัวแบบ

พิจารณาเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Akaike's Information Criterion : AIC) ที่น้อยที่สุด จะให้ตัวแบบที่ดีที่สุด

3) การประมาณค่าพารามิเตอร์

ประมาณค่าพารามิเตอร์สำหรับตัวแบบที่เหมาะสม ทดสอบเชิงสถิติว่าพารามิเตอร์นั้นควรมีอยู่ในตัวแบบหรือไม่ โดยมีพารามิเตอร์ ดังนี้

ϕ_1 แทน การถดถอยในตัวเองแบบไม่มีฤดูกาลอันดับ 1

Φ_1 แทน การถดถอยในตัวเองแบบมีฤดูกาลอันดับ 1

θ_1 แทน ค่าเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับ 1

Θ_1 แทน ค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ 1

4) การตรวจสอบความเหมาะสมของตัวแบบ

ทดสอบว่าค่าความคลาดเคลื่อนมีสหสัมพันธ์ในตัวเองหรือไม่ โดยใช้ตัวสถิติ

ทดสอบ Box-Pierce and Ljung-Box Test

โดยมีสมมติฐานในการทดสอบ ดังนี้

$$H_0 : \rho_e(1) = \rho_e(2) = \rho_e(3) = \dots = \rho_e(k) = 0$$

(ค่าความคลาดเคลื่อนของตัวแบบไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลา)

$$H_1 : \rho_e(k) \text{ อย่างน้อยหนึ่งค่าไม่เท่ากับ } 0 \text{ สำหรับ } k = 1, 2, 3, \dots, m$$

(ค่าความคลาดเคลื่อนของตัวแบบมีสหสัมพันธ์ในตัวเองทุก ช่วงเวลา)

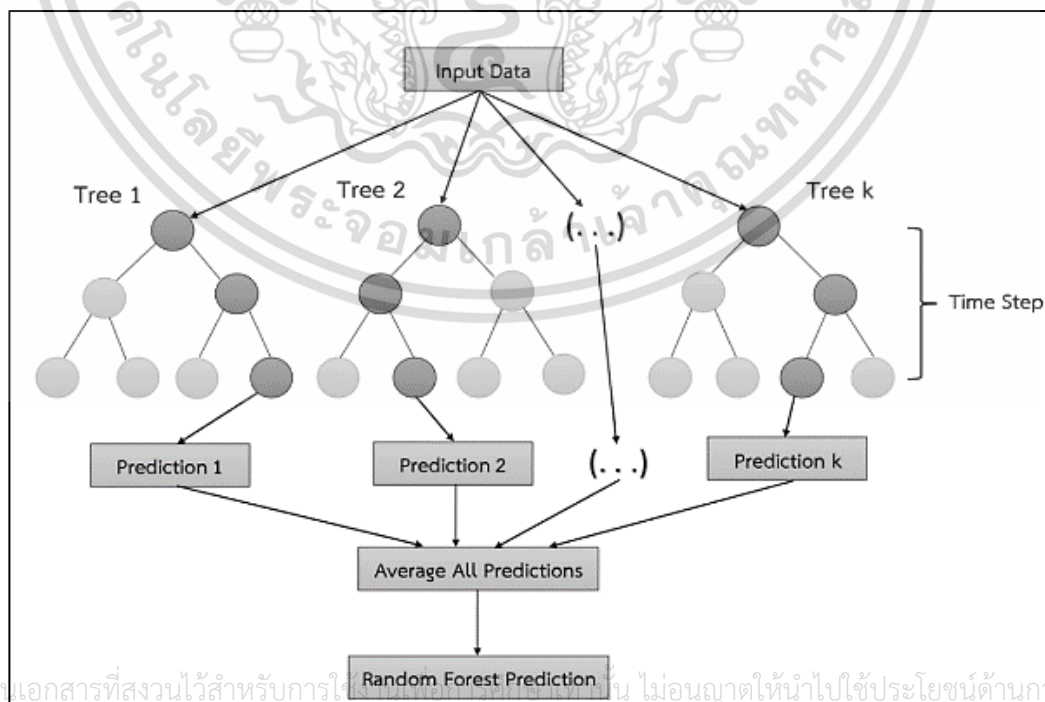
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ ถ้ายอมรับ H_0 จะสรุปได้ว่า ค่าความคลาดเคลื่อนของตัวแบบที่พิจารณาไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลา ดังนั้นตัวแบบที่ทำการทดสอบนี้เป็นตัวแบบที่เหมาะสม

3.6.2 ตัวแบบพยากรณ์วิธีบ็อก-เจนกินส์ด้วยตัวแบบ SARIMAX

- 1) การตรวจสอบข้อมูล
 - ตรวจสอบข้อมูล เช่น ข้อมูลได้รับผลกระทบมาจากปัจจัยอื่นๆ หรือไม่
- 2) การกำหนดตัวแบบ
- 3) การประมาณค่าพารามิเตอร์
 - นำตัวแบบที่เหมาะสมมาประมาณค่าพารามิเตอร์ ทดสอบเชิงสถิติว่าพารามิเตอร์นั้นควรมีอยู่ในตัวแบบหรือไม่
- 4) การตรวจสอบความเหมาะสมของตัวแบบ
 - ทดสอบว่าค่าความคลาดเคลื่อนมีสหสัมพันธ์ในตัวเองหรือไม่ โดยใช้สถิติทดสอบ Box - Pierce and Ljung - Box Tests

3.6.3 ตัวแบบพยากรณ์วิธีการเรียนรู้ของเครื่อง วิธีแรนดอมฟอเรส (Random Forest)

Geron (2017) กล่าวว่า ตัวแบบแรนดอมฟอเรส หรือ Random Forest เป็นตัวแบบการทำนายที่ใช้พื้นฐานจากต้นไม้ตัดสินใจ (Decision Tree) เป็นการทำนายแบบชุดของต้นไม้ตัดสินใจหลายๆ ต้น (Ensemble of Decision Trees) โดยสร้างจากการเลือกหน่วยตัวอย่างแบบใส่คืน (random sampling with replacement) เพื่อนำมาสร้างเป็นตัวแบบต้นไม้โดยแต่ละต้นมีลักษณะที่ไม่ซ้ำกัน โดยแต่ละแบบจำลองจะมีการทำนายผล ซึ่งผลจากการทำนายของต้นไม้แต่ละต้น จะทำการหาค่าเฉลี่ย ในกรณีRegression สถาปัตยกรรมของแรนดอมฟอเรส



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้ ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกที่รูปที่ 3.7 สถาปัตยกรรมของแรนดอมฟอเรสสำหรับอนุกรมเวลารั้งที่มีการนำไปใช้

จากรูปที่ 3.7 สถาปัตยกรรมของแรนดอมฟอเรสสำหรับอนุกรมเวลา โดยจะกำหนดจำนวนต้นไม้ตัดสินใจในการหาค่าพยากรณ์แต่ละต้นไม้ตัดสินใจ จำนวน k ต้น และจำนวนขั้นตอนเวลาก่อนหน้าให้กับตัวแบบพยากรณ์

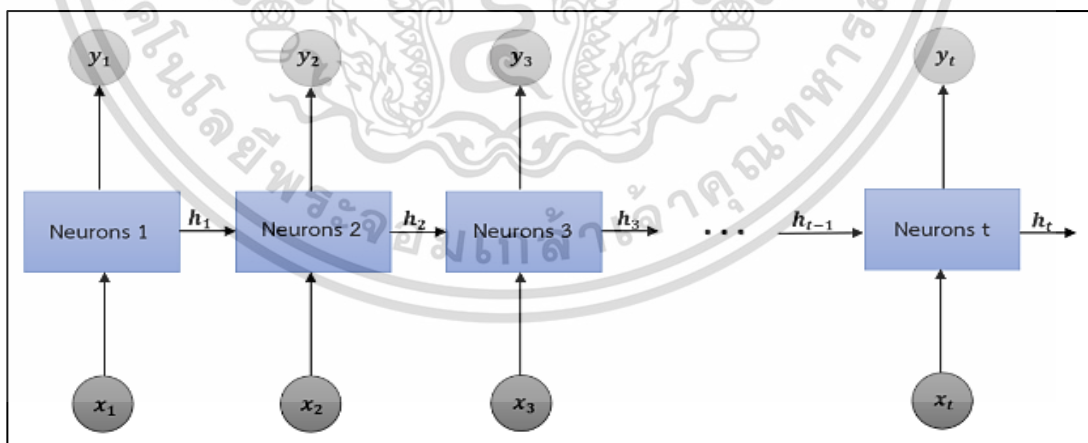
วิธีการสร้างตัวแบบแรนดอมฟอเรสโดยกำหนดค่าไฮเปอร์พารามิเตอร์หรือกระบวนการในการสร้างตัวแบบดังนี้

- 1) จำนวนขั้นตอนเวลา (Time Step) ก่อนหน้า
- 2) จำนวนต้นไม้ตัดสินใจ

ซึ่งผู้วิจัยได้ทำการศึกษากำหนดค่าไฮเปอร์พารามิเตอร์ของจำนวนขั้นตอนเวลา (Time Step) ก่อนหน้า อยู่ในช่วงตั้งแต่ 1 ถึง 10 ขั้นตอนเวลา และจำนวนต้นไม้ตัดสินใจ อยู่ในช่วงตั้งแต่ 200 ถึง 3000 ต้น โดยใช้กลยุทธ์การกำหนดค่าไฮเปอร์พารามิเตอร์แบบวิธีการค้นหาแบบสุ่ม (Random Search) เพื่อค้นหาขอบเขตแบบจำกัดที่เหมาะสม

3.6.4 ตัวแบบพยากรณ์วิธีการเรียนรู้เชิงลึก วิธีหน่วยความจำระยะสั้นระยะยาว

Hochreiter et al. (2004) กล่าวว่า ตัวแบบหน่วยความจำระยะสั้นระยะยาว (LSTM) เป็นตัวแบบที่พัฒนาขึ้นมาเพื่อแก้ปัญหา vanishing gradient ใน RNN โดยจะใช้ข้อมูล Cell State และ Hidden State ในการเก็บข้อมูลและส่งไปประมวลผลยังช่วงเวลาถัดไป ซึ่งอาศัย gate ต่าง ๆ ในการคำนวณค่าว่าจะเก็บรักษาข้อมูลภายใน Cell State และ Hidden State มากน้อยเพียงใด โดยประกอบด้วย Input Gate ,Output Gate และ Forget Gate แต่ละ Gate จะมีหน้าที่ในการตัดสินใจว่าจะให้ข้อมูลผ่านไปหรือไม่ตามค่าความสำคัญข้อมูล สถาปัตยกรรมของหน่วยความจำระยะสั้นระยะยาวที่ออกแบบสำหรับพัฒนาตัวแบบพยากรณ์แสดงดังรูปที่ 3.8



รูปที่ 3.8 สถาปัตยกรรมของหน่วยความจำระยะสั้นระยะยาวสำหรับอนุกรมเวลา

วิธีการสร้างตัวแบบการเรียนรู้เชิงลึกหน่วยความจำระยะสั้นระยะยาว โดยกำหนดค่าไฮเปอร์พารามิเตอร์หรือกระบวนการสร้างตัวแบบดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับ 1) หรืออัลกอริทึม เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้ง 2) มี Epoch ปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 3) Batch size
- 4) Neurons
- 5) Loss Function
- 6) Optimizer

ซึ่งผู้วิจัยได้ทำการศึกษากำหนดค่าไฮเปอร์พารามิเตอร์ของจำนวน Epoch อยู่ในช่วง ตั้งแต่ 1 ถึง 100 และกำหนดจำนวน Neurons ตั้งแต่ 100 ถึง 2000 โดยใช้กลยุทธ์การกำหนดค่าไฮเปอร์พารามิเตอร์แบบวิธีการค้นหาแบบสุ่ม (Random Search) เพื่อค้นหาขอบเขตแบบจำกัดที่เหมาะสม

3.7 การพยากรณ์รายวันล่วงหน้าเป็นระยะเวลาหนึ่งเดือน หรือ 30 วัน

การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ โดยพิจารณาจากค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (Mean Absolute Percentage Error : MAPE) โดยตัวแบบที่ให้ค่า MAPE ต่ำที่สุด จะเป็นตัวแบบที่ดีที่สุด และนำตัวแบบที่ดีที่สุดไปพยากรณ์ 30 วันข้างหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและการอภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการที่เหมาะสมสำหรับการพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ เพื่อเป็นแนวทางหรือข้อมูลเพิ่มเติมในการตัดสินใจในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ โดยจะทำการเลือกวิธีที่เหมาะสมจาก 4 ตัวแบบ ดังนี้

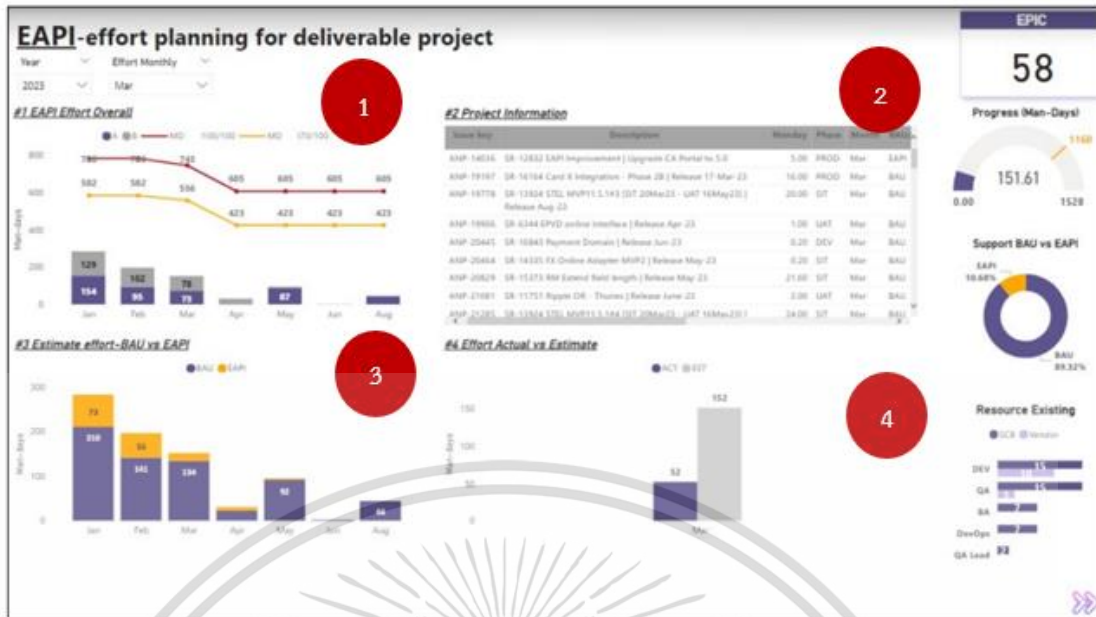
- 1) ตัวแบบ SARIMA
- 2) ตัวแบบ SARIMAX
- 3) ตัวแบบ Random Forest
- 4) ตัวแบบ LSTM

ทั้งนี้ข้อมูลจำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ที่ใช้ในงานวิจัยนี้ถูกแบ่งเป็น 2 ชุด ได้แก่ 1. ชุดข้อมูลฝึกสอนตั้งแต่วันที่ 1/01/2021 – 31/12/2022 (จำนวน 707 จุดเวลา) สำหรับสร้างตัวแบบและ 2. ชุดข้อมูลทดสอบตั้งแต่วันที่ 1/01/2023 - 31/03/2023 (จำนวน 90 จุดเวลา) สำหรับประเมินความถูกต้องของตัวแบบและวัดความแม่นยำของแต่ละตัวแบบด้วยค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (MAPE) โดยตัวแบบที่ดีที่สุดคือตัวแบบที่มีค่า MAPE น้อยที่สุด

4.1 การวิเคราะห์ข้อมูลและแสดงข้อมูลในรูปของแผนภาพหรือกราฟ (Visualization)

1) Effort Planning for Deliverable Project ผู้วิจัยวิเคราะห์ข้อมูลรวมทั้งออกแบบจัดทำรีพอร์ตเกี่ยวกับการวางแผนปริมาณ Man-Day ของ Project ทั้งหมดที่ทีมได้รับมาในแต่ละเดือน (ม.ค. - ธ.ค. 2023) เพื่อให้สามารถเห็นปริมาณที่เข้ามา เมื่อเทียบกับปริมาณ Resource (Man-Days) ที่มีอยู่ของทีม สามารถรองรับได้เพียงพอหรือต้องวางแผนในการจัดหา Resource เพิ่มเติม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.1 ภาพรวมปริมาณงานที่เข้ามาทั้งปีและปริมาณทรัพยากรที่มีอยู่

กราฟที่ 1 แสดงให้เห็นถึงภาพรวมปริมาณงานที่เข้ามาทั้งปี และปริมาณทรัพยากรที่มีอยู่ โดยชุดข้อมูล Resource จะรวมจำนวน Man-Day ของพนักงานและ Outsorce โดยแบ่งเส้นข้อมูลเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 Resource Man-Day 100% (เส้นสีแดง) ส่วนที่ 2 Resource Man-Day 70% (เส้นสีเหลือง) เพื่อใช้เป็นเส้นประเมินระหว่างปริมาณงานและ ปริมาณ Resource ที่สามารถรับได้ ซึ่งหากปริมาณงานในแต่ละเดือนยังไม่ถึงเส้น 70% (สีเหลือง) ก็ยังสามารถรับงานในเดือนนั้นได้อีก แต่หากปริมาณงานถึงเส้น 70% หรือเกิน ก็ตั้งสมมติฐานได้ว่าปริมาณ Resource ในเดือนนั้นอาจจะไม่เพียงพอ หรือต้องวางแผนการรับงานไปในเดือนถัดไป หรือสามารถวางแผนการเพิ่ม Resource เพื่อมารองรับงานในเดือนต่อไปได้ในอนาคต

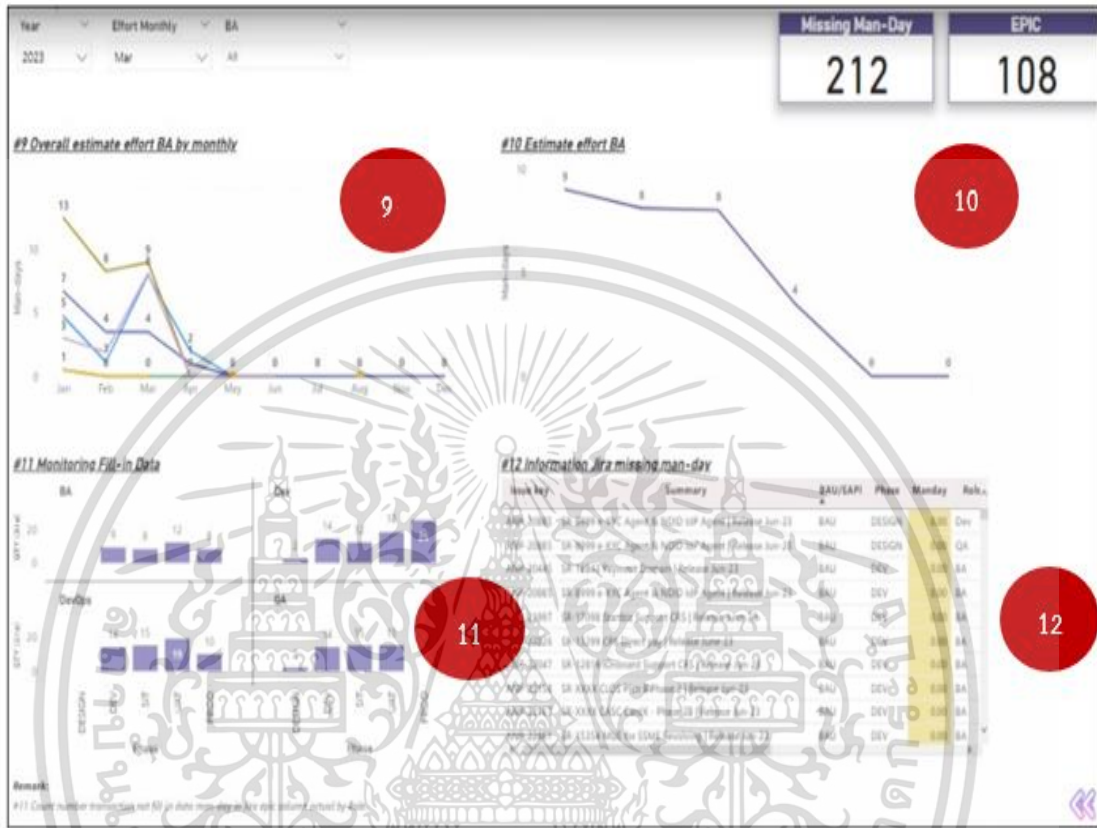
กราฟที่ 2 ตารางแสดงรายละเอียดโดยสังเขปของโปรเจกต์ทั้งหมดที่อยู่ในแผนงาน (Project Planning) ว่ามีรายละเอียดผู้รับผิดชอบ วัน และเวลาในการดำเนินการ เพื่อให้สามารถดูข้อมูลเบื้องต้นได้

กราฟที่ 3 แสดงปริมาณงานทั้งหมดที่ทีมรับมา โดยแบ่งออกเป็นสัดส่วนของงาน BAU (Business As Usual) และงาน EAPI (งาน Internal Improvement ของทีม) โดยส่วนมากในแต่ละเดือนจะเป็นงานที่ทำ Support Business เป็นหลัก และงานปรับปรุงแก้ปัญหาภายในทีมจะมีค่อนข้างน้อย เพื่อให้ทีมหรือผู้บริหารสามารถเห็นปริมาณงานในแต่ละเดือนว่ากลุ่มงานที่ทำฝ่ายสนับสนุนและสัดส่วน Resource Man-Day ที่ใช้ในการสนับสนุน งานแต่ละกลุ่ม

กราฟที่ 4 แสดงสัดส่วนของ Resource Man-day ที่ประมาณการล่วงหน้า (Estimate) และที่ใช้จริง (Actual) ในแต่ละเดือน เพื่อให้เห็นปริมาณ Resource ที่ประมาณการณ์ล่วงหน้า (Estimate) กับปริมาณ Resource ที่ใช้จริงแต่ละเดือน ยังเพียงพอต่อปริมาณงานที่เข้ามาในแต่ละเดือน อีกทั้งยังสามารถนำไปวิเคราะห์เพิ่มเติมในแต่ละเดือน ของค่าความคลาดเคลื่อน ของจำนวน Man-Day

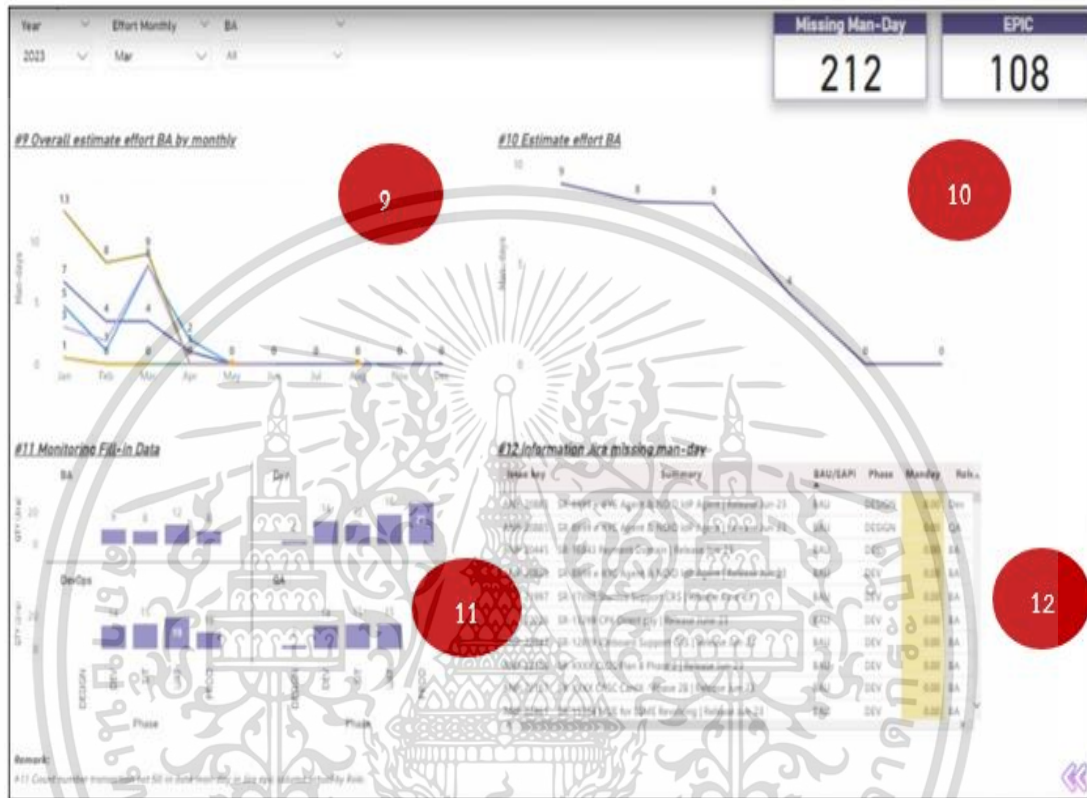
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามเผยแพร่ลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งหากนำไปใช้

ระหว่าง Man-Day คาดการณ์ล่วงหน้า (Estimate) และ Man -Day ที่ใช้จริง (Actual) ทำให้ทีมนำไปปรับปรุงเรื่องของการประมาณการล่วงหน้าในเดือนถัดไป เพื่อให้ Man-Day ของการประมาณการล่วงหน้า (Estimate) กับ ค่าที่ใช้จริง (Actual) มีความแม่นยำมากขึ้น



ต้องใช้ Support โครงการ มีแนวโน้มมากขึ้นหรือลดลง เพื่อให้ทีมนำไปจัดสรร Resource Man-Day ในการ Support งานให้ดีขึ้น

กราฟที่ 8 แสดงภาพรวมปริมาณของ Resource Man-day ที่ประมาณการล่วงหน้า (Estimate) ในแต่ละ Phase เพื่อให้เห็นปริมาณ Phase ที่ใช้ Man-Day ในการสนับสนุนงานมากที่สุด



รูปที่ 4.3 ภาพรวมปริมาณ Man-day ที่ประมาณการล่วงหน้า

กราฟที่ 9 แสดงภาพรวมปริมาณ Man-Day ที่ประมาณการล่วงหน้า (Estimate) ของ ตำแหน่ง BA (Business Analyst) เพื่อสามารถเห็นปริมาณของแต่ละคนในแต่ละเดือน ทำให้รายงานนี้ช่วยในการแบ่งและจัดสรรงาน หากพบว่ามีคนรับงานมากเกินไป เพื่อจัดการการแบ่งงานให้สอดคล้องกับจำนวนคน หากไม่เพียงพอ สามารถใช้รายงานนี้ในการอ้างอิง เพื่อนำเสนอผู้บริหารขอรับคนเพิ่มในอนาคตต่อไป

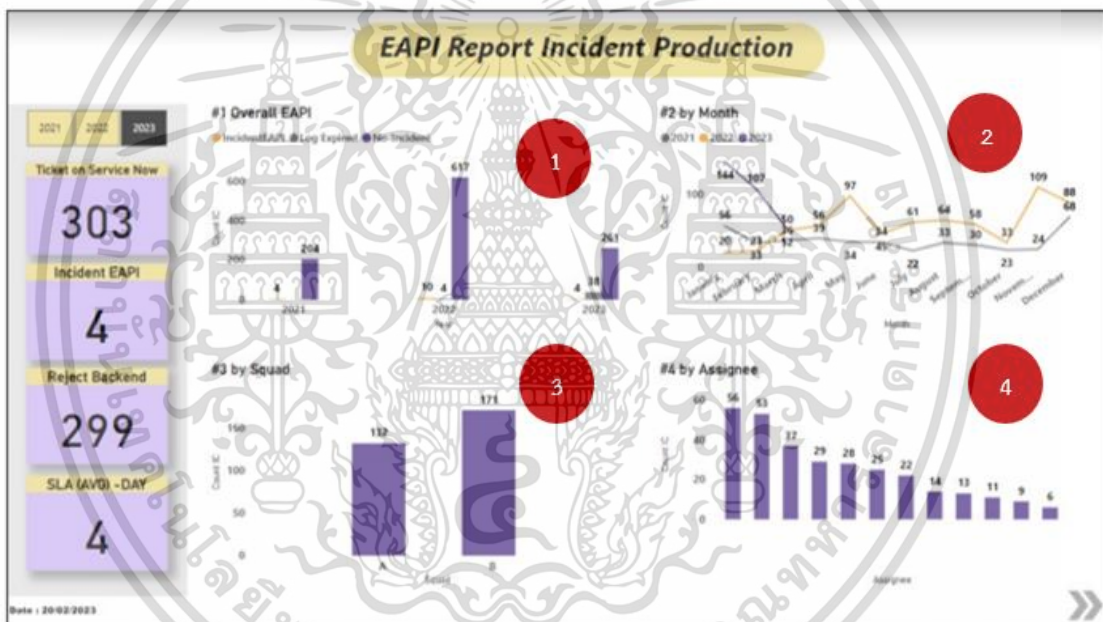
กราฟที่ 10 แสดงภาพรวมปริมาณ Man-Day ที่ประมาณการล่วงหน้า (Estimate) ของแต่ละคนในตำแหน่ง BA ในตัวอย่างรายงานนี้เป็นของเดือนมีนาคม รายงานนี้จะนำไปประเมินจำนวน Man-day ที่ BA support ในเดือนนั้น ๆ และสามารถให้ผู้บริหารหน่วยงานรับทราบถึงปริมาณงาน กับปริมาณ Man-Day ที่ BA ยังรับงานได้เพียงพอในแต่ละเดือน

กราฟที่ 11 แสดงปริมาณงานในแต่ละตำแหน่ง ยังไม่ใส่ ค่า Man-Day ของการประมาณการล่วงหน้า ของแต่ละ Phase (Cycle ของการ Implement) กราฟนี้จะช่วยให้ทีมสามารถติดตามไปว่า ภาระงานในแต่ละ Phase นั้นๆ มีแนวโน้มที่จะเพิ่มขึ้นหรือลดลงหรือไม่ หากพบว่ามีแนวโน้มที่จะเพิ่มขึ้นหรือลดลง ทีมสามารถเตรียมแผนรองรับได้ล่วงหน้า ไม่อย่างนั้น ทีมอาจต้องเผชิญกับปัญหาการขาดแคลนทรัพยากรในบาง Phase ได้

เจ็ทท์ที่รับผิดชอบว่ามีงานไหนบ้าง ยังไม่ใส่ Man-day เพื่อให้ทีมสะดวกในการมอนิเตอร์การใส่ค่าของ Man-day ให้ครบถ้วนและสมบูรณ์ เพื่อให้ข้อมูลมีความครบถ้วนมากที่สุดในแต่ละเดือน

กราฟที่ 12 ตารางแสดงรายละเอียดของโปรเจกต์ ที่แสดงตามการเลือกของผู้ใช้งาน โดยจะแสดงเฉพาะรายการของโปรเจกต์ที่ยังไม่ใส่ค่า Man-Day ประมาณการล่วงหน้า เพื่อให้ทีมสามารถส่งเป็นไฟล์ (นามสกุล csv) เพื่อส่งต่อการนำไปติดตามการใส่ข้อมูลให้สมบูรณ์ต่อไป

2) Incident Production ผู้วิจัยวิเคราะห์ข้อมูลออกมาและจัดทำรีพอร์ตเกี่ยวกับ ปริมาณของ Incident (ปัญหาการใช้งานของแอปพลิเคชัน) ที่แจ้งเข้ามาในแต่ละเดือน และสัดส่วนของ Incident ที่แก้ไขโดยทีม และจำนวน SLA ที่ใช้ระยะเวลาในการแก้ไขปัญหา รวมทั้งยังออกแบบเพื่อให้เห็นภาพรวมการรับงานincident ของผู้รับผิดชอบแต่ละคนในแต่ละเดือน เพื่อให้สามารถติดตามงาน และทราบถึง Man-day ที่ใช้ในการ support incident ทั้งหมด เพื่อนำเสนอต่อผู้บริหาร



รูปที่ 4.4 ภาพรวมปริมาณการแจ้งปัญหาการใช้งานของแอปพลิเคชัน

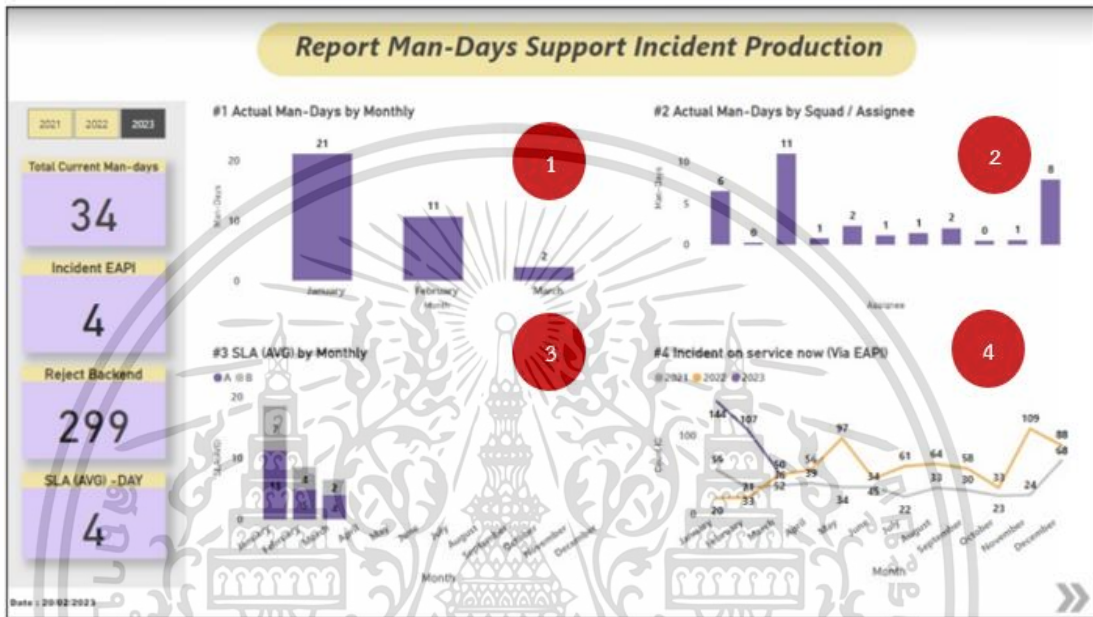
กราฟที่ 1 แสดงภาพรวมปริมาณการแจ้ง Incident (ปัญหาการใช้งานของแอปพลิเคชัน) ในช่วง 3 ปี (ข้อมูล 2021-2023) เพื่อให้เห็นภาพรวมตั้งแต่ปี 2021-2023 มีปริมาณการแจ้งปัญหาของแอปพลิเคชันเข้ามาเพิ่มขึ้น หรือ ลดลง โดยแบ่งสัดส่วนงานที่แจ้งเข้ามาเป็นงานที่ไม่ใช่ปัญหาของทีม (แห่งสีม่วง) และงานที่เป็นปัญหาของทีมที่จะต้องดำเนินการแก้ไขให้กับแอปพลิเคชัน (แห่งสีส้ม)

กราฟที่ 2 แสดงการเปรียบเทียบปริมาณงานตั้งแต่ปี 2021-2023 โดยแสดงเป็นรายเดือน เพื่อให้เห็น Trend (แนวโน้ม) ของข้อมูลในแต่ละปี เดือนไหนมีปริมาณของการแจ้ง Incident เข้ามามากหรือน้อย เพื่อนำไปวางแผนการจัด Resource Man-Day ในการ Support ช่วงเดือนที่มีปริมาณการแจ้งที่สูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กราฟที่ 3 กราฟแสดงปริมาณงานการแจ้ง Incident โดยแบ่งออกเป็น 2 กลุ่มได้แก่ Squad A (A) และ Squad B (B) เพื่อจะได้วางแผนการแบ่งผู้รับผิดชอบในการ Support Incident ให้เท่ากัน

กราฟที่ 4 แสดงปริมาณงาน โดยแบ่งตามผู้รับผิดชอบ เพื่อให้ผู้บริหารสามารถเห็นปริมาณงานของผู้รับผิดชอบแต่ละคน และนำไปวางแผนในการปรับลด-เพิ่ม จำนวนการ Support Incident ของเดือนถัดไปในอนาคต



รูปที่ 4.5 ปริมาณ Man-Day ที่ใช้ Support Incident จริง (Actual) ในแต่ละเดือน

กราฟที่ 1 แสดงปริมาณ Man-Day ที่ใช้ Support Incident จริง (Actual) ในแต่ละเดือนเพื่อให้ผู้บริหารหน่วยงานเห็นภาพรวมของการสนับสนุนในแต่ละเดือน

กราฟที่ 2 แสดงปริมาณ Man-Day ที่ใช้ Support Incident จริง (Actual) ที่แบ่งตามกลุ่มได้แก่ Squad A (A) และ Squad B (B) เพื่อให้เห็นภาพรวมของทั้ง 2 กลุ่ม ซึ่งผู้บริหารสามารถนำไปวางแผนการ Support ของแต่ละกลุ่มในอนาคต และกราฟยังสามารถดูรายละเอียด ปริมาณ Man-Day (Actual) ตามชื่อผู้รับผิดชอบ เพื่อให้สามารถดูข้อมูลในเชิงลึกได้ เพื่อเป็นข้อมูลประกอบการตัดสินใจในการร่วมกับการวางแผน

กราฟที่ 3 แสดงค่าเฉลี่ยของจำนวนวันในการปิดงาน Incident ในระบบ หรือ SLA (Service-level agreement) ว่าในแต่ละเดือนมีค่าเฉลี่ยปิดงานเมื่อเทียบกับเดือนที่ผ่านมาเพิ่มขึ้น หรือลดลง เพื่อนำมาปรับปรุง SLA ในการปิดงาน incident ปรับใช้กับเดือนถัดไปให้ดีขึ้น

กราฟที่ 4 อ้างอิงมาจากรีพอร์ตหน้าที่ 1 เพราะจะทำให้ง่ายและสะดวกในการดูภาพรวมทั้งหมดของปริมาณงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.6 ภาพรวมของปริมาณงานทั้งหมดที่ปิดงานในระบบ Service Now

กราฟที่ 1 แสดงภาพรวมของปริมาณงานทั้งหมดที่ปิดงานในระบบ Service Now (ระบบแจ้งปัญหา Incident) โดยแสดงเป็นสัดส่วน ได้แก่ งานที่ปิดใบบงานผ่านระบบเรียบร้อยแล้ว (Close_IC_Done) งานที่ยังไม่ปิดงานหรือรอปิดงาน (To_Be_Closed_IC) และงานที่รอดำเนินการแก้ไขหรือหาสาเหตุของปัญหาที่แจ้งเข้ามา (Pending) ทำให้ผู้บริหารทีมเห็นภาพรวมง่ายขึ้น

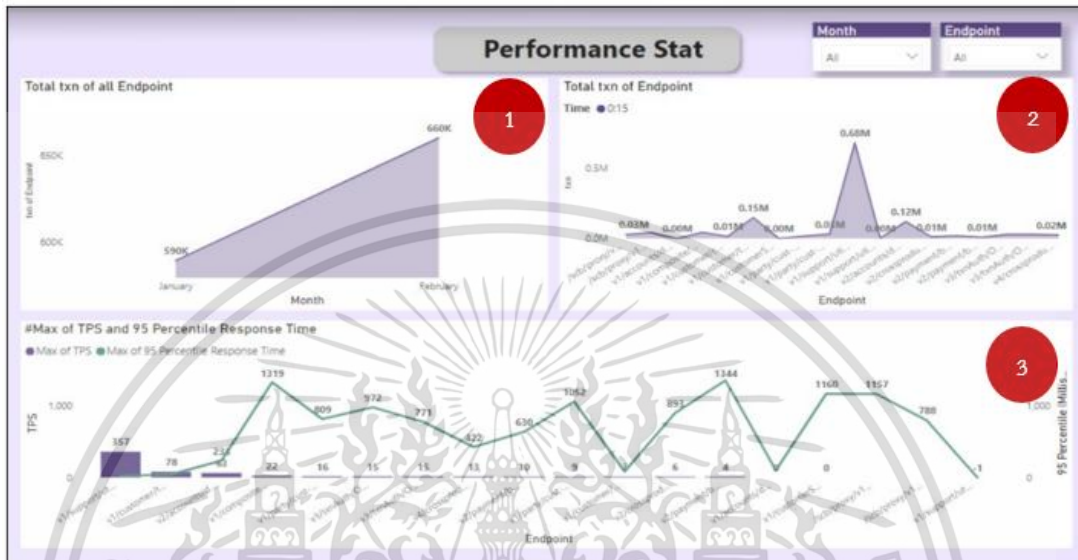
กราฟที่ 2 แสดงปริมาณงาน incident ที่แจ้งเข้ามา โดยแบ่งปริมาณงานออกเป็น 2 กลุ่ม ได้แก่ Squad A (A) และ Squad B (B) และแสดงสัดส่วนของข้อมูลออกตามสถานะของการปิดงาน Close_IC_Done, To_Be_Closed_IC และ Pending เพื่อให้ผู้บริหารทีมทราบว่าแต่ละ Squad มีปริมาณงานค้างปิด หรือยังรอแก้ไข มีมากหรือน้อย และสามารถนำไปประกอบเป็นข้อมูล KPI (Key Performance Indicator) ของทีมได้ในอนาคตต่อไป

กราฟที่ 3 แสดงปริมาณงานผู้รับผิดชอบแต่ละคน โดยแสดงสัดส่วนตามสถานการณ์ปิดงาน incident บนระบบ Service Now ซึ่งมีทั้งหมด 3 สถานะ ได้แก่ Close_IC_Done, To_Be_Closed_IC และ Pending เพื่อใช้เป็นรายงานติดตามงานกับผู้รับผิดชอบแต่ละคน ที่ยังไม่ปิดงาน incident บนระบบ โดยมุ่งเน้นไปที่สถานะ Pending (สีส้ม) ที่จะต้องติดตามเร่งด่วน เนื่องจากอาจจะเกินเวลา SLA ที่กำหนดไว้ (1-7 วัน) รายงานนี้จะช่วยให้ช่วยติดตามงานได้อย่างมีประสิทธิภาพมากขึ้น

กราฟที่ 4 แสดงปริมาณงานเฉพาะเจาะจงสถานะ งานที่รอดำเนินการแก้ไขหรือหาสาเหตุของปัญหาที่แจ้งเข้ามา (Pending) ของผู้รับผิดชอบแต่ละคนที่ยังไม่ดำเนินการ เพื่อติดตามงานและแจ้งผู้รับผิดชอบให้รีบดำเนินการก่อนครบกำหนด SLA ที่กำหนดไว้ (1-7 วัน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) Performance Stat ผู้วิจัยวิเคราะห์ข้อมูลออกมาและจัดทำรีพอร์ตเพื่อเปรียบเทียบข้อมูลเกี่ยวกับการทดสอบประสิทธิภาพ (Performance Test) ของการทำรายการเรียกใช้เซอร์วิสของไมโครเซอร์วิส (Transaction of microservice) โดยคัดเลือกมาจากเซอร์วิสที่มีจำนวนการเรียกที่เยอะที่สุด 1-20 อันดับ เพื่อใช้ในการวิเคราะห์และเปรียบเทียบในการทำวิจัยครั้งนี้



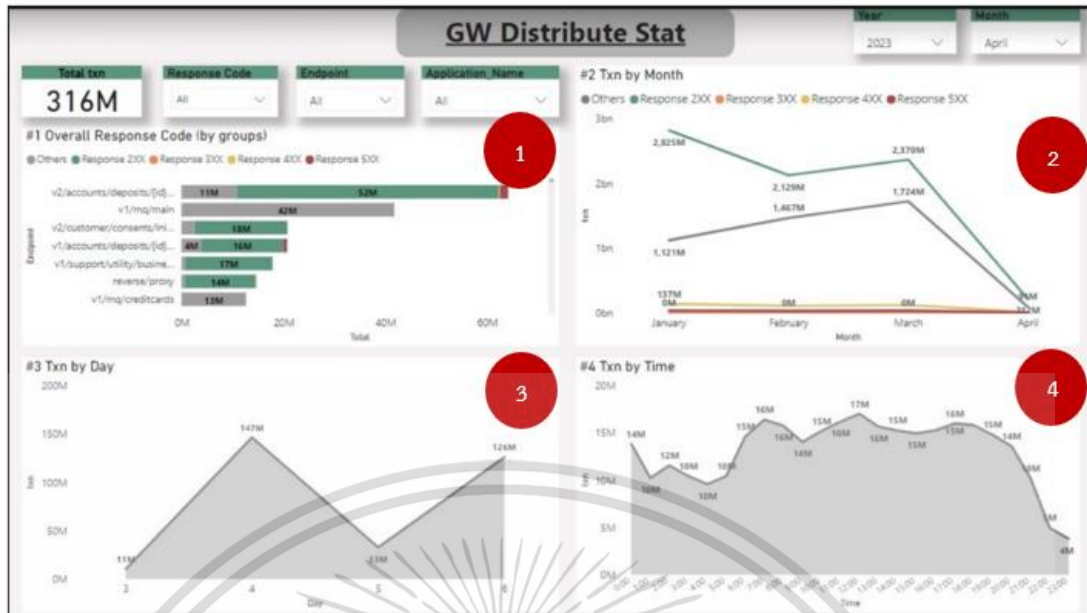
รูปที่ 4.7 ปริมาณการเรียกใช้เซอร์วิสของไมโครเซอร์วิส

กราฟที่ 1 แสดงปริมาณการเรียกใช้เซอร์วิสของไมโครเซอร์วิส (Transaction of Microservice) จากข้อมูลแสดงให้เห็นตั้งแต่เดือนมกราคม - เดือนกุมภาพันธ์ (ปี 2023) จะทำให้เห็นแนวโน้ม (Trend) ของการเรียกใช้ข้อมูลในแต่ละเดือน

กราฟที่ 2 แสดงภาพรวม Transaction of Microservice โดยแสดงตามชุดข้อมูล Endpoint (เป็น url path ของการเรียกใช้ไมโครเซอร์วิส) แสดง 1 - 20 ที่คัดเลือกมาตามที่กล่าวในเบื้องต้น โดยแสดงจำนวนทั้งหมดของ Transaction of Microservice เฉลี่ย 15 วินาที เพื่อเปรียบเทียบให้เห็นภาพรวมทั้งหมดว่ามีเซอร์วิสที่ทุก 15 วินาที ใช้มากที่สุด เพื่อนำ Endpoint ของเซอร์วิสนั้นไปใช้ในการวิเคราะห์ต่อไป

กราฟที่ 3 แสดงปริมาณมากที่สุด (Max) ของการทำรายการต่อวินาที (TPS หรือ Transaction per Second) ของ Transaction of Microservice โดยแสดงตามชุดข้อมูล Endpoint ทั้ง 1-20 โดยใช้กราฟเส้น (สีเขียว) เพื่อให้เห็นแนวโน้มของข้อมูลแต่ละ Endpoint ที่มีระยะเวลาการตอบกลับ (Response Time) ค่าสูงสุด โดยคิดเป็น 80% (80 Percentile) ทั้งนี้ผู้วิจัยได้วิเคราะห์จากกราฟนี้ โดยดูจากปริมาณการเรียกใช้เซอร์วิสของการทำรายการต่อวินาที (TPS) ที่มีจำนวนมากที่สุด เพื่อใช้ในการคาดการณ์สำหรับงานวิจัยนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.9 ปริมาณการเรียกใช้เซอร์วิสของเกตเวย์

กราฟที่ 1 แสดงปริมาณการเรียกใช้เซอร์วิสของเกตเวย์ จากข้อมูลแสดงให้เห็นรหัสการตอบกลับ (Response Code) คือ Response 2xx, Response 3xx, Response 4xx, และ Response 5xx ซึ่งจะทำให้เห็นถึงการตอบกลับของการเรียกใช้ในแต่ละข้อมูล

กราฟที่ 2 แสดงภาพรวมของเซอร์วิสที่มีการตอบกลับ โดยแบ่งออกตามเดือน เพื่อให้เห็นถึงแนวโน้มการตอบกลับของข้อมูล

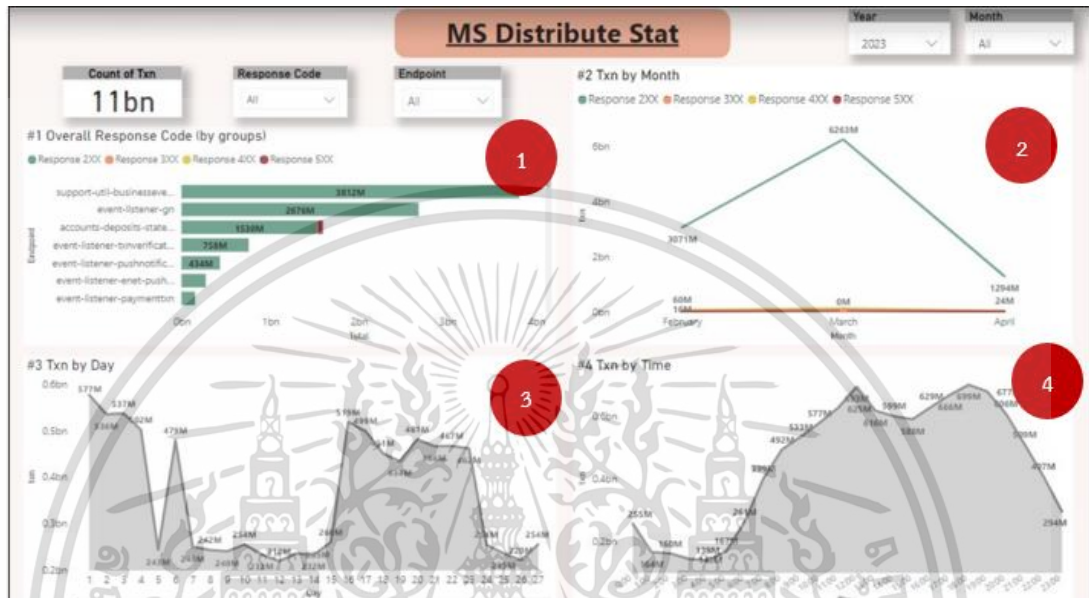
กราฟที่ 3 แสดงภาพรวมของเซอร์วิสที่มีการตอบกลับ โดยแบ่งออกตามวัน เพื่อให้เห็นถึงแนวโน้มการตอบกลับของข้อมูล

กราฟที่ 4 แสดงภาพรวมของเซอร์วิสที่มีการตอบกลับ โดยแบ่งออกตามช่วงเวลา เพื่อให้เห็นถึงแนวโน้มการตอบกลับของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กราฟที่ 3 แสดงปริมาณมากที่สุด (Max) ของการทำรายการต่อวินาที (TPS หรือ Transaction per Second) ในแต่ละช่วงเวลา ตั้งแต่ 00.00-23.00

กราฟที่ 4 แสดงระยะเวลาการตอบกลับ (Response Time) ค่าสูงสุด โดยคิดเป็น 95% (95 Percentile) ในแต่ละช่วงเวลา ตั้งแต่ 00.00-23.00



รูปที่ 4.12 ปริมาณการเรียกใช้เซอร์วิสของไมโครเซอร์วิส

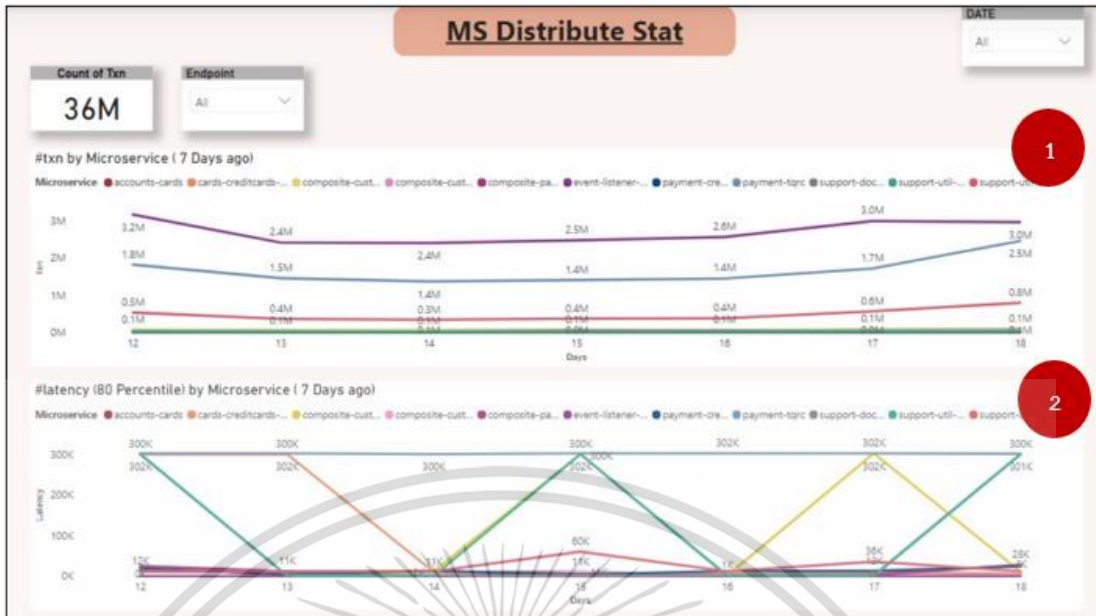
กราฟที่ 1 แสดงปริมาณการเรียกใช้เซอร์วิสของไมโครเซอร์วิส จากข้อมูลแสดงให้เห็นรหัสการตอบกลับ (Response Code) คือ Response 2xx, Response 3xx, Response 4xx, และ Response 5xx ซึ่งจะทำให้เห็นถึงการตอบกลับของการเรียกใช้ในแต่ละข้อมูล

กราฟที่ 2 แสดงภาพรวมของเซอร์วิสที่มีการตอบกลับ โดยแบ่งออกตามเดือน เพื่อให้เห็นถึงแนวโน้มการตอบกลับของข้อมูล

กราฟที่ 3 แสดงภาพรวมของเซอร์วิสที่มีการตอบกลับ โดยแบ่งออกตามวัน เพื่อให้เห็นถึงแนวโน้มการตอบกลับของข้อมูล

กราฟที่ 4 แสดงภาพรวมของเซอร์วิสที่มีการตอบกลับ โดยแบ่งออกตามช่วงเวลา เพื่อให้เห็นถึงแนวโน้มการตอบกลับของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.13 แสดงภาพรวม 7 วันย้อนหลัง โดยแสดงตามชุดข้อมูลไมโครเซอร์วิส

กราฟที่ 1 แสดงภาพรวม 7 วันย้อนหลัง Transaction of Microservice โดยแสดงตามชุดข้อมูลไมโครเซอร์วิส

กราฟที่ 2 แสดงภาพรวม 7 วันย้อนหลัง ระยะเวลาการตอบกลับ (Response Time) ค่าสูงสุด โดยคิดเป็น 80% (85 Percentile) โดยแสดงตามชุดข้อมูลไมโครเซอร์วิส

txn by Microservice		Latency by Microservice	
Microservice	Sum of Count	Microservice	Max of 80th percentile of Avg
event-listener-paymenttxn	57,893,580	composite-customer	305,384.00
payment-tqrc	37,847,561	payment-tqrc	303,119.00
support-uti-proxy	11,360,611	support-uti-bioauth	300,115.00
composite-customer	1,554,766	composite-payment	300,109.00
accounts-cards	1,551,378	support-docmngt-vchannel	300,106.00
support-uti-bioauth	602,007	cards-creditcards-statements	300,102.60
composite-payment	555,554	support-uti-proxy	184,900.60
support-docmngt-vchannel	364,510	composite-customer-svcs	30,029.80
payment-creditcardQR-web	227,348	payment-creditcardQR-web	24,380.00
composite-customer-svcs	188,537	accounts-cards	12,065.00
cards-creditcards-statements	154,578	event-listener-paymenttxn	6.00

รูปที่ 4.14 แสดงค่าเฉลี่ยระยะเวลาที่ดีที่สุด 10 อันดับแรกของไมโครเซอร์วิส

กราฟที่ 1 ตารางแสดงภาพรวม จำนวนรายการที่ต้องดำเนินการที่ดีที่สุด 10 อันดับแรกของไมโครเซอร์วิส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 กราฟที่ 2 ตารางแสดงภาพรวม ค่าเฉลี่ยระยะเวลาที่ดีที่สุด 10 อันดับแรกของไมโครเซอร์วิส
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 การพยากรณ์วิธีการบ็อกซ์ - เจนกินส์

เนื่องจากพบว่าข้อมูลมีความแปรผันทางฤดูกาลอย่างชัดเจน ผู้วิจัยจึงเลือกใช้วิธีบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMA สำหรับการพยากรณ์ซึ่งเป็นตัวแบบที่เหมาะสมกับข้อมูลที่มีลักษณะดังกล่าว โดยผลที่ได้จากการทดสอบเป็นดังนี้

4.2.1 ผลการทดลองวิธีบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMA

การพยากรณ์วิธีการบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMA ทำการทดสอบตามขั้นตอน ดังนี้

1) การตรวจสอบข้อมูล

จากข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส ที่แสดงในรูปที่ 3.2 จะพบว่าข้อมูลมีความแปรผันทางฤดูกาลทุกๆ 30 วัน และจากรูปที่ 3.3 และ 3.4 ค่า ACF และ PACF ตามลำดับ พบว่าอนุกรมเวลาไม่คงที่ กล่าวคือมีความแปรผันทางฤดูกาลเข้ามาเกี่ยวข้อง ดังนั้นจึงทำการหาผลต่างอันดับที่ 1 ของความแปรผันทางฤดูกาลของข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เพื่อให้อนุกรมเวลาดังกล่าวคงที่

2) การกำหนดตัวแบบที่เหมาะสม

จากการตรวจสอบข้อมูล พบว่าข้อมูลมีความแปรผันทางฤดูกาลทุกๆ 30 วัน จึงกำหนด ค่า $s = 30$ ในการพิจารณาตัวแบบที่เหมาะสม จะทำการพิจารณาตัวแบบที่มีค่า AIC น้อยที่สุด

ตารางที่ 4.1 ค่า AIC ในแต่ละอันดับของตัวแบบ ARIMA (p, d, q) SARIMA (P, D, Q)_s

ARIMA (p, d, q) SARIMA (P, D, Q) _s	AIC
ARIMA(0,0,0)x SARIMA(0,0,0) ₃₀	8754.419
ARIMA(0,0,0)x SARIMA(0,0,1) ₃₀	7681.964
ARIMA(0,0,0)x SARIMA(0,0,2) ₃₀	6893.5455
ARIMA(0,0,0)x SARIMA(0,1,0) ₃₀	5942.0099
ARIMA(0,0,0)x SARIMA(0,1,1) ₃₀	5652.0295
ARIMA(0,0,0)x SARIMA(0,1,2) ₃₀	5408.9028
ARIMA(0,0,0)x SARIMA(0,2,0) ₃₀	6271.3399
...	...
ARIMA(2,2,2)x SARIMA(1,2,0) ₃₀	5384.8721
ARIMA(2,2,2)x SARIMA(1,2,1) ₃₀	5048.8208
ARIMA(2, 2, 2)x SARIMA(1, 2, 2)₃₀	4783.2615
ARIMA(2,2,2)x SARIMA(2,0,0) ₃₀	4978.4256

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(2,2,2)x SARIMA(2,0,1) ₃₀	4982.5406
ARIMA(2,2,2)x SARIMA(2,0,2) ₃₀	4977.8950

เมื่อพิจารณาตารางที่ 4.1 ค่า AIC ที่น้อย พบว่าอันดับที่ดีที่สุด ได้แก่ $p = 2, d = 2, q = 2$ และ $P = 1, D = 2, Q = 2$ และตัวแบบที่เหมาะสมคือ ตัวแบบ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30}$ ซึ่งมีรูปแบบทั่วไปแสดงดังสมการที่ 4.1

$$(1 - \phi_1 B)(1 - B^{30})Y_t = (1 - \Theta_1 B^{30})\varepsilon_t \quad (4.1)$$

โดยกำหนด	Y_t	แทน ค่าพยากรณ์ ณ เวลา t
	ε_t	แทน ค่าความคลาดเคลื่อน ณ เวลาที่ t
	B	แทน ตัวดำเนินการย้อนหลัง
	ϕ_1	แทน การถดถอยในตัวเองแบบไม่มีฤดูกาลอันดับ 1
	Θ_1	แทน ค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ 1

3) การประมาณค่าพารามิเตอร์

หลังจากได้ตัวแบบที่เหมาะสมจากสมการที่ 4.1 แล้วจึงนำตัวแบบดังกล่าวมาประมาณค่าพารามิเตอร์แสดงดังตารางที่ 4.2

ตารางที่ 4.2 ทดสอบค่าพารามิเตอร์ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30}$

Statistics	Coef.	SE Coef.	t	p-value
ϕ_1	0.6531	0.068	9.672	0.000
Θ_1	-1.7126	0.065	-26.408	0.000

สมมติฐานในการทดสอบ

$$H_0 : \phi_1 = 0$$

$$H_1 : \phi_1 \neq 0$$

จากตารางที่ 4.2 พบว่าค่า p - value = 0.000 < $\alpha = 0.05$ จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปคือ ค่าพารามิเตอร์ ϕ_1 ในตัวแบบมีค่าไม่เท่ากับ 0 หรือ พารามิเตอร์ ϕ_1 ควรมีในตัวแบบ

$$H_0 : \Theta_1 = 0$$

$$H_1 : \Theta_1 \neq 0$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.2 พบว่าค่า p - value = 0.000 < α = 0.05 จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปคือ ค่าพารามิเตอร์ θ_1 ในตัวแบบมีค่าไม่เท่ากับ 0 หรือ พารามิเตอร์ θ_1 ควรมีในตัวแบบ ดังนั้นตัวแบบ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30}$ เป็นตัวแบบที่เหมาะสม

4) การตรวจสอบความเหมาะสมของตัวแบบ

จากนั้นทำการตรวจสอบค่าความคลาดเคลื่อนว่าจะต้องไม่มีสหสัมพันธ์ในตัวเองด้วย Box-Pierce and Ljung-Box Tests ดังตารางที่ 4.3

ตารางที่ 4.3 ทดสอบความเป็นอิสระของค่าความคลาดเคลื่อนของตัวแบบ ด้วย Box-Pierce and Ljung-Box Tests

Modified Box-Pierce (Ljung-Box) Chi-Square Statistics	
Q	2.12
p-value	0.15

สมมติฐานในการทดสอบ

$$H_0 : \rho_e(1) = \rho_e(2) = \rho_e(3) = \dots = \rho_e(k) = 0$$

(ค่าความคลาดเคลื่อนของตัวแบบไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลา)

$$H_1 : \rho_e(k) \text{ อย่างน้อยหนึ่งค่าไม่เท่ากับ } 0 \text{ สำหรับ } k = 1, 2, 3, \dots, m$$

(ค่าความคลาดเคลื่อนของตัวแบบมีสหสัมพันธ์ในตัวเองทุกช่วงเวลา)

จากตารางที่ 4.3 พบว่าค่า p - value = 0.15 > α = 0.05 จึงยอมรับ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปว่า ค่าความคลาดเคลื่อนของตัวแบบที่พิจารณาไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลา ดังนั้นแสดงว่าตัวแบบ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30}$ เป็นตัวแบบที่เหมาะสม จากการทดสอบจะได้พารามิเตอร์ที่เหมาะสมที่สุดดังตารางที่ 4.4

ตารางที่ 4.4 ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดของตัวแบบ SARIMA

ตัวแบบ	พารามิเตอร์					
	p	d	q	P	D	Q
ARIMA (p, d, q) SARIMA (P, D, Q)s	2	2	2	1	2	2

ดังนั้น สมการพยากรณ์ด้วยตัวแบบ SARIMA คือ

$$Y_t = Y_{t-30} - \phi_1 Y_{t-1} + \phi_1 Y_{t-31} + \varepsilon_t - \theta_1 \varepsilon_{t-30}$$

4.2.2 ผลการทดลองวิธีบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMAX

การพยากรณ์วิธีการบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMAX โดยทำการทดสอบตามขั้นตอนดังนี้
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) การตรวจสอบข้อมูล

จากจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส ตั้งแต่วันที่ 1/01/2021 – 31/12/2022 พบว่าจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสน้อยในวันเสาร์ - อาทิตย์ ทางผู้วิจัยได้เพิ่มตัวแปรอิสระหรือตัวแปรภายนอก(Exogenous) ขึ้นมาโดยที่

X_1 แทน วันเสาร์ - อาทิตย์

2) การกำหนดตัวแบบที่เหมาะสม

พิจารณาตัวแบบจากวิธี SARIMA จากสมการ 4.1 จะได้ตัวแบบที่เหมาะสม คือ ตัวแบบ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30}$ ซึ่งมีรูปแบบทั่วไปแสดงดังสมการที่ 4.2

$$(1 - \phi_1 B)(1 - B)(1 - B^{30})Y_t = (1 - \theta_1)(1 - \Theta_1 B^{30})\varepsilon_t + \alpha_1 X_1 \quad (4.2)$$

โดยกำหนด Y_t แทน ค่าพยากรณ์ ณ เวลา t
 ε_t แทน ค่าความคลาดเคลื่อน ณ เวลาที่ t
 B แทน ตัวดำเนินการย้อนหลัง
 ϕ_1 แทน การถดถอยในตัวเองแบบไม่มีฤดูกาลอันดับ 1
 θ_1 แทน ค่าเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับ 1
 Θ_1 แทน ค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ 1
 α_1 แทน พารามิเตอร์ของ X_1 (วันเสาร์ - อาทิตย์)

3) การประมาณค่าพารามิเตอร์

หลังจากได้ตัวแบบที่เหมาะสมแล้ว จึงนำตัวแบบดังกล่าวมาประมาณค่าพารามิเตอร์ แสดงดังตารางที่ 4.5

ตารางที่ 4.5 ทดสอบค่าพารามิเตอร์ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30} + \alpha_1 + X_1$

Statistics	Coef.	SE Coef.	t	p-value
ϕ_1	0.3114	0.072	4.348	0.000
θ_1	-0.0143	0.010	-1.500	0.003
Θ_1	-2.2136	1.970	-1.123	0.002
α_1	-8.9561	0.878	-10.195	0.000

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมมติฐานในการทดสอบ

$$H_0 : \phi_1 = 0$$

$$H_1 : \phi_1 \neq 0$$

จากตารางที่ 4.5 พบว่าค่า p - value = 0.000 < $\alpha = 0.05$ จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปคือ ค่าพารามิเตอร์ ϕ_1 ในตัวแบบมีค่าไม่เท่ากับ 0 หรือ พารามิเตอร์ ϕ_1 ควรมีในตัวแบบ

$$H_0 : \theta_1 = 0$$

$$H_1 : \theta_1 \neq 0$$

จากตารางที่ 4.5 พบว่าค่า p - value = 0.001 < $\alpha = 0.05$ จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปคือ ค่าพารามิเตอร์ θ_1 ในตัวแบบมีค่าไม่เท่ากับ 0 หรือ พารามิเตอร์ θ_1 ควรมีในตัวแบบ

$$H_0 : \Theta_1 = 0$$

$$H_1 : \Theta_1 \neq 0$$

จากตารางที่ 4.5 พบว่าค่า p - value = 0.000 < $\alpha = 0.05$ จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปคือ ค่าพารามิเตอร์ Θ_1 ในตัวแบบมีค่าไม่เท่ากับ 0 หรือ พารามิเตอร์ Θ_1 ควรมีในตัวแบบ

$$H_0 : \alpha_1 = 0$$

$$H_1 : \alpha_1 \neq 0$$

จากตารางที่ 4.5 พบว่าค่า p - value = 0.000 < $\alpha = 0.05$ จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปคือ ค่าพารามิเตอร์ α_1 ในตัวแบบมีค่าไม่เท่ากับ 0 หรือ พารามิเตอร์ α_1 ควรมีในตัวแบบ ดังนั้นตัวแบบ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30} + \alpha_1 + X_1$ เป็นตัวแบบที่เหมาะสม

4) การตรวจสอบความเหมาะสมของตัวแบบ

จากนั้นทำการตรวจสอบค่าความคลาดเคลื่อนว่าจะต้องไม่มีสหสัมพันธ์ในตัวเอง ด้วย Box-Pierce and Ljung-Box Tests แสดงดังตารางที่ 4.6

ตารางที่ 4.6 ทดสอบความเป็นอิสระของค่าความคลาดเคลื่อนของตัวแบบ ด้วย Box-Pierce and Ljung Box Test

Modified Box-Pierce (Ljung-Box) Chi-Square Statistics	
Q	0.01
p-value	0.94

เอกสารนี้เป็นเอกสารที่สงวนไว้สมมติฐานในการทดสอบศึกษาเท่านั้น ไม่นุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมี $H_0 : \rho_e(1) = \rho_e(2) = \rho_e(3) = \dots = \rho_e(k) = 0$ ในทุกครั้งที่มีการนำไปใช้

(ค่าความคลาดเคลื่อนของตัวแบบไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลา)

$$H_1 : \rho_c(k) \text{ อย่างน้อยหนึ่งค่าไม่เท่ากับ } 0 \text{ สำหรับ } k = 1, 2, 3, \dots, m$$

(ค่าความคลาดเคลื่อนของตัวแบบมีสหสัมพันธ์ในตัวเองทุกช่วงเวลา)

จากตารางที่ 4.6 พบว่าค่า p -value = 0.94 > $\alpha = 0.05$ จึงยอมรับ H_0

ที่ระดับนัยสำคัญ 0.05 สรุปว่า ค่าความคลาดเคลื่อนของตัวแบบที่พิจารณาไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลา ดังนั้นแสดงว่าตัวแบบ $ARIMA(2, 2, 2) \times SARIMA(1, 2, 2)_{30} + \alpha_1 + X_1$ เป็นตัวแบบที่เหมาะสม จากการทดสอบจะได้พารามิเตอร์ที่เหมาะสมที่สุดดังตารางที่ 4.7

ตารางที่ 4.7 ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดของตัวแบบ SARIMA

ตัวแบบ	พารามิเตอร์						
	p	d	q	P	D	Q	Exogenous Variables
ARIMA (p, d, q) SARIMA (P, D, Q)s With Exogenous	2	2	2	1	2	2	α_1 วันเสาร์ - อาทิตย์

ดังนั้นสมการพยากรณ์ด้วยตัวแบบ SARIMAX คือ

$$Y_t = (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} - \phi_1 Y_{t-31} - (1 + \phi_1)Y_{t-31} + \phi_1 Y_{t-32} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} + \theta_1 \Theta_1 \varepsilon_{t-31} + 4555.765 X_1$$

4.3 ผลการทดลองวิธี Random Forest

จากที่กล่าวใน 3.6.3 การพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพวิธี Random Forest จะทดสอบเพื่อให้ได้ตัวแบบที่เหมาะสมในการพยากรณ์โดยเลือกกำหนดค่าไฮเปอร์พารามิเตอร์ ดังตารางที่ 4.8

ตารางที่ 4.8 กำหนดค่าไฮเปอร์พารามิเตอร์หรือกระบวนการในการสร้างตัวแบบ

ข้อกำหนดต่าง ๆ ในการสร้างตัวแบบ	ค่าไฮเปอร์พารามิเตอร์
จำนวนขั้นตอนเวลาก่อนหน้า (Time Step)	1
จำนวนต้นไม้ตัดสินใจ (Decision Tree)	1000

จากตารางที่ 4.8 แสดงการกำหนดค่าไฮเปอร์พารามิเตอร์สำหรับการสร้างตัวแบบ โดยผู้วิจัยได้ข้อกำหนดต่าง ๆ ในการสร้างตัวแบบไว้ดังนี้

- 1) จำนวนขั้นตอนเวลาก่อนหน้า (Time Step) : กำหนดจำนวนขั้นตอนเวลาก่อนหน้าของทุก ๆ ต้นไม้ตัดสินใจ โดยผู้จัดทำได้ทำการศึกษากำหนดจำนวนขั้นตอนเวลาก่อนหน้า ตั้งแต่ 1 ถึง 10 ขั้นตอนเวลา โดยพิจารณาเปรียบเทียบจากค่า MAPE ของตัวแบบที่มีจำนวนขั้นตอนเวลาก่อนหน้า เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พบว่า เมื่อจำนวนขั้นตอนเวลาก่อนหน้า เท่ากับ 1 จะให้ค่า MAPE ที่ต่ำที่สุด ซึ่งจะได้ตัวแบบที่ดีที่สุด ดังนั้นผู้วิจัยจึงเลือกใช้จำนวนขั้นตอนเวลาก่อนหน้า เท่ากับ 1

2) จำนวนต้นไม้ตัดสินใจ (Decision Tree) : กำหนดจำนวนต้นไม้ตัดสินใจ โดยผู้จัดทำได้ทำการศึกษากำหนดจำนวนต้นไม้ตัดสินใจ ตั้งแต่ 200 ถึง 3000 ต้น โดยพิจารณาเปรียบเทียบจากค่า MAPE ของตัวแบบที่มีจำนวนต้นไม้ตัดสินใจ พบว่าเมื่อจำนวนต้นไม้ตัดสินใจ เท่ากับ 1000 ต้นจะให้ค่า MAPE ที่ต่ำที่สุด ซึ่งจะได้ตัวแบบที่ดีที่สุด ดังนั้นผู้วิจัยจึงเลือกใช้จำนวนต้นไม้ตัดสินใจเท่ากับ 1000

ตารางที่ 4.9 ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดของตัวแบบแรนดอมฟอเรส (Random Forest)

ตัวแบบ	พารามิเตอร์	
	Time Step Previous	Decision Tree
แรนดอมฟอเรส	1	1000

4.4 ผลการทดลองวิธีหน่วยความจำระยะสั้นระยะยาว (LSTM)

จากที่กล่าวใน 3.6.4 การพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพวิธีหน่วยความจำระยะสั้นระยะยาว จะทดสอบเพื่อให้ได้ตัวแบบที่เหมาะสมในการพยากรณ์ โดยเลือกกำหนดค่าไฮเปอร์พารามิเตอร์ ได้ดังตารางที่ 4.10

ตารางที่ 4.10 กำหนดค่าไฮเปอร์พารามิเตอร์หรือกระบวนการในการสร้างตัวแบบ

ข้อกำหนดต่าง ๆ ในการสร้างตัวแบบ	ค่าไฮเปอร์พารามิเตอร์
อัลกอริทึม (Algorithm)	LSTM
จำนวน Epoch	20
จำนวน Batch size	1
จำนวน Neurons	900
ฟังก์ชันการสูญเสีย (Loss Function)	Mean Absolute error
Optimizer	Adam

จากตารางที่ 4.10 แสดงการกำหนดค่าไฮเปอร์พารามิเตอร์สำหรับการสร้างตัวแบบ โดยผู้วิจัยได้ข้อกำหนดต่างๆ ในการสร้างตัวแบบไว้ดังนี้

- 1) อัลกอริทึม (Algorithm) : หน่วยความจำระยะสั้นระยะยาว
 - 2) จำนวน Epoch : ทำการกำหนดจำนวน Epoch ตั้งแต่ 1 ถึง 1000 โดยพิจารณาเปรียบเทียบจากค่า MAPE ของตัวแบบที่มีจำนวน Epoch อยู่ในช่วงที่ศึกษา พบว่าเมื่อจำนวน Epoch เท่ากับ 20 จะให้ค่า MAPE ที่ต่ำที่สุด ซึ่งจะได้ตัวแบบที่ดีที่สุด ดังนั้นผู้วิจัยจึงเลือกใช้จำนวน Epoch เท่ากับ 20
- ไม่ว่ากรณีใดๆ ทั้งสิ้นผู้วิจัยจึงเลือกใช้จำนวน Epoch เท่ากับ 20 ถึงเจ้าของเอกสารทุกครั้งที่มีกรนำมาใช้

- 3) จำนวน Batch size : เนื่องจากข้อมูลจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เป็นข้อมูลอนุกรมเวลาตัวแปรเดียว (Univariate Time Series) จึงกำหนดจำนวน Batch size เท่ากับ 1
 - 4) จำนวน Neurons : ทำการกำหนดจำนวน Neurons ตั้งแต่ 100 ถึง 900 โดยพิจารณาเปรียบเทียบจากค่า MAPE ของตัวแบบที่มีจำนวน Neurons อยู่ในช่วงที่ศึกษา พบว่าเมื่อจำนวน Neurons เท่ากับ 900 จะให้ค่า MAPE ที่ต่ำที่สุด ซึ่งจะได้ตัวแบบที่ดีที่สุด ดังนั้นผู้วิจัยจึงเลือกใช้จำนวน Neurons เท่ากับ 900
 - 5) ฟังก์ชันการสูญเสีย (Loss Function) : ผู้วิจัยได้กำหนดเป็น Mean Absolute error
 - 6) Optimizer : ผู้วิจัยได้กำหนดเป็น Adam
- จากการทดลองจะได้พารามิเตอร์ที่เหมาะสมที่สุดดังตารางที่ 4.11

ตารางที่ 4.11 ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดของตัวแบบการเรียนรู้เชิงลึก วิธีหน่วยความจำระยะสั้น ระยะยาว

ตัวแบบ	พารามิเตอร์		
	Epoch	Batch size	Neurons
การเรียนรู้เชิงลึก วิธีหน่วยความจำระยะสั้น ระยะยาว	20	1	900

4.5 การเปรียบเทียบผลประสิทธิภาพของตัวแบบพยากรณ์

หลังจากทดลองหาค่าพารามิเตอร์ที่ดีที่สุดของตัวแบบทั้ง 4 วิธี ได้แก่ ตัวแบบ SARIMA, SARIMAX, Random Forest และตัวแบบ LSTM จากนั้นทำการเปรียบเทียบความแม่นยำของค่าพยากรณ์โดยใช้ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ ดังตารางที่ 4.12

ตารางที่ 4.12 ค่า MAPE ของแต่ละตัวแบบ

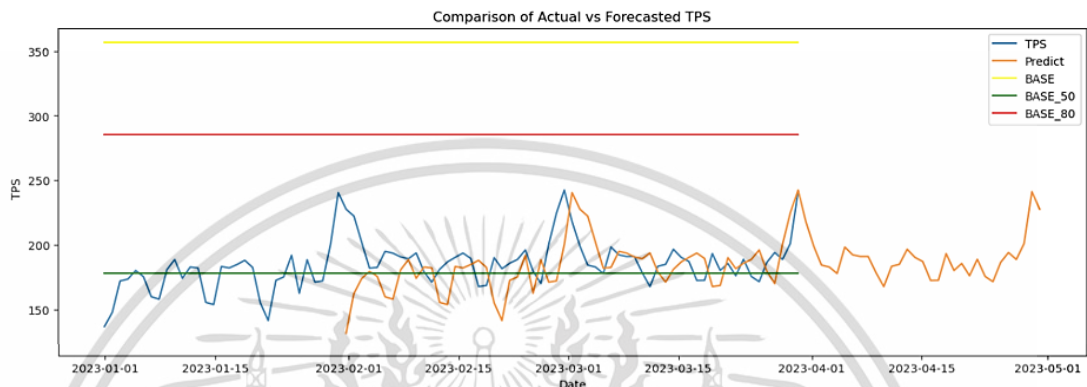
ตัวแบบ	MAPE
SARIMA	5.3796%
SARIMAX	4.9051%
Random Forest	4.8174%
LSTM	5.0942%

เมื่อพิจารณาค่า MAPE ที่ดีที่สุดของแต่ละตัวแบบ เพื่อเลือกแบบจำลองที่มีค่าความคลาดเคลื่อนที่น้อยที่สุดนั้น จะเห็นว่า แบบจำลองที่เหมาะสมที่สุดในจำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส ได้แบบจำลองที่ดีที่สุด คือ วิธีการเรียนรู้ของเครื่อง (Machine Learning) ตัวแบบ Random Forest ที่มีค่า MAPE เท่ากับ 4.8174%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.6 ผลการทดลองจากตัวแบบที่ดีที่สุดในการพยากรณ์รายวันใน 30 วันข้างหน้า

การเปรียบเทียบผลประสิทธิภาพของตัวแบบพยากรณ์ โดยทำการคัดเลือกตัวแบบการพยากรณ์ที่ดีที่สุด ซึ่งจากการเปรียบเทียบผลประสิทธิภาพการพยากรณ์ของตัวแบบพยากรณ์ พบว่าตัวแบบ Random Forest ให้ค่า MAPE เท่ากับ 4.8174% จึงนำ ตัวแบบ Random Forest ไปใช้พยากรณ์รายวันใน 30 วันข้างหน้า แสดงดังรูปที่ 4.15



รูปที่ 4.15 ผลการพยากรณ์รายวันในอีก 30 วันข้างหน้า

หมายเหตุ	ค่าจริง	ค่าพยากรณ์	50 เปอร์เซ็นต์ของขีดความสามารถที่รับได้ก่อนระบบจะล่ม	ขีดความสามารถที่รับได้ก่อนระบบจะล่ม	80 เปอร์เซ็นต์ของขีดความสามารถที่รับได้ก่อนระบบจะล่ม
	■	■	■	■	■

จากรูปที่ 4.15 การพยากรณ์รายวันใน 30 วันข้างหน้า ได้การพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ด้วยตัวแบบ Random Forest โดยใช้ผลการพยากรณ์ของวันก่อนหน้าเป็นข้อมูลนำเข้า ซึ่งสามารถนำรูปแบบจำลองจากตัวแบบที่ได้ไปเป็นแนวทางหรือข้อมูลเพิ่มเติมในการตัดสินใจในการปรับปรุงการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อที่จะศึกษาตัวแบบที่เหมาะสมสำหรับการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส เพื่อนำไปตัวแบบที่ได้ไปเป็นแนวทางหรือข้อมูลเพิ่มเติมในการตัดสินใจในการวางแผนเพื่อช่วยปรับปรุงค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ ดังนั้นผู้วิจัยได้ทำการศึกษาและเลือกตัวแบบที่เหมาะสมที่สุดในการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส ด้วยการเปรียบเทียบความแม่นยำของค่าพยากรณ์ที่ได้จาก 4 ตัวแบบ คือ ตัวแบบ SARIMA, ตัวแบบ SARIMAX, ตัวแบบ Random Forest และตัวแบบ LSTM โดยพิจารณาจากค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error: MAPE) ซึ่งตัวแบบที่มีค่าความคลาดเคลื่อนในการพยากรณ์น้อยที่สุดนั้นคือ ตัวแบบที่มีค่า MAPE ต่ำที่สุด จึงสามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

ผลการเปรียบเทียบของตัวแบบพยากรณ์ทั้ง 4 ตัวแบบ สามารถสรุปผลการดำเนินวิจัยได้ดังนี้

1) ตัวแบบ SARIMA

การวิเคราะห์หารูปแบบการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสพบว่า รูปแบบที่เหมาะสมที่สุด คือ $ARIMA(2,2,2) \times SARIMA(1,2,2)_{30}$ เนื่องจากมีค่า AIC ที่ต่ำเท่ากับ 4783.262 และเมื่อทดสอบค่าพารามิเตอร์และ Ljung-Box พบว่าตัวแบบมีความเหมาะสม กล่าวคือ ค่าความคลาดเคลื่อนของตัวแบบที่พิจารณาไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลาและเมื่อนำไปทดสอบความคลาดเคลื่อนของตัวแบบให้ค่า MAPE เท่ากับ 5.3796%

2) ตัวแบบ SARIMAX

การวิเคราะห์หารูปแบบการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสพบว่า รูปแบบที่เหมาะสมที่สุด คือ $ARIMA(2,2,2) \times SARIMA(1,2,2)_{30}$ และตัวแปรปัจจัยภายนอก ได้แก่ X1 คือ วันเสาร์ – อาทิตย์ และเมื่อทดสอบค่าพารามิเตอร์และ Ljung-Box พบว่าตัวแบบมีความเหมาะสม กล่าวคือ ค่าความคลาดเคลื่อนของตัวแบบที่พิจารณาไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลา และเมื่อนำไปทดสอบความคลาดเคลื่อนของตัวแบบให้ค่า MAPE เท่ากับ 4.9051% Lewis-Beck Rice (1982)

การพยากรณ์โดยวิธี Box-Jenkins เป็นการพยากรณ์ที่เหมาะสมข้อมูลอนุกรมเวลาทุกประเภทและใช้ข้อมูลในการพยากรณ์จำนวนมาก เพื่อให้ค่าพยากรณ์ที่มีความแม่นยำสูง ตัวแบบที่ศึกษาในงานวิจัย คือ ตัวแบบ SARIMA และ SARIMAX ซึ่งเป็นตัวแบบที่เหมาะสมกับข้อมูลที่มีความแปร

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อเผยแพร่ให้ท่านไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมีเหตุขัดแย้งและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งหากท่านไปใช้

ผ่านทางฤดูกาลมาที่เกี่ยวข้อง และงานวิจัยครั้งนี้มีจำนวนข้อมูลที่ใช้ศึกษามี 797 ค่า ทำให้ค่าความคลาดเคลื่อนในการพยากรณ์มีค่าอยู่ระหว่าง 4.90% - 5.37% ซึ่งมีค่าความคลาดเคลื่อนสูงกว่า ตัวแบบ Random Forest และ ตัวแบบ LSTM

3) ตัวแบบ Random Forest

การวิเคราะห์หารูปแบบการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสพบว่า รูปแบบที่เหมาะสมที่สุด คือ ตัวแบบ Random Forest ที่มีจำนวนขั้นตอนเวลาก่อนหน้า 1 ขั้นตอนเวลา และมีจำนวนต้นไม้ตัดสินใจจำนวน 1000 ต้น และเมื่อนำไปทดสอบความคลาดเคลื่อนของตัวแบบให้ค่า MAPE เท่ากับ 4.8174% Lewis-Beck Rice (1982)

การพยากรณ์โดยตัวแบบ Random Forest สามารถใช้ได้กับข้อมูลที่หลากหลาย เช่น ตาราง รูปภาพ และข้อความ เป็นตัวแบบที่นำข้อมูลมาสร้างต้นไม้ตัดสินใจแบบสุ่ม ในแต่ละต้นไม้ตัดสินใจข้อมูลนำเข้าจะมีค่าไม่เหมือนกันและเลือกผลลัพธ์จากค่าคลาดเคลื่อนที่น้อยที่สุดมาแสดงผล ในงานวิจัยครั้งนี้ค่าคลาดเคลื่อนของตัวแบบ Random Forest มีค่าอยู่ที่ 4.8174% ซึ่งเป็นตัวแบบที่ให้ค่าคลาดเคลื่อนการพยากรณ์ต่ำที่สุดสำหรับในงานวิจัยนี้

4) ตัวแบบ LSTM

การวิเคราะห์หารูปแบบการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสพบว่า รูปแบบที่เหมาะสมที่สุด คือ ตัวแบบ LSTM ที่มีจำนวน Epoch เท่ากับ 20, มีจำนวน Batch size เท่ากับ 1, มีจำนวน Neurons เท่ากับ 900 และเมื่อนำไปทดสอบความคลาดเคลื่อนของตัวแบบให้ค่า MAPE เท่ากับ 5.0942%

การพยากรณ์โดยตัวแบบ LSTM เป็นตัวที่มีการทำงานแบบจำลองเซลล์ประสาทของมนุษย์ที่มีความยุ่งยากซับซ้อน เข้าใจได้ยาก LSTM สามารถจัดการกับข้อมูลได้ว่า ข้อมูลที่ควรจดจำหรือควรลืม ข้อมูลบางอย่างไม่เกี่ยวข้องก็จะไม่นำมาพิจารณา ทำให้ตัวแบบ LSTM มีความคลาดเคลื่อนในการพยากรณ์ต่ำ สำหรับงานวิจัยการพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิส มีค่าความคลาดเคลื่อนอยู่ที่ 5.0942% ซึ่งมีค่าที่ดีกว่าวิธีการพยากรณ์บ็อก - เจนกินส์

เมื่อทำการเปรียบเทียบค่า MAPE ของตัวแบบทั้ง 4 วิธี โดยพิจารณาค่า MAPE ที่ต่ำที่สุด ซึ่งแสดงว่าตัวแบบนั้นมีค่าความคลาดเคลื่อนน้อยที่สุด พบว่า ตัวแบบ Random Forest ให้ค่า MAPE ต่ำที่สุด จึงนำตัวแบบที่เหมาะสมนี้ไปพยากรณ์ใน 30 วันข้างหน้า เพื่อนำไปเป็นแนวทางหรือข้อมูลเพิ่มเติมในการตัดสินใจในการวางแผนเพื่อช่วยปรับปรุงค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ เมื่อเราสามารถพยากรณ์จำนวนการเรียกใช้เซอร์วิสของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซอร์วิสที่จะเกิดขึ้นในอนาคตได้ องค์กรก็จะสามารถวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์ได้ เพราะการพยากรณ์ล่วงหน้าจะทำให้องค์กรทราบว่าในอีกกี่เดือนข้างหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอญูยาดให้นำไปใช้ประโยชน์ด้านการค้า
ระบบจะล้ม
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 อภิปรายผล

จากการวิจัยการศึกษาและเปรียบเทียบการพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบประสิทธิภาพของเซิร์ฟเวอร์ ด้วยวิธีการของบ็อกซ์เจนกินส์ โดยใช้ตัวแบบ SARIMA ตัวแบบ SARIMAX ตัวแบบ Random Forest และตัวแบบ LSTM โดยพิจารณาจากค่า MAPE ผลการวิจัยพบว่าการพยากรณ์โดยใช้ตัวแบบ Random Forest ให้ค่าความคลาดเคลื่อนต่ำที่สุด ซึ่งสอดคล้องกับงานวิจัยของภรณ์ยา (2559) ที่ได้ศึกษาการเพิ่มประสิทธิภาพ เทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุล โดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดต่ออินเทอร์เน็ต และงานวิจัยของอัศวพล (2562) ที่ได้ศึกษาเทคนิคพยากรณ์การได้รับปัจจัยพื้นฐานนักเรียนยากจนของนักเรียน โรงเรียนวัดพระขาว (ประชาชนุเคราะห์) ด้วยเทคนิคเหมืองข้อมูล ซึ่งงานวิจัยทั้ง 2 งานวิจัยให้ผลการวิจัยว่าตัวแบบ Random Forest ให้ค่าความคลาดเคลื่อนน้อยในการพยากรณ์น้อยกว่าวิธีการของบ็อกซ์เจนกินส์ โดยใช้ตัวแบบ SARIMA, ตัวแบบ SARIMAX และตัวแบบ LSTM

5.3 ข้อเสนอแนะ

- 1) นำข้อมูลเซิร์ฟเวอร์ทั้งหมดมาพยากรณ์ เพื่อใช้ในการวางแผนค่าใช้จ่ายทรัพยากรของเซิร์ฟเวอร์
- 2) นำผลการพยากรณ์ที่ได้ ไปจัดทำรีพอร์ตเพื่อใช้เป็นข้อมูลประกอบการวางแผน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- กัลยา วานิชย์บัณฑิต. 2554. **หลักสถิติ**. พิมพ์ครั้งที่ 6. กรุงเทพมหานคร.: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย
- กฤษดา. 2556 . **ขั้นตอนในการพยากรณ์**. กรุงเทพฯ.: อินเทอร์เน็ตมีเดีย.
- กสิณ คงเกียรติขจร. **Forecasting (การพยากรณ์)**. [Online]. เข้าถึงได้จาก <https://sites.google.com/site/kmaths55><https://sites.google.com/site/kmaths55/mba>.
- จักภัทรสรณ์ย จุฑาธิปไตย. 2561. **การพยากรณ์ปริมาณการใช้ไฟฟ้าของโรงเรียนในสังกัดเมืองพัทยา โดยใช้ตัวแบบ SARIMA กับตัวแบบ SARIMAX (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต**. ม.บูรพา ชลบุรี.
- จินตามาส สุทธิชัยเมธี. 2554. **การประยุกต์ใช้ ARIMA MODEL เพื่อการวิจัย**. วารสารสุทธปริทัศน์. 25(76) : 109-111.
- เฉลิมพล จตุพร, พัฒนา สุขประเสริฐ. 2559. **ตัวแบบพยากรณ์ผลผลิตและปริมาณส่งออกยางพาราของประเทศไทย**. แก่นเกษตร 44(2)ม 219-228
- ชวริย์ รัตนวงศ์ศรี. 2556. **การพยากรณ์ราคาขายพาราในตลาดสินค้าเกษตรล่วงหน้าแห่งประเทศไทย**. วารสารมนุษยศาสตร์สังคมศาสตร์และวิทยาศาสตร์. 6(1) : 840-846.
- ทรงศิริ แต่สมบัติ. 2549. **การพยากรณ์เชิงปริมาณ**. กรุงเทพฯ. สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์. ธนาคารแห่งประเทศไทย. 2021. **การใช้ดิจิทัลแบงก์กิ้งค์**. [Online]. เข้าถึงได้จาก <https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages>.
- นิภา นีรุตติกุล. 2551. **การพยากรณ์การขาย**. กรุงเทพฯ. สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์
- บุญฤทธิ์ ชูประดิษฐ์ และเสาวภา ชัยพิทักษ์. 2561. **ตัวแบบพยากรณ์ปริมาณการส่งออกมะม่วงของประเทศไทย**. วารสารวิทยาศาสตร์และเทคโนโลยี. 26(2) : 74-58.
- มุกดา แม่นมิน. 2549. **อนุกรมเวลาและการพยากรณ์**. กรุงเทพมหานคร. สำนักพิมพ์ประกายพริก.
- วิชราภักษ์และวัชร. 2562. **การสร้างตัวแบบทางคณิตศาสตร์เพื่อพยากรณ์ปริมาณขยะในเขตเทศบาลเมืองบุรีรัมย์**. [Online]. <http://math.bru.ac.th/bru-research/การสร้างตัวแบบทางคณิต>.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

- วรางคณา กীরติวิบูลย์. 2559. การพยากรณ์ปริมาณการส่งออกเนื้อไก่แช่แข็งและแช่เย็น. วารสารวิชาการและวิจัย มทร.พระนคร, 10 : 37-50.
- สมศรี บัณฑิตวิไล. (2559). การวิเคราะห์อนุกรมเวลา: เอกสารประกอบการสอน. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. กรุงเทพมหานคร.
- Akaike, H. 1978. **A Bayesian analysis of the minimum AIC procedure.** Annals of the Institute of Statistical Mathematics. 30(1) : 9–14.
- Box, G.E.P. Jenkins, G.M. and Reinsel, G.C. 1994. **Time Series Analysis: Forecasting and Control.** 3rd ed. New Jersey : Prentice Hall.
- Big Data Thailand. 2021. การเตรียมข้อมูลเบื้องต้นด้วย Power BI (Basic Data Preparation with Power BI). [Online].
https://bigdata.go.th/big-data101/data_preparation_power_bi/
- Leo, B. 2001. **Random Forest.** [Online].
<https://www.stat.berkeley.edu/users/breiman/randomforest2001.pdf>.
- Lewis and Colin, D. 1982. **Industrial and business forecasting methods : a practical guide to exponential smoothing and curve fitting.** London : Butterworth Scientific.
- Michael, J, Natalie, P, Matthew, S and Peter, R. 2014. **Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks.** [Online].
<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-15-276>
- Oinkina and Hakyll. 2015. **Understanding LSTM Networks.** [Online].
<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
- Premjith B, Soman K. P and Prabakaran Poornachandran. 2018. **A deep learning based Part-of-Speech (POS) tagger for Sanskrit language by embedding character level features.** [Online].
<https://www.semanticscholar.org/paper/A-deep-learning-based-Part-of-Speech>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง (ต่อ)

- Qiang D, Qin D, and William P. 2002. **Decision tree classification of spatial data streams using Peano Count Trees**. Proceedings of the 2002 ACM symposium on Applied computing. : 413-417.
- Quang H, Shih-Kuei L, Jeng-Fung C, Chi-Luan Le and Luong Hoang Anh. 2020. **Forecasting Air Passenger Demand: A Comparison of LSTM and SARIMA**. Journal of Computer Science. 6(7) : 1063.1084.
- Robert, H. **Theory of the Backpropagation Neural Network**. [Online].
<https://www.semanticscholar.org/paper/Theory-of-the-backpropagation-neural-network>
- Selva, P. 2017. **ARIMA Model-Complete Guide to Time Series Forecasting in Python**. [Online]. <https://www.machinelearningplus.com/time-series/arma-model-time-series>
- Sima S, Neda T and Akbar S. 2018. **A Comparison of ARIMA and LSTM in Forecasting Time Series**. [Online].
https://www.researchgate.net/publication/330477082_A_Comparison_of_ARIMA_and_LSTM_in_Forecasting_Time_Series. (เข้าถึงเมื่อ 30 มีนาคม 2564)
- Stefan K, Tomas M, Martin K and Lukas B. 2010. **Recurrent neural network based language model**. [Online]. <https://www.semanticscholar.org/paper/Recurrent-neural-network-based>
- Umematsu, T., Sano, A., Taylor, S., and Picard, R. 2019. **Improving Students' Daily Life Stress Forecasting using LSTM Neural Networks**. [Online].
<https://www.media.mit.edu/publications/improving-students-daily-life-stress-forecasting>
- Varun, G and Ankit, P. 2018. **Crude Oil Price Prediction Using LSTM Networks**. Engineering and Technology International Journal of Computer and Information Engineering. 12(3): 226-230.
- Vasily, D, Andriy, M, Nataliia D, Vitalii, B and Albert, A. 2020. **Machine learning approaches for financial time series forecasting**. CURE-WS. 2173: 434-450
- William H. Kruskal and W. Allen Wallis. 1952. **USE OF RANKS IN ONE-CRITERION VARIANCE ANALYSIS**. Journal of the American Statistical Association. 260(47) : 584-621.

เอกสารนี้ Witten, I.H. and Frank, E. 2005. **Data mining: practical machine learning tools and techniques**. San Francisco: Elsevier. ไม่ว่ากรณีใดๆ ที่เข้าถึงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

ตารางภาคผนวก ก ที่ 1 ตารางค่า AIC ในแต่ละอันดับของตัวแบบ
ARIMA(p,d,q)SARIMA(P,D,Q)s

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(0,0,0)x SARIMA(0,0,0) ₃₀	8754.419
ARIMA(0,0,0)x SARIMA(0,0,1) ₃₀	7681.964
ARIMA(0,0,0)x SARIMA(0,0,2) ₃₀	6893.5455
ARIMA(0,0,0)x SARIMA(0,1,0) ₃₀	5942.0099
ARIMA(0,0,0)x SARIMA(0,1,1) ₃₀	5652.0295
ARIMA(0,0,0)x SARIMA(0,1,2) ₃₀	5408.9028
ARIMA(0,0,0)x SARIMA(0,2,0) ₃₀	6271.3399
ARIMA(0,0,0)x SARIMA(0,2,1) ₃₀	5484.8945
ARIMA(0,0,0)x SARIMA(0,2,2) ₃₀	5125.7574
ARIMA(0,0,0)x SARIMA(1,0,0) ₃₀	5928.3614
ARIMA(0,0,0)x SARIMA(1,0,1) ₃₀	5808.6207
ARIMA(0,0,0)x SARIMA(1,0,2) ₃₀	5565.6586
ARIMA(0,0,0)x SARIMA(1,1,0) ₃₀	5661.5403
ARIMA(0,0,0)x SARIMA(1,1,1) ₃₀	5652.7642
ARIMA(0,0,0)x SARIMA(1,1,2) ₃₀	5388.2325
ARIMA(0,0,0)x SARIMA(1,2,0) ₃₀	5763.2234
ARIMA(0,0,0)x SARIMA(1,2,1) ₃₀	5416.4973
ARIMA(0,0,0)x SARIMA(1,2,2) ₃₀	5127.3602
ARIMA(0,0,0)x SARIMA(2,0,0) ₃₀	5602.5270
ARIMA(0,0,0)x SARIMA(2,0,1) ₃₀	5573.8689
ARIMA(0,0,0)x SARIMA(2,0,2) ₃₀	5561.2251
ARIMA(0,0,0)x SARIMA(2,1,0) ₃₀	5417.3206
ARIMA(0,0,0)x SARIMA(2,1,1) ₃₀	5416.8953
ARIMA(0,0,0)x SARIMA(2,1,2) ₃₀	5366.3193
ARIMA(0,0,0)x SARIMA(2,2,0) ₃₀	5312.3130
ARIMA(0,0,0)x SARIMA(2,2,1) ₃₀	5128.8582
ARIMA(0,0,0)x SARIMA(2,2,2) ₃₀	5121.9113
ARIMA(0,0,1)x SARIMA(0,0,0) ₃₀	7856.2074
ARIMA(0,0,1)x SARIMA(0,0,1) ₃₀	6934.9437
ARIMA(0,0,1)x SARIMA(0,0,2) ₃₀	6330.4472

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาวิจัยเท่านั้น ไม่อนุญาตให้นำไปเผยแพร่โดยไม่ได้รับอนุญาต การค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่มีให้ตีพิมพ์ลงนิตยสาร และต้องยังอิงถึงเจ้าของเอกสารทุกครั้งที่มีกรนำไปใช้

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(0,0,1)x SARIMA(0,1,0) ₃₀	5634.7594
ARIMA(0,0,1)x SARIMA(0,1,1) ₃₀	5312.4039
ARIMA(0,0,1)x SARIMA(0,1,2) ₃₀	5085.9698
ARIMA(0,0,1)x SARIMA(0,2,0) ₃₀	5995.0514
ARIMA(0,0,1)x SARIMA(0,2,1) ₃₀	5215.7765
ARIMA(0,0,1)x SARIMA(0,2,2) ₃₀	4843.1637
ARIMA(0,0,1)x SARIMA(1,0,0) ₃₀	5645.7995
ARIMA(0,0,1)x SARIMA(1,0,1) ₃₀	5488.1758
ARIMA(0,0,1)x SARIMA(1,0,2) ₃₀	5236.5533
ARIMA(0,0,1)x SARIMA(1,1,0) ₃₀	5344.3463
ARIMA(0,0,1)x SARIMA(1,1,1) ₃₀	5313.6865
ARIMA(0,0,1)x SARIMA(1,1,2) ₃₀	5083.9884
ARIMA(0,0,1)x SARIMA(1,2,0) ₃₀	5509.2980
ARIMA(0,0,1)x SARIMA(1,2,1) ₃₀	5143.2432
ARIMA(0,0,1)x SARIMA(1,2,2) ₃₀	4841.7213
ARIMA(0,0,1)x SARIMA(2,0,0) ₃₀	5317.3965
ARIMA(0,0,1)x SARIMA(2,0,1) ₃₀	5259.5446
ARIMA(0,0,1)x SARIMA(2,0,2) ₃₀	5232.9938
ARIMA(0,0,1)x SARIMA(2,1,0) ₃₀	5088.1329
ARIMA(0,0,1)x SARIMA(2,1,1) ₃₀	5090.1329
ARIMA(0,0,1)x SARIMA(2,1,2) ₃₀	5063.0576
ARIMA(0,0,1)x SARIMA(2,2,0) ₃₀	5081.5046
ARIMA(0,0,1)x SARIMA(2,2,1) ₃₀	4859.3924
ARIMA(0,0,1)x SARIMA(2,2,2) ₃₀	4841.5239
ARIMA(0,0,2)x SARIMA(0,0,0) ₃₀	7274.2867
ARIMA(0,0,2)x SARIMA(0,0,1) ₃₀	6536.8419
ARIMA(0,0,2)x SARIMA(0,0,2) ₃₀	6059.1412
ARIMA(0,0,2)x SARIMA(0,1,0) ₃₀	5600.3997
ARIMA(0,0,2)x SARIMA(0,1,1) ₃₀	5253.5136
ARIMA(0,0,2)x SARIMA(0,1,2) ₃₀	5029.0530
ARIMA(0,0,2)x SARIMA(0,2,0) ₃₀	5972.2606
ARIMA(0,0,2)x SARIMA(0,2,1) ₃₀	5188.5757
ARIMA(0,0,2)x SARIMA(0,2,2) ₃₀	4794.9283
ARIMA(0,0,2)x SARIMA(1,0,0) ₃₀	5622.7171
ARIMA(0,0,2)x SARIMA(1,0,1) ₃₀	5429.2878
ARIMA(0,0,2)x SARIMA(1,0,2) ₃₀	5178.5823

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(0,0,2)x SARIMA(1,1,0) ₃₀	5305.2055
ARIMA(0,0,2)x SARIMA(1,1,1) ₃₀	5255.1340
ARIMA(0,0,2)x SARIMA(1,1,2) ₃₀	5026.2919
ARIMA(0,0,2)x SARIMA(1,2,0) ₃₀	5490.8531
ARIMA(0,0,2)x SARIMA(1,2,1) ₃₀	5109.2023
ARIMA(0,0,2)x SARIMA(1,2,2) ₃₀	4796.0177
ARIMA(0,0,2)x SARIMA(2,0,0) ₃₀	5288.2108
ARIMA(0,0,2)x SARIMA(2,0,1) ₃₀	5223.7533
ARIMA(0,0,2)x SARIMA(2,0,2) ₃₀	5175.5903
ARIMA(0,0,2)x SARIMA(2,1,0) ₃₀	5037.9654
ARIMA(0,0,2)x SARIMA(2,1,1) ₃₀	5039.1238
ARIMA(0,0,2)x SARIMA(2,1,2) ₃₀	5012.4134
ARIMA(0,0,2)x SARIMA(2,2,0) ₃₀	5060.4529
ARIMA(0,0,2)x SARIMA(2,2,1) ₃₀	4829.5019
ARIMA(0,0,2)x SARIMA(2,2,2) ₃₀	4795.5245
ARIMA(0,1,0)x SARIMA(0,0,0) ₃₀	5512.9082
ARIMA(0,1,0)x SARIMA(0,0,1) ₃₀	5266.8412
ARIMA(0,1,0)x SARIMA(0,0,2) ₃₀	5048.3432
ARIMA(0,1,0)x SARIMA(0,1,0) ₃₀	5672.3612
ARIMA(0,1,0)x SARIMA(0,1,1) ₃₀	5100.3864
ARIMA(0,1,0)x SARIMA(0,1,2) ₃₀	4879.3506
ARIMA(0,1,0)x SARIMA(0,2,0) ₃₀	6086.9246
ARIMA(0,1,0)x SARIMA(0,2,1) ₃₀	5274.3772
ARIMA(0,1,0)x SARIMA(0,2,2) ₃₀	4781.8523
ARIMA(0,1,0)x SARIMA(1,0,0) ₃₀	5274.6466
ARIMA(0,1,0)x SARIMA(1,0,1) ₃₀	5268.7777
ARIMA(0,1,0)x SARIMA(1,0,2) ₃₀	5050.3380
ARIMA(0,1,0)x SARIMA(1,1,0) ₃₀	5299.9725
ARIMA(0,1,0)x SARIMA(1,1,1) ₃₀	5101.8446
ARIMA(0,1,0)x SARIMA(1,1,2) ₃₀	4867.6818
ARIMA(0,1,0)x SARIMA(1,2,0) ₃₀	5559.5167
ARIMA(0,1,0)x SARIMA(1,2,1) ₃₀	5167.3958
ARIMA(0,1,0)x SARIMA(1,2,2) ₃₀	4783.5310
ARIMA(0,1,0)x SARIMA(2,0,0) ₃₀	5056.2327
ARIMA(0,1,0)x SARIMA(2,0,1) ₃₀	5058.0382
ARIMA(0,1,0)x SARIMA(2,0,2) ₃₀	5051.4494

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(0,1,0)x SARIMA(2,1,0) ₃₀	4961.8228
ARIMA(0,1,0)x SARIMA(2,1,1) ₃₀	4885.3947
ARIMA(0,1,0)x SARIMA(2,1,2) ₃₀	4869.6244
ARIMA(0,1,0)x SARIMA(2,2,0) ₃₀	5104.6142
ARIMA(0,1,0)x SARIMA(2,2,1) ₃₀	4837.9976
ARIMA(0,1,0)x SARIMA(2,2,2) ₃₀	4807.0100
ARIMA(0,1,1)x SARIMA(0,0,0) ₃₀	5452.2158
ARIMA(0,1,1)x SARIMA(0,0,1) ₃₀	5208.0182
ARIMA(0,1,1)x SARIMA(0,0,2) ₃₀	4994.7838
ARIMA(0,1,1)x SARIMA(0,1,0) ₃₀	5552.7715
ARIMA(0,1,1)x SARIMA(0,1,1) ₃₀	5010.0509
ARIMA(0,1,1)x SARIMA(0,1,2) ₃₀	4793.7752
ARIMA(0,1,1)x SARIMA(0,2,0) ₃₀	5967.7028
ARIMA(0,1,1)x SARIMA(0,2,1) ₃₀	5165.2258
ARIMA(0,1,1)x SARIMA(0,2,2) ₃₀	4694.8304
ARIMA(0,1,1)x SARIMA(1,0,0) ₃₀	5222.6119
ARIMA(0,1,1)x SARIMA(1,0,1) ₃₀	5210.0164
ARIMA(0,1,1)x SARIMA(1,0,2) ₃₀	4995.1954
ARIMA(0,1,1)x SARIMA(1,1,0) ₃₀	5195.5491
ARIMA(0,1,1)x SARIMA(1,1,1) ₃₀	5011.2890
ARIMA(0,1,1)x SARIMA(1,1,2) ₃₀	4784.1293
ARIMA(0,1,1)x SARIMA(1,2,0) ₃₀	5457.9453
ARIMA(0,1,1)x SARIMA(1,2,1) ₃₀	5058.5157
ARIMA(0,1,1)x SARIMA(1,2,2) ₃₀	4694.4348
ARIMA(0,1,1)x SARIMA(2,0,0) ₃₀	5008.5068
ARIMA(0,1,1)x SARIMA(2,0,1) ₃₀	5009.2245
ARIMA(0,1,1)x SARIMA(2,0,2) ₃₀	4989.9826
ARIMA(0,1,1)x SARIMA(2,1,0) ₃₀	4858.7892
ARIMA(0,1,1)x SARIMA(2,1,1) ₃₀	4811.3929
ARIMA(0,1,1)x SARIMA(2,1,2) ₃₀	4785.2610
ARIMA(0,1,1)x SARIMA(2,2,0) ₃₀	4995.0075
ARIMA(0,1,1)x SARIMA(2,2,1) ₃₀	4739.0127
ARIMA(0,1,1)x SARIMA(2,2,2) ₃₀	4695.4101
ARIMA(0,1,2)x SARIMA(0,0,0) ₃₀	5403.1674
ARIMA(0,1,2)x SARIMA(0,0,1) ₃₀	5145.0359
ARIMA(0,1,2)x SARIMA(0,0,2) ₃₀	4933.7805

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น

ARIMA (p, d, q) SARIMA (P, D, Q) _s	AIC
ARIMA(0,1,2)x SARIMA(0,1,0) ₃₀	5482.2455
ARIMA(0,1,2)x SARIMA(0,1,1) ₃₀	4941.0046
ARIMA(0,1,2)x SARIMA(0,1,2) ₃₀	4722.8535
ARIMA(0,1,2)x SARIMA(0,2,0) ₃₀	5893.9989
ARIMA(0,1,2)x SARIMA(0,2,1) ₃₀	5098.2466
ARIMA(0,1,2)x SARIMA(0,2,2) ₃₀	4633.9470
ARIMA(0,1,2)x SARIMA(1,0,0) ₃₀	5165.4861
ARIMA(0,1,2)x SARIMA(1,0,1) ₃₀	5146.7172
ARIMA(0,1,2)x SARIMA(1,0,2) ₃₀	4929.8236
ARIMA(0,1,2)x SARIMA(1,1,0) ₃₀	5138.0468
ARIMA(0,1,2)x SARIMA(1,1,1) ₃₀	4941.0629
ARIMA(0,1,2)x SARIMA(1,1,2) ₃₀	4710.0518
ARIMA(0,1,2)x SARIMA(1,2,0) ₃₀	5399.2393
ARIMA(0,1,2)x SARIMA(1,2,1) ₃₀	4996.8615
ARIMA(0,1,2)x SARIMA(1,2,2) ₃₀	4633.4855
ARIMA(0,1,2)x SARIMA(2,0,0) ₃₀	4952.0248
ARIMA(0,1,2)x SARIMA(2,0,1) ₃₀	4951.0653
ARIMA(0,1,2)x SARIMA(2,0,2) ₃₀	4923.3356
ARIMA(0,1,2)x SARIMA(2,1,0) ₃₀	4811.9143
ARIMA(0,1,2)x SARIMA(2,1,1) ₃₀	4746.4954
ARIMA(0,1,2)x SARIMA(2,1,2) ₃₀	4717.5538
ARIMA(0,1,2)x SARIMA(2,2,0) ₃₀	4959.9870
ARIMA(0,1,2)x SARIMA(2,2,1) ₃₀	4695.5421
ARIMA(0,1,2)x SARIMA(2,2,2) ₃₀	4635.4568
ARIMA(0,2,0)x SARIMA(0,0,0) ₃₀	6119.1039
ARIMA(0,2,0)x SARIMA(0,0,1) ₃₀	5838.6106
ARIMA(0,2,0)x SARIMA(0,0,2) ₃₀	5588.2332
ARIMA(0,2,0)x SARIMA(0,1,0) ₃₀	6264.4693
ARIMA(0,2,0)x SARIMA(0,1,1) ₃₀	5661.6898
ARIMA(0,2,0)x SARIMA(0,1,2) ₃₀	5409.1812
ARIMA(0,2,0)x SARIMA(0,2,0) ₃₀	6645.8226
ARIMA(0,2,0)x SARIMA(0,2,1) ₃₀	5810.3522
ARIMA(0,2,0)x SARIMA(0,2,2) ₃₀	5291.5459
ARIMA(0,2,0)x SARIMA(1,0,0) ₃₀	5847.2411
ARIMA(0,2,0)x SARIMA(1,0,1) ₃₀	5840.5402
ARIMA(0,2,0)x SARIMA(1,0,2) ₃₀	5590.2043

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษามาก่อน ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(0,2,0)x SARIMA(1,1,0) ₃₀	5870.7875
ARIMA(0,2,0)x SARIMA(1,1,1) ₃₀	5661.2941
ARIMA(0,2,0)x SARIMA(1,1,2) ₃₀	5396.6332
ARIMA(0,2,0)x SARIMA(1,2,0) ₃₀	6096.3236
ARIMA(0,2,0)x SARIMA(1,2,1) ₃₀	5710.4770
ARIMA(0,2,0)x SARIMA(1,2,2) ₃₀	5290.1397
ARIMA(0,2,0)x SARIMA(2,0,0) ₃₀	5597.1244
ARIMA(0,2,0)x SARIMA(2,0,1) ₃₀	5599.0582
ARIMA(0,2,0)x SARIMA(2,0,2) ₃₀	5592.1646
ARIMA(0,2,0)x SARIMA(2,1,0) ₃₀	5513.9858
ARIMA(0,2,0)x SARIMA(2,1,1) ₃₀	5413.9772
ARIMA(0,2,0)x SARIMA(2,1,2) ₃₀	5401.7043
ARIMA(0,2,0)x SARIMA(2,2,0) ₃₀	5637.9040
ARIMA(0,2,0)x SARIMA(2,2,1) ₃₀	5363.5491
ARIMA(0,2,0)x SARIMA(2,2,2) ₃₀	5287.2318
ARIMA(0,2,1)x SARIMA(0,0,0) ₃₀	5506.6260
ARIMA(0,2,1)x SARIMA(0,0,1) ₃₀	5260.3364
ARIMA(0,2,1)x SARIMA(0,0,2) ₃₀	5041.0561
ARIMA(0,2,1)x SARIMA(0,1,0) ₃₀	5664.9499
ARIMA(0,2,1)x SARIMA(0,1,1) ₃₀	5094.4096
ARIMA(0,2,1)x SARIMA(0,1,2) ₃₀	4873.4025
ARIMA(0,2,1)x SARIMA(0,2,0) ₃₀	6077.3933
ARIMA(0,2,1)x SARIMA(0,2,1) ₃₀	5267.1226
ARIMA(0,2,1)x SARIMA(0,2,2) ₃₀	4777.0447
ARIMA(0,2,1)x SARIMA(1,0,0) ₃₀	5276.2367
ARIMA(0,2,1)x SARIMA(1,0,1) ₃₀	5262.1740
ARIMA(0,2,1)x SARIMA(1,0,2) ₃₀	5043.0295
ARIMA(0,2,1)x SARIMA(1,1,0) ₃₀	5301.2191
ARIMA(0,2,1)x SARIMA(1,1,1) ₃₀	5096.4096
ARIMA(0,2,1)x SARIMA(1,1,2) ₃₀	4860.9494
ARIMA(0,2,1)x SARIMA(1,2,0) ₃₀	5559.8688
ARIMA(0,2,1)x SARIMA(1,2,1) ₃₀	5162.6708
ARIMA(0,2,1)x SARIMA(1,2,2) ₃₀	4779.0445
ARIMA(0,2,1)x SARIMA(2,0,0) ₃₀	5057.5782
ARIMA(0,2,1)x SARIMA(2,0,1) ₃₀	5059.1711
ARIMA(0,2,1)x SARIMA(2,0,2) ₃₀	5044.3998

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษามาก่อน ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต การค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(0,2,1)x SARIMA(2,1,0) ₃₀	4963.1902
ARIMA(0,2,1)x SARIMA(2,1,1) ₃₀	4889.9627
ARIMA(0,2,1)x SARIMA(2,1,2) ₃₀	4864.8894
ARIMA(0,2,1)x SARIMA(2,2,0) ₃₀	5105.2579
ARIMA(0,2,1)x SARIMA(2,2,1) ₃₀	4842.7222
ARIMA(0,2,1)x SARIMA(2,2,2) ₃₀	4781.0371
ARIMA(0,2,2)x SARIMA(0,0,0) ₃₀	5446.1892
ARIMA(0,2,2)x SARIMA(0,0,1) ₃₀	5201.9075
ARIMA(0,2,2)x SARIMA(0,0,2) ₃₀	4988.6608
ARIMA(0,2,2)x SARIMA(0,1,0) ₃₀	5546.2267
ARIMA(0,2,2)x SARIMA(0,1,1) ₃₀	5004.5602
ARIMA(0,2,2)x SARIMA(0,1,2) ₃₀	4787.2544
ARIMA(0,2,2)x SARIMA(0,2,0) ₃₀	5959.3983
ARIMA(0,2,2)x SARIMA(0,2,1) ₃₀	5158.0852
ARIMA(0,2,2)x SARIMA(0,2,2) ₃₀	4689.9684
ARIMA(0,2,2)x SARIMA(1,0,0) ₃₀	5225.6167
ARIMA(0,2,2)x SARIMA(1,0,1) ₃₀	5203.8623
ARIMA(0,2,2)x SARIMA(1,0,2) ₃₀	4990.1386
ARIMA(0,2,2)x SARIMA(1,1,0) ₃₀	5198.9853
ARIMA(0,2,2)x SARIMA(1,1,1) ₃₀	5006.5619
ARIMA(0,2,2)x SARIMA(1,1,2) ₃₀	4776.7852
ARIMA(0,2,2)x SARIMA(1,2,0) ₃₀	5460.5510
ARIMA(0,2,2)x SARIMA(1,2,1) ₃₀	5053.8987
ARIMA(0,2,2)x SARIMA(1,2,2) ₃₀	4691.1496
ARIMA(0,2,2)x SARIMA(2,0,0) ₃₀	5009.4585
ARIMA(0,2,2)x SARIMA(2,0,1) ₃₀	5010.9618
ARIMA(0,2,2)x SARIMA(2,0,2) ₃₀	4985.4631
ARIMA(0,2,2)x SARIMA(2,1,0) ₃₀	4862.2159
ARIMA(0,2,2)x SARIMA(2,1,1) ₃₀	4812.8735
ARIMA(0,2,2)x SARIMA(2,1,2) ₃₀	4780.1871
ARIMA(0,2,2)x SARIMA(2,2,0) ₃₀	4997.7459
ARIMA(0,2,2)x SARIMA(2,2,1) ₃₀	4745.1582
ARIMA(0,2,2)x SARIMA(2,2,2) ₃₀	4698.1327
ARIMA(1,0,0)x SARIMA(0,0,0) ₃₀	5520.7035
ARIMA(1,0,0)x SARIMA(0,0,1) ₃₀	5274.4255
ARIMA(1,0,0)x SARIMA(0,0,2) ₃₀	5056.0843

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษามาก่อน ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(1,0,0)x SARIMA(0,1,0) ₃₀	5559.5688
ARIMA(1,0,0)x SARIMA(0,1,1) ₃₀	5084.9489
ARIMA(1,0,0)x SARIMA(0,1,2) ₃₀	4864.9740
ARIMA(1,0,0)x SARIMA(0,2,0) ₃₀	5963.0064
ARIMA(1,0,0)x SARIMA(0,2,1) ₃₀	5166.2837
ARIMA(1,0,0)x SARIMA(0,2,2) ₃₀	4716.5660
ARIMA(1,0,0)x SARIMA(1,0,0) ₃₀	5275.2549
ARIMA(1,0,0)x SARIMA(1,0,1) ₃₀	5276.3877
ARIMA(1,0,0)x SARIMA(1,0,2) ₃₀	5058.0692
ARIMA(1,0,0)x SARIMA(1,1,0) ₃₀	5207.0400
ARIMA(1,0,0)x SARIMA(1,1,1) ₃₀	5086.3912
ARIMA(1,0,0)x SARIMA(1,1,2) ₃₀	4855.7851
ARIMA(1,0,0)x SARIMA(1,2,0) ₃₀	5442.1619
ARIMA(1,0,0)x SARIMA(1,2,1) ₃₀	5068.4906
ARIMA(1,0,0)x SARIMA(1,2,2) ₃₀	4717.8184
ARIMA(1,0,0)x SARIMA(2,0,0) ₃₀	5057.0896
ARIMA(1,0,0)x SARIMA(2,0,1) ₃₀	5058.9102
ARIMA(1,0,0)x SARIMA(2,0,2) ₃₀	5062.6075
ARIMA(1,0,0)x SARIMA(2,1,0) ₃₀	4895.7713
ARIMA(1,0,0)x SARIMA(2,1,1) ₃₀	4860.2867
ARIMA(1,0,0)x SARIMA(2,1,2) ₃₀	4854.1522
ARIMA(1,0,0)x SARIMA(2,2,0) ₃₀	4994.8715
ARIMA(1,0,0)x SARIMA(2,2,1) ₃₀	4746.8076
ARIMA(1,0,0)x SARIMA(2,2,2) ₃₀	4717.9029
ARIMA(1,0,1)x SARIMA(0,0,0) ₃₀	5461.0576
ARIMA(1,0,1)x SARIMA(0,0,1) ₃₀	5216.8586
ARIMA(1,0,1)x SARIMA(0,0,2) ₃₀	5004.4537
ARIMA(1,0,1)x SARIMA(0,1,0) ₃₀	5537.9468
ARIMA(1,0,1)x SARIMA(0,1,1) ₃₀	5015.4247
ARIMA(1,0,1)x SARIMA(0,1,2) ₃₀	4799.2930
ARIMA(1,0,1)x SARIMA(0,2,0) ₃₀	5949.4171
ARIMA(1,0,1)x SARIMA(0,2,1) ₃₀	5151.8037
ARIMA(1,0,1)x SARIMA(0,2,2) ₃₀	4688.9389
ARIMA(1,0,1)x SARIMA(1,0,0) ₃₀	5224.4828
ARIMA(1,0,1)x SARIMA(1,0,1) ₃₀	5218.8586
ARIMA(1,0,1)x SARIMA(1,0,2) ₃₀	5004.6810

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น

ARIMA (p, d, q) SARIMA (P, D, Q)s	AIC
ARIMA(1,0,1)x SARIMA(1,1,0) ₃₀	5181.4226
ARIMA(1,0,1)x SARIMA(1,1,1) ₃₀	5017.1597
ARIMA(1,0,1)x SARIMA(1,1,2) ₃₀	4789.6489
ARIMA(1,0,1)x SARIMA(1,2,0) ₃₀	5436.8420
ARIMA(1,0,1)x SARIMA(1,2,1) ₃₀	5048.8086
ARIMA(1,0,1)x SARIMA(1,2,2) ₃₀	4690.3855
ARIMA(1,0,1)x SARIMA(2,0,0) ₃₀	5010.4826
ARIMA(1,0,1)x SARIMA(2,0,1) ₃₀	5011.2071
ARIMA(1,0,1)x SARIMA(2,0,2) ₃₀	4999.2942
ARIMA(1,0,1)x SARIMA(2,1,0) ₃₀	4849.4235
ARIMA(1,0,1)x SARIMA(2,1,1) ₃₀	4808.6066
ARIMA(1,0,1)x SARIMA(2,1,2) ₃₀	4791.4665
ARIMA(1,0,1)x SARIMA(2,2,0) ₃₀	4974.2812
ARIMA(1,0,1)x SARIMA(2,2,1) ₃₀	4723.3270
ARIMA(1,0,1)x SARIMA(2,2,2) ₃₀	4692.3764
ARIMA(1,0,2)x SARIMA(0,0,0) ₃₀	5411.8877
ARIMA(1,0,2)x SARIMA(0,0,1) ₃₀	5153.8156
ARIMA(1,0,2)x SARIMA(0,0,2) ₃₀	4942.6204
ARIMA(1,0,2)x SARIMA(0,1,0) ₃₀	5476.9227
ARIMA(1,0,2)x SARIMA(0,1,1) ₃₀	4948.1752
ARIMA(1,0,2)x SARIMA(0,1,2) ₃₀	4729.9689
ARIMA(1,0,2)x SARIMA(0,2,0) ₃₀	5885.3792
ARIMA(1,0,2)x SARIMA(0,2,1) ₃₀	5092.7766
ARIMA(1,0,2)x SARIMA(0,2,2) ₃₀	4639.3638
ARIMA(1,0,2)x SARIMA(1,0,0) ₃₀	5167.4278
ARIMA(1,0,2)x SARIMA(1,0,1) ₃₀	5155.4791
ARIMA(1,0,2)x SARIMA(1,0,2) ₃₀	4938.9667
ARIMA(1,0,2)x SARIMA(1,1,0) ₃₀	5130.6610
ARIMA(1,0,2)x SARIMA(1,1,1) ₃₀	4947.7071
ARIMA(1,0,2)x SARIMA(1,1,2) ₃₀	4720.2770
ARIMA(1,0,2)x SARIMA(1,2,0) ₃₀	5386.3548
ARIMA(1,0,2)x SARIMA(1,2,1) ₃₀	4994.0917
ARIMA(1,0,2)x SARIMA(1,2,2) ₃₀	4635.0412
ARIMA(1,0,2)x SARIMA(2,0,0) ₃₀	4957.0139
ARIMA(1,0,2)x SARIMA(2,0,1) ₃₀	4953.4932
ARIMA(1,0,2)x SARIMA(2,0,2) ₃₀	4932.3403

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษามาก่อน ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น

ARIMA (p, d, q) SARIMA (P, D, Q) _s	AIC
ARIMA(1,0,2)x SARIMA(2,1,0) ₃₀	4807.7410
ARIMA(1,0,2)x SARIMA(2,1,1) ₃₀	4746.0256
ARIMA(1,0,2)x SARIMA(2,1,2) ₃₀	4722.1675
ARIMA(1,0,2)x SARIMA(2,2,0) ₃₀	4947.3810
ARIMA(1,0,2)x SARIMA(2,2,1) ₃₀	4686.6751
ARIMA(1,0,2)x SARIMA(2,2,2) ₃₀	4634.8266
ARIMA(1,1,0)x SARIMA(0,0,0) ₃₀	5486.5681
ARIMA(1,1,0)x SARIMA(0,0,1) ₃₀	5245.5616
ARIMA(1,1,0)x SARIMA(0,0,2) ₃₀	5029.9997
ARIMA(1,1,0)x SARIMA(0,1,0) ₃₀	5641.5658
ARIMA(1,1,0)x SARIMA(0,1,1) ₃₀	5073.4656
ARIMA(1,1,0)x SARIMA(0,1,2) ₃₀	4855.5622
ARIMA(1,1,0)x SARIMA(0,2,0) ₃₀	6060.6536
ARIMA(1,1,0)x SARIMA(0,2,1) ₃₀	5247.9943
ARIMA(1,1,0)x SARIMA(0,2,2) ₃₀	4756.9055
ARIMA(1,1,0)x SARIMA(1,0,0) ₃₀	5246.3665
ARIMA(1,1,0)x SARIMA(1,0,1) ₃₀	5247.5105
ARIMA(1,1,0)x SARIMA(1,0,2) ₃₀	5031.9813
ARIMA(1,1,0)x SARIMA(1,1,0) ₃₀	5260.7189
ARIMA(1,1,0)x SARIMA(1,1,1) ₃₀	5075.3803
ARIMA(1,1,0)x SARIMA(1,1,2) ₃₀	4844.8064
ARIMA(1,1,0)x SARIMA(1,2,0) ₃₀	5524.2462
ARIMA(1,1,0)x SARIMA(1,2,1) ₃₀	5136.7921
ARIMA(1,1,0)x SARIMA(1,2,2) ₃₀	4759.1715
ARIMA(1,1,0)x SARIMA(2,0,0) ₃₀	5030.8505
ARIMA(1,1,0)x SARIMA(2,0,1) ₃₀	5032.7722
ARIMA(1,1,0)x SARIMA(2,0,2) ₃₀	5032.2633
ARIMA(1,1,0)x SARIMA(2,1,0) ₃₀	4918.8451
ARIMA(1,1,0)x SARIMA(2,1,1) ₃₀	4855.3505
ARIMA(1,1,0)x SARIMA(2,1,2) ₃₀	4844.9702
ARIMA(1,1,0)x SARIMA(2,2,0) ₃₀	5056.1978
ARIMA(1,1,0)x SARIMA(2,2,1) ₃₀	4795.3097
ARIMA(1,1,0)x SARIMA(2,2,2) ₃₀	4785.8746

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

1) โค้ดพยากรณ์วิธีบ็อกซ์และเจนกินส์ (Box and Jenkins)

ตัวแบบSARIMA

```

#เรียกPACKAGE
from dateutil.parser import parse
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import warnings
import itertools
import statsmodels.api as sm
from statsmodels.tsa.stattools import acf, pacf
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
#เรียกข้อมูล
df = pd.read_csv("TPS.csv")
#แบ่งข้อมูล
train = df[:707]
test = df[707:]
train
#พล็อตกราฟข้อมูลทั้งหมด
import seaborn as sns
def plot_df(df, x, y, title="", xlabel='DATE', ylabel='Value', dpi=100,
freq='D'):
    df[x] = pd.to_datetime(df[x])
    df.set_index(x, inplace=True)
    df = df.resample(freq).mean()
    plt.figure(figsize=(16,5), dpi=dpi)
    sns.lineplot(data=df[y])
    plt.gca().set(title=title, xlabel=xlabel, ylabel=ylabel)
    plt.legend(y, loc='upper left')
    plt.show()

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plot_df(df, x='DATE', y=['TPS', 'BASE', 'BASE_50', 'BASE_80'], title='The
Application through API and the Performance of Services', ylabel='TPS,
BASE, BASE_50, BASE_80', freq='D')
```

#พล็อตกราฟข้อมูลTrain

```
def plot_train(df, x, y, title="", xlabel='DATE', ylabel='Value', dpi=100,
freq='D'):
```

```
    df[x] = pd.to_datetime(df[x])
    df.set_index(x, inplace=True)
    df = df.resample(freq).mean()
    plt.figure(figsize=(16,5), dpi=dpi)
    sns.lineplot(data=df[y])
    plt.gca().set(title=title, xlabel=xlabel, ylabel=ylabel)
    plt.legend(y, loc='upper left')
    plt.show()
```

```
plot_train(train, x='DATE', y=['TPS', 'BASE', 'BASE_50', 'BASE_80'],
title='The Application through API and the Performance of Services',
ylabel='TPS, BASE, BASE_50, BASE_80', freq='D')
```

#Plot ACF/PACF

```
plot_acf(train["TPS"],lags=40, alpha=0.05)
plot_pacf(train["TPS"],lags=40, alpha=0.05)
plt.show()
```

#ตรวจสอบความเป็นStationary

```
from statsmodels.tsa.stattools import adfuller #importing adfuller tool
from statsmodels
#statsmodels provide adfuller() function to implement stationarity test
of a time series
adf = adfuller(train["TPS"])
print(adf)
print("\nADF = ',str(adf[0])) #more towards negative value the better
print("\np-value = ',str(adf[1]))
print("\nCritical Values:')
for key, val in adf[4].items(): #for loop to print the p-value (1%, 5%
and 10%) and their respective values
```

```
    print(key,':',val)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

print('Null Hypothesis Rejected.Time Series is Stationary')
else:
    print('Null Hypothesis Accepted.Time Series is not Stationary')
# ทำ differencing ของชุดข้อมูล
train_diff = train["TPS"].diff().dropna()
# ตรวจสอบความเป็นStationary หลังจากทำ differencing
adf = adfuller(train_diff)
print(adf)
print("\nADF = ',str(adf[0])) #more towards negative value the better
print("\np-value = ',str(adf[1]))
print("\nCritical Values:')
for key, val in adf[4].items(): #for loop to print the p-value (1%, 5%
and 10%) and their respective values
print(key,':',val)
    if adf[0] < val:
        print('Null Hypothesis Rejected.Time Series is Stationary')
    else:
        print('Null Hypothesis Accepted.Time Series is not Stationary')
#Plot ACF/PACF
plot_acf(train_diff, lags=40, alpha=0.05)
plot_pacf(train_diff, lags=40, alpha=0.05)
plt.show()
#ตัวแบบที่เหมาะสม
p = d = q = range(0,3)
pdq = list(itertools.product(p,d,q))
seasonal_pdq = [(x[0], x[1], x[2], 30) for x in
list(itertools.product(p,d,q))]
warnings.filterwarnings("ignore")
for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod =
            sm.tsa.statespace.SARIMAX(train.TPS,order=para
m,seasonal_order=param_seasonal,enforce_stati
onarity=False,enforce_invertibility=False)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

results = mod.fit()
print('ARIMA{}*SARIMA{} - AIC:{}'.format(param, param_seasonal,
results.aic))
except:
    continue
#ModelSARIMA
modelsarima = sm.tsa.statespace.SARIMAX(train.TPS,
    order=(2,1,3),
    seasonal_order=(3,1,3,7),
    enforce_stationarity=False,
    enforce_invertibility=False).fit()
print(modelsarima.summary())
print(modelsarima.aic)print(modelsarima.summary())
print(modelsarima.aic)
forecast = modelsarima.forecast(steps=120)
print(forecast)
lforecast = forecast.tolist()
print(type(lforecast))
forecast_dates = pd.date_range(start=train.index[-
1]+pd.Timedelta(days=1),periods=120,freq='D')
lforecast_dates = forecast_dates.tolist()
print(type(lforecast_dates))
forecast1 = pd.Series(lforecast,index=lforecast_dates)
print(type(forecast1))
forecast = pd.DataFrame(forecast1,columns=['VALUES'])
test['forecast']=forecast1
print(test)
forecast.plot()
print(forecast.index)
test = test.set_index(pd.to_datetime(test['DATE']))
print(test.index)
# พล็อตกราฟเปรียบเทียบค่าจริงและค่าที่พยากรณ์ได้
import matplotlib.dates as mdates
fig, ax = plt.subplots(figsize=(16,5))

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

ax.plot(pd.date_range(start=test.DATE.tolist()[0], periods=len(test.BASE),
freq='D'), test.BASE, label='BASE', color='orange')
ax.plot(pd.date_range(start=test.DATE.tolist()[0],
periods=len(test.BASE_50), freq='D'), test.BASE_50, label='BASE_50',
color='green')
ax.plot(pd.date_range(start=test.DATE.tolist()[0],
periods=len(test.BASE_80), freq='D'), test.BASE_80, label='BASE_80',
color='red')
ax.plot(pd.date_range(start=test.DATE.tolist()[0], periods=len(test.TPS),
freq='D'), test.TPS, label='actual')
ax.plot(pd.date_range(start=test.DATE.tolist()[0],
periods=len(forecast.values), freq='D'), forecast.values, label='forecast')
ax.set_title('Comparison of Actual vs Forecasted TPS')
ax.set_xlabel('Date')
ax.set_ylabel('TPS')
plt.legend()
plt.show()
#ทดสอบMAPE
def mape(y_true,y_pred):
    mape = np.mean(abs((y_true-y_pred)/y_true))*100
    return mape
test['forecast1'] = modelsarima.forecast(step=97)
print("MAPE : %.2f" % mape(test[:, 'TPS'],test[:, 'forecast']) + '%')

```

ตัวแบบSARIMAX

```

#เรียกPACKAGE
from dateutil.parser import parse
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import warnings
import itertools
import statsmodels.api as sm

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
#อ่านข้อมูล TPS
df = pd.read_csv("TPS.csv")
#แบ่งข้อมูล
train = df[:707]
test = df[707:]
train
#อ่านข้อมูลปัจจัยภายนอก
days = pd.read_csv('Days.csv')
days['DATE'] = pd.to_datetime(days['DATE'])
days
#รวมข้อมูล
df['DATE'] = pd.to_datetime(df['DATE'])
merged_data = pd.concat([df.set_index('DATE'), days.set_index('DATE')],
axis=1, join='inner').reset_index()
ext_var_list = ['TPS','X1']
exog_data = merged_data[ext_var_list]
exog_data
# แบ่งข้อมูลเป็นชุดฝึกและชุดทดสอบ
train = merged_data.iloc[:707]
test = merged_data.iloc[-90:]
endog = train['TPS']
exog = train['X1']
sarimax = sm.tsa.statespace.SARIMAX(endog=endog,
exog=exog,
order=(2, 1, 3),
seasonal_order=(3, 1, 3, 7),
enforce_stationarity=False,
enforce_invertibility=False).fit()
print(sarimax.summary())
start_index = test.index[0]
end_index = test.index[-1]
# ทำนายค่าในชุดทดสอบ
predictions = sarimax.predict(start=start_index, end=end_index,
exog=test['X1'])

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# แสดงผลลัพธ์การทำนาย
print(predictions)
# สร้างตารางผลลัพธ์เปรียบเทียบค่าทำนายกับค่าจริง
results = pd.DataFrame({'Actual': test['TPS'], 'Predicted': predictions})
# แสดงผลลัพธ์
print(results)
import matplotlib.pyplot as plt
# พล็อตกราฟเปรียบเทียบค่าทำนายกับค่าจริง
plt.figure(figsize=(10, 6))
plt.plot(results.index, results['Actual'], label='Actual')
plt.plot(results.index, results['Predicted'], label='Predicted')
plt.xlabel('Date')
plt.ylabel('TPS')
plt.title('Comparison of Actual vs Forecasted TPS')
plt.legend()
plt.show()
# คำนวณค่า MAPE
mape = np.mean(np.abs((results['Actual'] - results['Predicted']) /
results['Actual'])) * 100
print('MAPE:', mape)
# พล็อตกราฟเปรียบเทียบค่าจริงและค่าที่พยากรณ์ได้
import matplotlib.dates as mdates
fig, ax = plt.subplots(figsize=(16,5))
ax.plot(pd.date_range(start=test.DATE.tolist()[0], periods=len(test.BASE),
freq='D'), test.BASE, label='BASE', color='orange')
ax.plot(pd.date_range(start=test.DATE.tolist()[0],
periods=len(test.BASE_50), freq='D'), test.BASE_50, label='BASE_50',
color='green')
ax.plot(pd.date_range(start=test.DATE.tolist()[0],
periods=len(test.BASE_80), freq='D'), test.BASE_80, label='BASE_80',
color='red')
ax.plot(pd.date_range(start=test.DATE.tolist()[0], periods=len(test.TPS),
freq='D'), test.TPS, label='actual')

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

ax.plot(pd.date_range(start=test.DATE.tolist()[0],
periods=len(predictions.values), freq='D'), predictions.values,
label='forecast')
ax.set_title('Comparison of Actual vs Forecasted TPS')
ax.set_xlabel('Date')
ax.set_ylabel('TPS')
plt.legend()
plt.show()

```

2) โค้ดพยากรณ์ตัวแบบ Random Forest

#เรียก Package

```

from numpy import asarray
from pandas import read_csv
from pandas import DataFrame
from pandas import concat
from sklearn.metrics import mean_absolute_error
from sklearn.ensemble import RandomForestRegressor
from matplotlib import pyplot
import pandas as pd
from math import sqrt
from sklearn.model_selection import train_test_split

```

#แปลงชุดข้อมูลอนุกรมเวลาเป็นชุดข้อมูลการเรียนรู้แบบมีผู้สอน

```

def series_to_supervised(data, n_in=1, n_out=1, dropnan=True, n_vars=1):
    if not isinstance(data, pd.DataFrame):
        data = pd.DataFrame(data)
    n_vars = n_vars if type(n_vars) == list else list(range(n_vars))
    cols, names = list(), list()

    # input sequence (t-n, ..., t-1)
    for i in range(n_in, 0, -1):
        cols.append(data.shift(i))
        names += [f'{data.columns[j]}(t-{i})' for j in n_vars]

    # forecast sequence (t, t+1, ..., t+n)
    for i in range(0, n_out):
        cols.append(data.shift(-i))

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

names += [f'{data.columns[j]}(t)' for j in n_vars]
else:
names += [f'{data.columns[j]}(t+{i})' for j in n_vars]

# put it all together
agg = pd.concat(cols, axis=1)
agg.columns = names

# drop rows with NaN values
if dropnan:
agg.dropna(inplace=True)
return agg

# แบ่งข้อมูล
def train_test_split(data, n_test):
return data[:len(data)-n_test], data[len(data)-n_test:]

# Random Forest Model
def random_forest_forecast(train, testX):
train = asarray(train)
trainX, trainy = train[:, :-1], train[:, -1]
model = RandomForestRegressor(n_estimators=1000)
model.fit(trainX, trainy)
testX = asarray(testX).reshape(1, -1) # แปลงข้อมูล testX เป็นรูปแบบที่
yhat = model.predict(testX)
return yhat[0]

# Import numpy module to use 'concatenate'
from numpy import concatenate
def walk_forward_validation(data, n_test, n_output):
predictions = list()

# split dataset
train, test = train_test_split(data, n_test)

# extract the number of input columns
n_input = train.shape[1] - 1

# define the feature columns
cols = [i for i in range(n_input)]

# define the target column

```

ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

cols.append(train.shape[1] - 1)
# walk forward over time steps in test
for i in range(len(test)):
# split test row into input and output columns
testX, testy = test.iloc[i, :n_input].values, test.iloc[i, n_input:].values
# fit model on history and make a prediction
yhat = random_forest_forecast(train, testX)

# store forecast in list of predictions
predictions.append(yhat)
# add actual observation to history for the next loop
train = pd.concat([train, test.iloc[i:i+n_output]])
# summarize progress
print('>expected=%.1f, predicted=%.1f' % (testy, yhat))
# estimate prediction error
error = mean_absolute_error(test.iloc[:, n_input:].values, predictions)
print("mae", error)
return test, predictions
# load dataset
series = read_csv('TPS.csv')
values = series.values
print (series)
df = pd.DataFrame(values,columns =
[DATE','TPS','BASE','BASE_50','BASE_80'])
# transform data to supervised learning
data = series_to_supervised(df['TPS'], n_in=1)
# evaluate
y, yhat = walk_forward_validation(data, 60, 90)
print("y",y)
print("yhat",yhat)
import matplotlib.dates as mdates
import matplotlib.pyplot as plt
fig, ax = plt.subplots(figsize=(16,5))
ax.plot(pd.date_range(start=series.DATE.tolist()[-1], periods=len(y["TPS(t)"]),
freq='D'), y["TPS(t)"], label='TPS')

```

เอกสารนี้เป็นเอกสารสงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอญาตรหน้าไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

ax.plot(pd.date_range(start=series.DATE.tolist()[-1], periods=len(yhat), freq='D'),
yhat, label='Predict')
ax.plot(pd.date_range(start=series.DATE.tolist()[-1], periods=len(yhat), freq='D'),
series["BASE"][0]*len(yhat), label='BASE', color='yellow')
[
ax.plot(pd.date_range(start=series.DATE.tolist()[-1], periods=len(yhat), freq='D'),
[series["BASE_50"][0]*len(yhat), label='BASE_50', color='green')
ax.plot(pd.date_range(start=series.DATE.tolist()[-1], periods=len(yhat), freq='D'),
[series["BASE_80"][0]*len(yhat), label='BASE_80', color='red')
ax.set_title('Comparison of Actual vs Forecasted TPS')
ax.set_xlabel('Date')
ax.set_ylabel('TPS')
plt.legend()
plt.show()
#หาค่า MAPE
import numpy as np
def mean_absolute_percentage_error(y_true, y_pred):
y_true, y_pred = np.array(y_true), np.array(y_pred)
return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
actual = y["TPS(t)"].values
predicted = yhat
mape = mean_absolute_percentage_error(actual, predicted)
print("MAPE:", mape)

```

3) โค้ดพยากรณ์ตัวแบบ LSTM

```

#เรียกPackage
!pip install keras
!pip install tensorflow
from pandas import DataFrame
from pandas import Series
from pandas import concat
from pandas import read_csv
from sklearn.metrics import mean_squared_error
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

from keras.layers import LSTM
from matplotlib import pyplot
import numpy
from sklearn.model_selection import train_test_split
# ฟังก์ชันสำหรับแปลง Time series เป็น supervised learning
def timeseries_to_supervised(data, lag=1):
    df = DataFrame(data)
    columns = [df.shift(i) for i in range(1, lag+1)]
    columns.append(df)
    df = concat(columns, axis=1)
    df.fillna(0, inplace=True)
    return df
# ฟังก์ชันสำหรับทำให้ข้อมูลเป็น stationary
def difference(dataset, interval=1):
    diff = list()
    for i in range(interval, len(dataset)):
        value = dataset[i] - dataset[i-interval]
        diff.append(value)
    return Series(diff)
# ฟังก์ชันสำหรับกลับกระบวนการทำให้ข้อมูล stationary
def inverse_difference(history, yhat, interval=1):
    return yhat + history[-interval]

# ฟังก์ชันสำหรับปรับ scale ข้อมูล
def scale(train, test):
    # fit scaler
    scaler = MinMaxScaler(feature_range = (-1,1))
    scaler = scaler.fit(train)

    # transform train
    train = train.reshape(train.shape[0],train.shape[1])
    train_scaled = scaler.transform(train)

    # transform test
    test = test.reshape(test.shape[0],test.shape[1])
    test_scaled = scaler.transform(test)
    return scaler, train_scaled, test_scaled

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลับสู่มาตราส่วนเดิม เพื่อให้ตีความผลลัพธ์ได้

```
def invert_scale(scaler, x, value):
    new_row = [x for x in x] + [value]
    array = numpy.array(new_row)
    array = array.reshape(1, len(array))
    inverted = scaler.inverse_transform(array)
    return inverted[0, -1]

def fit_lstm(train, batch_size, nb_epoch, neurons):
    x, y = train[:, 0:-1], train[:, -1]
    x = x.reshape(x.shape[0], 1, x.shape[1])
    model = Sequential()
    model.add(LSTM(neurons, batch_input_shape=(batch_size, x.shape[1],
    x.shape[2]), stateful=True))
    model.add(Dense(1))
    model.compile(loss='mean_squared_error', optimizer='adam')
    for i in range(nb_epoch):
        model.fit(x, y, epochs=1, batch_size=batch_size, verbose=0, shuffle=False)
    model.reset_states()
    return model
```

make a one-step forecast

```
def forecast_lstm(model, batch_size, x):
    x = x.reshape(1,1,len(x))
    yhat = model.predict(x,batch_size = batch_size)
    return yhat[0,0]
```

โหลดข้อมูลเข้ามา

```
series = read_csv('TPS.csv', header=0, index_col=0)
raw_values = series.values
```

แปลง time series เป็น supervised learning

```
supervised = timeseries_to_supervised(raw_values, 1)
supervised_values = supervised.values
```

ทำให้ข้อมูล stationary

```
diff_values = difference(supervised_values, 1)
```

แบ่งข้อมูล train/test

```
train, test = supervised_values[0:-90],supervised_values[-90:]
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยฺาตให้นำไปใช้ประโยชน์ด้านการค้า

#transform the scale of the data

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

scaler,train_scaled,test_scaled = scale(train,test)
#ปรับ scaler ด้วย train set
scaler = MinMaxScaler(feature_range=(-1, 1))
scaler = scaler.fit(train)
#ทำให้ข้อมูล train/test เป็นสเกลแล้ว
train_scaled = scaler.transform(train)
test_scaled = scaler.transform(test)
import numpy as np

# fit the model
lstm_model = fit_lstm(train_scaled, 1, 30, 900)
# make a one-step forecast for each training data point to build up state for
forecasting
train_reshaped = train_scaled[:, 0].reshape(len(train_scaled), 1, 1)
train_reshaped = np.repeat(train_reshaped, 7, axis=2) # repeat feature values
lstm_model.predict(train_reshaped, batch_size=1)
# forecast the test dataset
predictions = []
for i in range(len(test_scaled)):
# make one-step forecast
x, y = test_scaled[i, 0:-1], test_scaled[i, -1]
yhat = forecast_lstm(lstm_model, 1, x)
# invert scaling
yhat = invert_scale(scaler, x, yhat)
predictions.append(yhat)
# invert differencing
def inverse_difference(history, yhat, interval=1):
    return yhat + history[-interval]
yhat = inverse_difference(raw_values, yhat, len(test_scaled)+1-i)
# store forecast
predictions.append(yhat[0])
expected = raw_values[len(train)+i+1]
print('Day=%d, Predicted=%f, Expected=%f' %(i+1, yhat[0], expected))
print(predictions)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 y = series[-90:]
 ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

y
import matplotlib.dates as mdates
import matplotlib.pyplot as plt
import pandas as pd
fig, ax = plt.subplots(figsize=(16,5))
ax.plot(pd.date_range(start='2023-03-01', periods=len(y.TPS), freq='D'), y.TPS,
label='TPS')
ax.plot(pd.date_range(start='2023-03-01', periods=len(predictions), freq='D'),
predictions, label='predictions')
ax.plot(pd.date_range(start='2023-03-01', periods=len(predictions), freq='D'),
[y.BASE[0]] * len(predictions), label='BASE', color='yellow')
ax.plot(pd.date_range(start='2023-03-01', periods=len(predictions), freq='D'),
[y.BASE_50[0]] * len(predictions), label='BASE_50', color='green')
ax.plot(pd.date_range(start='2023-03-01', periods=len(predictions), freq='D'),
[y.BASE_80[0]] * len(predictions), label='BASE_80', color='red')
ax.set_title('Comparison of Actual vs Forecasted TPS')
ax.set_xlabel('Date')
ax.set_ylabel('TPS')
plt.legend()
plt.show()

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้วิจัย



ลลิตภัทร จินะพรม (น้ำ)

ที่อยู่ : 186/24 หมู่ที่ 1 ซอยก้งสดาล 2 ตำบลองครักษ์ อำเภองครักษ์ จังหวัดนครนายก 26120

เบอร์โทรศัพท์ : 0958291722

E-mail : numlalit019@gmail.com

ประวัติการศึกษา

- ระดับมัธยมศึกษา
โรงเรียนเบญจมราชรังสฤษฎ์ ๓ ฉะเชิงเทรา
- ระดับปริญญาตรี
วิทยาศาสตร์ สาขาสถิติประยุกต์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



งานทะเบียนคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
คำรับรองเล่มสหกิจศึกษา

วันที่ 9 เดือน กรกฎาคม พ.ศ. 2566

ข้าพเจ้า นางสาวลลิตภัทร จินะพรม รหัสประจำตัว 62050825

นักศึกษาหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชา สถิติประยุกต์ ภาควิชา สถิติ ขอรับรองว่าโครงการ
สหกิจศึกษา เรื่อง

ชื่อภาษาไทย การพยากรณ์จำนวนการเรียกใช้เซิร์ฟเวอร์ของแอปพลิเคชันและการทดสอบ
ประสิทธิภาพของเซิร์ฟเวอร์ เพื่อเป็นแนวทางในการวางแผนค่าใช้จ่าย
ทรัพยากรของ เซิร์ฟเวอร์ได้อย่างมีประสิทธิภาพ

ชื่อภาษาอังกฤษ Forecasting Calling the Applications through API and the
Performance of Services to Guide Server Cost
Management

ปีการศึกษา 2565

เป็นผลงานวิจัยที่มีได้คัดลอกหรือละเมิดลิขสิทธิ์ของผู้อื่นและได้ผ่านการตรวจสอบความซ้ำซ้อน
เรียบร้อยแล้ว และได้แนบเอกสารการตรวจสอบการลอกเลียนวรรณกรรมที่ตรวจสอบจากเล่ม
โครงการสหกิจศึกษาฉบับสมบูรณ์แล้ว

โปรแกรมอักขรวิสุทธิ 9.84%

ลงชื่อ.....**ลลิตภัทร จินะพรม**.....

(นางสาวลลิตภัทร จินะพรม)

นักศึกษา

ข้าพเจ้า ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์ อาจารย์ที่ปรึกษาโครงการสหกิจศึกษา ได้ตรวจสอบ โครงการสห
กิจศึกษาของนักศึกษาข้างต้นแล้ว ขอรับรองว่าเป็นผลงานวิจัยของนักศึกษาและมีเนื้อหาสมบูรณ์ จึง
ลงชื่อไว้เป็นหลักฐาน

ลงชื่อ.....**พรพิมล ชัยวุฒิศักดิ์**.....

อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้