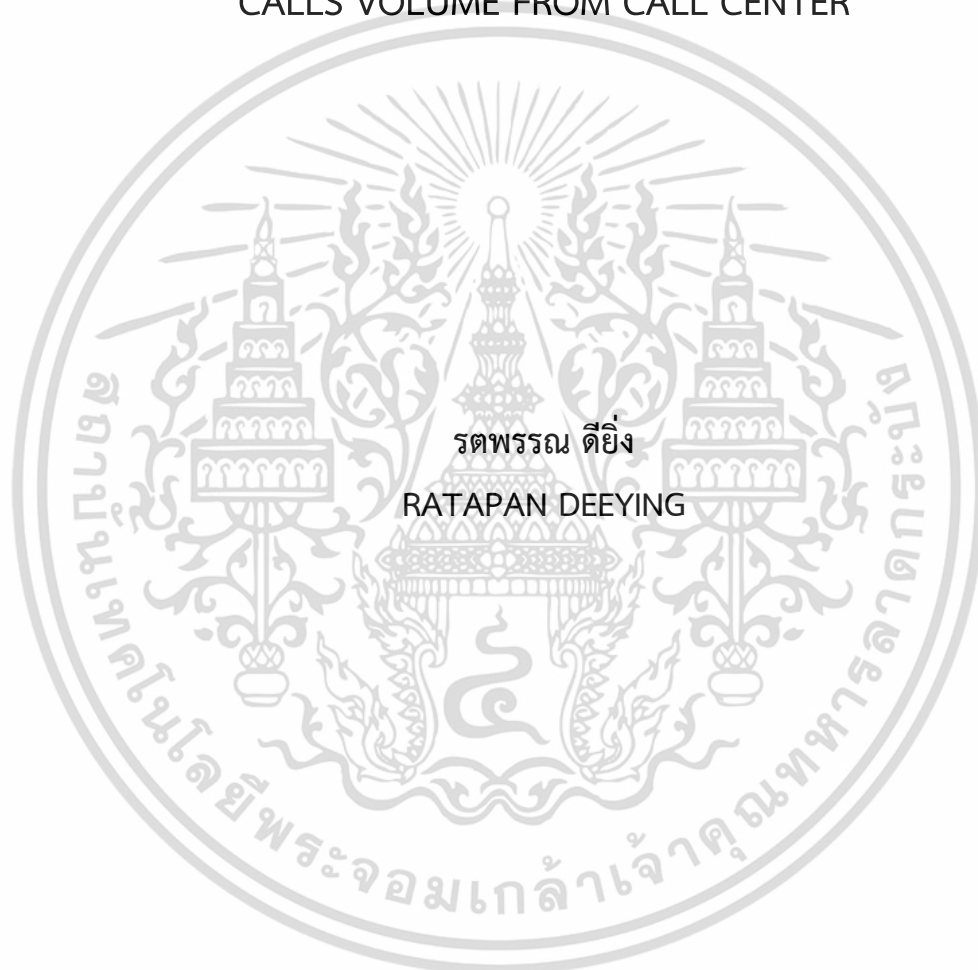


การเปรียบเทียบประสิทธิภาพการพยากรณ์ระหว่างวิธีบอกซ์-เจนกินส์  
วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม: กรณีศึกษาปริมาณสายโทรเข้า  
ศูนย์บริการข้อมูลลูกค้าของบริษัทแห่งหนึ่ง

PERFORMANCE COMPARISON BETWEEN BOX-JENKINS METHOD,  
RANDOM FOREST, AND NEURAL NETWORK: A CASE STUDY OF THE  
CALLS VOLUME FROM CALL CENTER



การค้นคว้าอิสระเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาสถิติและการวิเคราะห์ธุรกิจ  
ภาควิชาสถิติ คณะวิทยาศาสตร์  
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2565

KMITL-2022-SC-M-050-094

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

PERFORMANCE COMPARISON BETWEEN BOX-JENKINS METHOD,  
RANDOM FOREST, AND NEURAL NETWORK: A CASE STUDY OF THE  
CALLS VOLUME FROM CALL CENTER



AN INDEPENDENT RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR  
THE DEGREE OF MASTER OF SCIENCE PROGRAM IN STATISTICS AND BUSINESS  
ANALYTICS DEPARTMENT OF STATISTICS SCHOOL OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2022

KMITL-2022-SC-M-050-094

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การเปรียบเทียบประสิทธิภาพการพยากรณ์ระหว่างวิธีบอกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม: กรณีศึกษาปริมาณสายโทรเข้า ศูนย์บริการข้อมูลลูกค้าของบริษัทแห่งหนึ่ง
ชื่อนักศึกษา	รตพรรณ ตี๋ยิ่ง
รหัสประจำตัว	63605117
ปริญญา	วิทยาศาสตร์มหาบัณฑิต (สถิติและการวิเคราะห์ธุรกิจ)
ภาควิชา	สถิติ
พ.ศ.	2565
อาจารย์ที่ปรึกษา	ดร.ยุวดี กล่อมวิเศษ

### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ระหว่างวิธีบอกซ์-เจนกินส์ วิธีป่าสุ่มและวิธีโครงข่ายประสาทเทียม โดยใช้ข้อมูลปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าของบริษัทแห่งหนึ่ง ตั้งแต่วันที่ 1 เดือนกันยายน พ.ศ. 2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ. 2563 รวมทั้งสิ้น 91 วัน โดยทำการแบ่งข้อมูลเป็นชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบในอัตราส่วน 70 และ 30 ตามลำดับ สำหรับการเปรียบเทียบประสิทธิภาพของวิธีบอกซ์-เจนกินส์ วิธีป่าสุ่มและวิธีโครงข่ายประสาทเทียม จะเปรียบเทียบโดยวัดจากเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE)

จากการเปรียบเทียบประสิทธิภาพในการพยากรณ์ทั้ง 3 วิธีพบว่า วิธีบอกซ์-เจนกินส์ ได้ค่า MAPE เท่ากับ 3.681% วิธีป่าสุ่ม ได้ค่า MAPE เท่ากับ 4.653% และวิธีโครงข่ายประสาทเทียม ได้ค่า MAPE เท่ากับ 2.604% วิธีโครงข่ายประสาทเทียมจึงเป็นวิธีที่มีความเหมาะสมในการพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าเนื่องจากมีเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำที่สุด

**คำสำคัญ:** การพยากรณ์ การเปรียบเทียบประสิทธิภาพ โครงข่ายประสาทเทียม บอกซ์-เจนกินส์ ปริมาณสายโทรเข้า ป่าสุ่ม

**Independent Study Title** PERFORMANCE COMPARISON BETWEEN BOX-JENKINS METHOD, RANDOM FOREST, AND NEURAL NETWORK: A CASE STUDY OF THE CALLS VOLUME FROM CALL CENTER

**Student Name** Miss Ratapan Deeying

**Student ID** 63605117

**Degree** Master of Science (Statistics and Business Analytics)

**Department** Statistics

**Year** 2022

**Independent Study Advisor** Dr.Yuwadee Klomwises

### Abstract

The purpose of this research is to compare efficiency of forecasting daily incoming calls by three methods including Box-Jenkins, random forest, and neural network. The data for this study was daily incoming calls at a call center from September 1, 2020 to November 30, 2020, which consist of 91 days. Data set are divided into training data and testing data in a 70 : 30 ratio. In addition, the efficiencies of Box-Jenkins, random forest and neural network are compared with the mean absolute percentage error (MAPE).

By comparing the performance of forecasting using three methods, the results can be summarized as follows. Box-Jenkins, random forest, and neural network have MAPE values as 3.681%, 4.653%, and 2.604% respectively. In conclusion, that neural network can be considered as the most suitable model for forecasting daily incoming calls at a call center related to the lowest value of mean absolute error percentage.

**Keywords:** Forecasting, Calls Volume, Performance Comparison, Neural Network, Box-Jenkins, Random Forest



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

การค้นคว้าอิสระในครั้งนี้ สำเร็จลุล่วงได้ด้วยความกรุณาจาก ดร.ยุวดี กล่อมวิเศษ อาจารย์ที่ปรึกษาการค้นคว้าอิสระซึ่งได้ให้ความรู้ การชี้แนะแนวทางการศึกษา ตรวจสอบและแก้ไขข้อบกพร่องในงาน ตลอดจนการให้คำปรึกษาซึ่งเป็นประโยชน์ในการวิจัยจนการค้นคว้าอิสระครั้งนี้มีความสมบูรณ์ครบถ้วนสำเร็จไปได้ด้วยดี

ขอขอบพระคุณ รองศาสตราจารย์ ดร.อัชฌา อระวีพร และดร.บุญญสิทธิ วรจันทร์ กรรมการการค้นคว้าอิสระที่ท่านได้ให้คำแนะนำและคำปรึกษาเพื่อแก้ไขข้อบกพร่องของการค้นคว้าอิสระเล่มนี้ เพื่อให้การค้นคว้าอิสระเล่มนี้สมบูรณ์ยิ่งขึ้น

รวมถึงคณะอาจารย์ประจำภาควิชาสถิติทุกท่านที่ได้ถ่ายทอดวิชาความรู้พร้อมทั้งคำแนะนำทำให้สามารถนำความรู้ทางวิชาการต่างๆ มาประยุกต์ใช้ในการศึกษาวิจัยครั้งนี้ ผู้วิจัยจึงขอขอบพระคุณเป็นอย่างสูง มาไว้ ณ โอกาสนี้

รตพรรณ ดิยिंग

# สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ง
สารบัญ	จ
สารบัญตาราง	ช
สารบัญรูป	ฅ
<b>บทที่ 1 บทนำ</b>	<b>1</b>
1.1 ความสำคัญและที่มาของการศึกษา	1
1.2 วัตถุประสงค์ของการวิจัย	3
1.3 ขอบเขตการศึกษา	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
1.5 คำนิยามศัพท์เฉพาะ	3
<b>บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง</b>	<b>4</b>
2.1 ทฤษฎีเกี่ยวกับการพยากรณ์	4
2.1.1 ความหมายของการพยากรณ์	4
2.1.2 ขั้นตอนในการพยากรณ์	4
2.2 ทฤษฎีเกี่ยวกับอนุกรมเวลา	5
2.2.1 ความหมายของอนุกรมเวลา	5
2.2.2 ส่วนประกอบของอนุกรมเวลา	6
2.3 วิธี Box และ Jenkins	7
2.3.1 ขั้นตอนการพยากรณ์โดยวิธีของ Box และ Jenkins	9
2.3.2 รูปแบบ ARIMA	10
2.3.3 รูปแบบ ARIMAX	10
2.3.4 สหสัมพันธ์ Autocorrelation Function	11
2.3.5 ตัวแบบของอนุกรมเวลา	11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ (ต่อ)

	หน้า
2.3.6 การทดสอบความนิ่งของข้อมูล (Unit Root Test)	12
2.5 วิธีปาสุ่ม	13
2.6 วิธีโครงข่ายประสาทเทียม	14
2.7 การวัดประสิทธิภาพของการพยากรณ์	16
2.8 ศูนย์บริการข้อมูลลูกค้า	16
2.9 งานวิจัยที่เกี่ยวข้อง	18
<b>บทที่ 3 วิธีดำเนินงานวิจัย</b>	<b>22</b>
3.1 ขั้นตอนการดำเนินงานวิจัย	22
3.2 ข้อมูลที่ใช้ในงานวิจัย	22
3.3 สถิติที่ใช้ในการวิเคราะห์ข้อมูล	23
3.3.1 สถิติสำหรับวิเคราะห์	23
3.3.2 สถิติสำหรับวัดประสิทธิภาพของตัวแบบพยากรณ์	26
3.4 ขั้นตอนการทำ Dashboard	27
<b>บทที่ 4 ผลการวิจัย</b>	<b>28</b>
4.1 วิธีบอกซ์-เจนกินส์	28
4.2 วิธีปาสุ่ม	36
4.3 วิธีโครงข่ายประสาทเทียม	39
4.4 การเปรียบเทียบค่าจริงกับค่าพยากรณ์แบบรายวัน	42
4.5 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์	43
4.6 การเปรียบเทียบประสิทธิภาพของการพยากรณ์กับชุดข้อมูล	43
<b>บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ</b>	<b>45</b>
5.1 สรุปผลการวิจัย	45
5.2 อภิปรายผล	45
5.3 ข้อเสนอแนะสำหรับการศึกษาคั้งต่อไป	45
5.4 Dashboard	46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ (ต่อ)

เอกสารอ้างอิง

หน้า

49

ภาคผนวก ก

52



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

ตารางที่	หน้า
3.1 กำหนดค่าไฮเปอร์พารามิเตอร์สำหรับตัวแบบป่าสุ่ม	25
3.2 กำหนดค่าไฮเปอร์พารามิเตอร์สำหรับตัวแบบโครงข่ายประสาทเทียม	26
4.1 เปรียบเทียบประสิทธิภาพตัวแบบบ็อกซ์-เจนกินส์	29
4.2 ผลการทดสอบค่าพารามิเตอร์ของตัวแบบ ARIMAX(0,0,0)	30
4.3 ตรวจสอบตัวแบบ ARIMAX(0,0,0)	33
4.4 ตารางแสดงค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีบ็อกซ์และเจนกินส์	35
4.5 เปรียบเทียบประสิทธิภาพตัวแบบป่าสุ่ม	36
4.6 ตารางแสดงค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีป่าสุ่ม	38
4.7 เปรียบเทียบประสิทธิภาพตัวแบบโครงข่ายประสาทเทียม	39
4.8 ตารางแสดงค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีโครงข่ายประสาทเทียม	41
4.9 ตารางแสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลกับค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลด้วยวิธีบ็อกซ์และเจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม	42
4.10 เปรียบเทียบความแม่นยำระหว่างตัวแบบพยากรณ์	43
4.11 การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายโดยใช้ชุดข้อมูลย้อนหลัง (Lagged)	44

## สารบัญรูปร่างภาพ

รูปที่	หน้า
2.1 ขั้นตอนกระบวนการพยากรณ์	5
2.2 แบบจำลองป่าสุ่ม	14
2.3 แบบจำลองโครงข่ายประสาทเทียม	15
2.4 โครงสร้างเบื้องต้น PBX Call Center	17
4.1 การเคลื่อนไหวของอนุกรมเวลา	28
4.2 คอเรลโรแกรมของ ACF และ PACF ของปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า	29
4.3 กราฟแสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีบอกซ์-เจนกินส์	34
4.4 กราฟแสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีป่าสุ่ม	37
4.5 กราฟแสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีโครงข่ายประสาทเทียม	40
5.1 Dashboard แสดงภาพรวมของข้อมูลและการพยากรณ์	46
5.2 Dashboard แสดงปริมาณสายโทรเข้าและระยะเวลา	46
5.3 Dashboard แสดงปริมาณสายโทรเข้าแบ่งตามวันทำงาน	47
5.4 Dashboard แสดงปริมาณสายโทรเข้า	47
5.5 Dashboard แสดงปริมาณสายโทรเข้าและค่าพยากรณ์ด้วยวิธีบอกซ์-เจนกินส์	47
5.6 Dashboard แสดงปริมาณสายโทรเข้าและค่าพยากรณ์ด้วยวิธีป่าสุ่ม	48
5.7 Dashboard แสดงปริมาณสายโทรเข้าและค่าพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียม	48

# บทที่ 1

## บทนำ

### 1.1 ความสำคัญและที่มาของการศึกษา

ปัจจุบันการติดต่อสื่อสารโดยใช้โทรศัพท์เป็นการสื่อสารที่ประหยัด สะดวก และรวดเร็วที่สุด ในการทำธุรกรรมด้านการบริการหรือสอบถามข้อมูล และเป็นส่วนสำคัญสำหรับการทำธุรกิจ ซึ่งแต่ละประเทศทั่วโลกต่างก็มีองค์กรต่างๆ ที่สนใจลงทุนทำธุรกิจด้านการสื่อสารมากมาย เพื่อตอบสนองความต้องการของลูกค้า ดังนั้นศูนย์บริการข้อมูลลูกค้านับเป็นการบริการทางหนึ่ง ซึ่งองค์กรที่ทำธุรกิจด้านการสื่อสาร จัดตั้งขึ้นเพื่อเพิ่มประสิทธิภาพในการให้บริการลูกค้า สร้างความประทับใจ และทำให้ลูกค้าเกิดความภักดี ตลอดจนบอกต่อผู้อื่นเพื่อมาใช้บริการ

ปัญหาที่พบได้โดยทั่วไปของงานศูนย์บริการข้อมูลลูกค้า คือ ปริมาณสายที่เข้ามาตลอดเวลาและไม่สามารถคาดการณ์ได้ การที่องค์กรยังไม่เคยมีการพยากรณ์ปริมาณสายมาก่อน จึงทำให้เกิดภาระงานขึ้น และจำนวนของพนักงานผู้รับสายนั้นอาจจะไม่เพียงพอต่อปริมาณสายที่โทรเข้ามา ในงานศูนย์บริการข้อมูลลูกค้าที่เน้นการบริการด้วยความสุภาพ แต่ต้องสนทนาด้วยความกระชับ เนื่องจากมีลูกค้าที่รอรับบริการเป็นจำนวนมาก ดังนั้นการที่จำนวนพนักงานรับสายไม่เพียงพอปริมาณสายที่เข้ามาอาจนำไปสู่ความสูญเสียรายได้เนื่องจากอาจมีลูกค้าไม่สามารถรอสายเพื่อสนทนากับพนักงานรับบริการได้ การพยากรณ์ปริมาณสายโทรเข้านั้นควรนำมาเป็นกลยุทธ์ในการวางแผนจัดตารางงานของพนักงานให้เพียงพอกับปริมาณในทุกช่วงเวลา เพื่อลดโอกาสที่ลูกค้ารอสายเพื่อรับบริการนานเกินไปและเพื่อลดค่าใช้จ่ายขององค์กรในการจ้างพนักงานรับสายล่วงเวลา การกำหนดให้มีจำนวนพนักงานรับสายสอดคล้องกับความต้องการของลูกค้าในอนาคต จะเป็นการเพิ่มประสิทธิภาพในการทำงานของพนักงานรับสาย และสามารถช่วยในการวางแผนเพื่อจัดตารางงานสำหรับพนักงานรับสายซึ่งมีจำนวนพนักงานรับสายที่เหมาะสมตามปริมาณสายที่เข้ามาคือสิ่งสำคัญในการบริการลูกค้าในงานศูนย์บริการข้อมูลลูกค้า

ดังนั้น ศูนย์บริการข้อมูลลูกค้าจึงมีบทบาทที่สำคัญต่อการให้บริการ เนื่องจากเป็นช่องทางในการเพิ่มประสิทธิภาพในการให้บริการแก่ลูกค้าการบริการข่าวสาร ติดต่อสอบถามและให้ความช่วยเหลือด้านต่างๆ ให้กับลูกค้าโดยตรงส่งผลต่อภาพลักษณ์และความเชื่อมั่นในการรับบริการ การที่ให้บริการลูกค้าที่มีประสิทธิภาพที่ดีนั้น พนักงานต้องให้บริการที่สุภาพ รวดเร็ว สามารถตอบคำถามได้อย่างถูกต้อง ครบถ้วนและรวดเร็วภายในระยะเวลาที่กระชับเหมาะสมและที่สำคัญคือต้องสามารถรับสายลูกค้า

ได้ทันที เวลาที่ลูกค้ามารับบริการยิ่งน้อยยิ่งดี ดังนั้นทางองค์กรจึงต้องมีการพยากรณ์ปริมาณสายโทรเข้าที่ถูกต้องและแม่นยำ เพื่อจะวางแผนในการจัดพนักงานให้เพียงพอต่อปริมาณสายได้อย่างเหมาะสม

จากการศึกษางานวิจัยตัวอย่างที่เกี่ยวกับการพยากรณ์ที่ผ่านมาพบว่า ศิริเทพ จันทร์บุญแก้ว (2560) ได้ศึกษาการพยากรณ์ปริมาณสายโทรเข้าของธนาคารพาณิชย์สำหรับการพยากรณ์รายเดือนทำการเปรียบเทียบระหว่างวิธีเฉลี่ยเคลื่อนที่อย่างง่าย (Simple Moving Average) วิธีปรับให้เรียบเอ็กโปเนนเชียล (Exponential Smoothing Methods) วิธีการของบอซ-เจนกินส์ (ARIMA) และวิธีปัจจุบันของธนาคาร ผลการวิจัยพบว่าการพยากรณ์ด้วยวิธีการพยากรณ์แบบบอซ-เจนกินส์ให้ค่าพยากรณ์ที่ดีที่สุด และจากการศึกษางานวิจัยของ Albrecht, T. et al. (2021) ได้ศึกษาการพยากรณ์สายโทรเข้าของคอลเซ็นเตอร์ โดยเปรียบเทียบความแม่นยำในการพยากรณ์ระหว่างวิธี ARIMA, Random Forest, Gradient boosting, K-nearest neighbor และ Support Vector Regression ใช้ข้อมูลคอลเซ็นเตอร์ของผู้ค้าปลีกออนไลน์ชั้นนำของประเทศเยอรมัน ซึ่งเก็บรวบรวมข้อมูลจำแนกเป็นรายครึ่งชั่วโมง ตั้งแต่วันที่ 2 เดือนมกราคม ค.ศ. 2016 ถึงวันที่ 7 เดือนพฤษภาคม ค.ศ. 2019 ผลการศึกษาพบว่า วิธีที่มีความเหมาะสม คือวิธีการพยากรณ์โดย Random Forest

นอกจากวิธีแบบดั้งเดิม เช่น การพยากรณ์แบบ Autoregressive Integrated Moving Average (ARIMA) ที่เป็นตัวแบบการพยากรณ์จากข้อมูลในอดีตหรือที่เรียกว่า อนุกรมเวลา ในปัจจุบันตัวแบบการพยากรณ์โดยอาศัยการเรียนรู้ของเครื่อง เป็นอีกตัวแบบหนึ่งที่ทำให้ความแม่นยำสูง และมีความหลากหลายของการสร้างตัวแบบเพื่อให้การพยากรณ์มีความแม่นยำมากขึ้น การพยากรณ์โดยอาศัยการเรียนรู้ของเครื่องจะทำการเรียนรู้และสร้างตัวแบบในการพยากรณ์จากข้อมูลที่นำเข้าและผลลัพธ์ที่เกิดขึ้นจากข้อมูลนั้น และนำตัวแบบในการพยากรณ์ที่สร้างขึ้นจากการเรียนรู้มาทำการพยากรณ์เมื่อมีการนำเข้าข้อมูลใหม่

เพื่อค้นหาตัวแบบการพยากรณ์ที่ให้ผลการพยากรณ์ที่แม่นยำที่สุดในการพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า ผู้วิจัยจึงจัดทำงานวิจัยนี้เพื่อศึกษาการเปรียบเทียบการพยากรณ์ระหว่างวิธีบอซ-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม กรณีศึกษาปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าของบริษัทแห่งหนึ่ง

## 1.2 วัตถุประสงค์ของการวิจัย

- 1) เพื่อศึกษาตัวแบบการพยากรณ์รายวันของปริมาณสายโทรเข้าโดยใช้วิธีบอกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม
- 2) เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ระหว่างวิธีบอกซ์-เจนกินส์ วิธีป่าสุ่มและวิธีโครงข่ายประสาทเทียม

## 1.3 ขอบเขตการศึกษา

- 1) ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลปริมาณสายโทรเข้ารายวันของบริษัทแห่งหนึ่ง ตั้งแต่วันที่ 1 เดือนกันยายน พ.ศ. 2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563
- 2) เปรียบเทียบประสิทธิภาพเทคนิคการพยากรณ์ที่ใช้ศึกษา ได้แก่ วิธีบอกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม โดยวัดจากเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE)

## 1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ได้วิธีที่เหมาะสมในการพยากรณ์ปริมาณสายโทรเข้า
- 2) ช่วยให้การวางแผนจัดพนักงานรับสายให้มีประสิทธิภาพมากขึ้น ทั้งยังสามารถลดความน่าจะเป็นที่จำนวนพนักงานจะไม่เพียงพอต่อการให้บริการลูกค้า หรือจำนวนพนักงานมากเกินไป ซึ่งช่วยลดต้นทุนในการจ้างพนักงาน

## 1.5 คำนิยามศัพท์เฉพาะ

ศูนย์บริการข้อมูลลูกค้า (Call Center) หมายถึง หน่วยงานที่มีเจ้าหน้าที่คอยให้บริการข้อมูลข่าวสารรวมถึงรับทำรายการทางธุรกิจต่างๆ โดยมีช่องทางการให้บริการทางโทรศัพท์

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยเรื่อง การเปรียบเทียบประสิทธิภาพการพยากรณ์ระหว่างวิธีบอกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม : กรณีศึกษาปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าบริษัทแห่งหนึ่ง ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี ความหมาย และเอกสารที่เกี่ยวข้อง ดังนี้

- 2.1 ทฤษฎีเกี่ยวกับการพยากรณ์
- 2.2 ทฤษฎีเกี่ยวกับอนุกรมเวลา
- 2.3 วิธี Box และ Jenkins
- 2.4 วิธีป่าสุ่ม
- 2.5 วิธีโครงข่ายประสาทเทียม
- 2.6 การวัดประสิทธิภาพของการพยากรณ์
- 2.7 งานวิจัยที่เกี่ยวข้อง

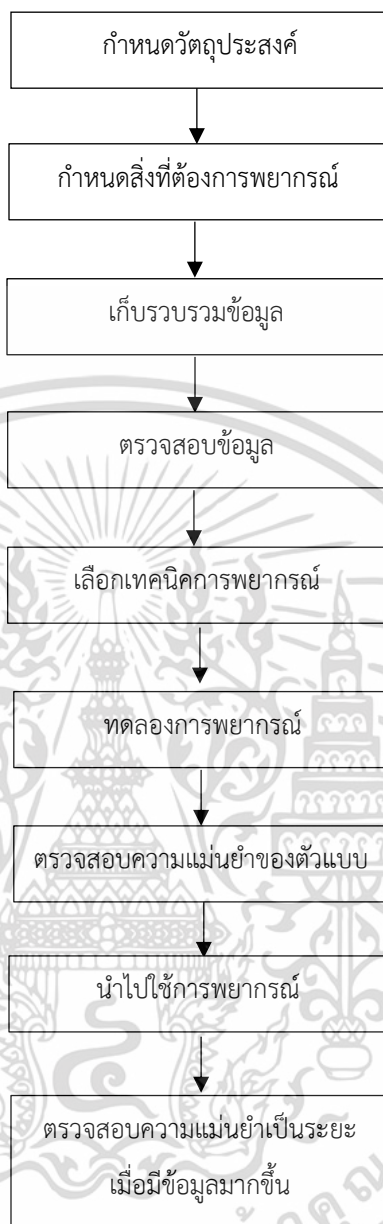
#### 2.1 ทฤษฎีเกี่ยวกับการพยากรณ์

##### 2.1.1 ความหมายของการพยากรณ์

การพยากรณ์ หมายถึง การคาดคะเนหรือทำนายการเกิดของเหตุการณ์หรือสภาพการณ์ต่าง ๆ ในอนาคต โดยการพยากรณ์จะทำจากการศึกษาแนวโน้มและรูปแบบการเกิดของเหตุการณ์หรือสภาพการณ์ จากข้อมูลในอดีตหรือใช้ความรู้ ความสามารถ ประสบการณ์ และวิจรณ์ญาณของผู้พยากรณ์ การพยากรณ์มีความจำเป็นอย่างยิ่งต่อการวางแผนและการตัดสินใจเกี่ยวกับการดำเนินงานของบุคคลทุกสาขาอาชีพ และของทุกองค์กรที่มีกิจกรรมในสาขาต่างๆ เช่น ธุรกิจอุตสาหกรรม เกษตร การเมือง การสาธารณสุข เป็นต้น

##### 2.1.2 ขั้นตอนในการพยากรณ์

กระบวนการพยากรณ์สามารถแบ่งออกเป็นขั้นตอนจะเห็นได้ว่ากระบวนการพยากรณ์ทั้ง 9 ขั้นตอน ต้องมีการพยากรณ์ซ้ำๆ และตรวจสอบผลการพยากรณ์เป็นระยะๆ เพื่อให้ได้ค่าพยากรณ์ที่ถูกต้องมากที่สุด



รูปที่ 2.1 ขั้นตอนกระบวนการพยากรณ์ (ปฎิมาพร อุดม, 2556)

## 2.2 ทฤษฎีเกี่ยวกับอนุกรมเวลา

### 2.2.1 ความหมายของอนุกรมเวลา

อนุกรมเวลา หมายถึง ค่าข้อมูลหรือค่าสังเกตที่เก็บรวบรวมตามลำดับเวลาที่เกิดขึ้น อย่างต่อเนื่องช่วงเวลาที่เก็บรวบรวมข้อมูลอาจห่างเท่ากันหรือไม่เท่ากันก็ได้ แต่ส่วนใหญ่ช่วงเวลาจะห่างเท่ากัน โดยอาจจะเป็นรายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือรายปี ก็ได้ เช่น ราคาหุ้นของ

ธนาकारที่เปลี่ยนแปลง ไปเป็นรายวัน ยอดขายของห้างสรรพสินค้ารายปี จำนวนผู้ป่วยโรคไข้เลือดออก รายสัปดาห์

## 2.2.2 ส่วนประกอบของอนุกรมเวลา

ข้อมูลอนุกรมเวลามีส่วนประกอบที่สำคัญ 4 ส่วน คือ

### 1) แนวโน้ม (Trend)

แนวโน้ม หมายถึง การเคลื่อนไหวของอนุกรมเวลาในระยะยาวว่าน่าจะมีแนวโน้มเพิ่มขึ้น หรือ ลดลงและลักษณะแนวโน้มนั้นอาจจะมีลักษณะเป็นเส้นตรงหรือเส้นโค้งก็ได้ ระยะเวลาที่จะทำให้สามารถ เห็นแนวโน้มส่วนใหญ่ควรจะไมต่ำกว่า 10 ช่วงเวลา การอธิบายแนวโน้มจะอธิบายจากเส้นที่เรียกว่า เส้นแนวโน้ม และค่าที่ได้จากแนวโน้มเราเรียกว่า ค่าแนวโน้ม ลักษณะเด่นของเส้นแนวโน้มคือจะต้องเรียบ ไม่มีการหักมุม ณ ที่ใด ๆ ของเส้นแนวโน้ม ไม่ว่าเส้นแนวโน้มนั้นจะเป็นเส้นตรงหรือเส้นโค้งก็ตาม

### 2) ความผันแปรตามฤดูกาล (Seasonal Variation)

ความผันแปรตามฤดูกาล หมายถึง การเปลี่ยนแปลงของข้อมูลที่เกิดขึ้น เนื่องจากอิทธิพลของ ฤดูกาล ซึ่งจะเกิดขึ้นซ้ำๆกันในช่วงเดียวกันของแต่ละปี โดยทั่วไปช่วงเวลาของฤดูกาลหนึ่งๆ มักจะสั้นกว่า 1 ปี

### 3) ความผันแปรตามวัฏจักร (Cyclical Variation)

ความผันแปรตามวัฏจักร หมายถึงการเคลื่อนไหวของข้อมูลที่มีลักษณะซ้ำๆกัน คล้ายกับความผันแปรตามฤดูกาลต่างกันที่ระยะเวลาของการเคลื่อนไหวของข้อมูลจะมีระยะเวลานานกว่า 1 ปี โดยทั่วไป ความผันแปรตามวัฏจักรมักจะพบในวัฏจักรของธุรกิจซึ่งมีแบบแผนของการเปลี่ยนแปลง 4 ระยะ คือ ระยะเจริญรุ่งเรือง (Prosperity) ระยะถดถอยหรือชะงักงัน (Recession) ระยะตกต่ำหรือหยุดอยู่กับที่ (Depression) ระยะฟื้นตัว (Recovery)

### 4) ความผันแปรเนื่องจากเหตุการณ์ที่ผิดปกติ (Irregular Variation)

ความผันแปรเนื่องจากเหตุการณ์ที่ผิดปกติ หมายถึงการเคลื่อนไหวของข้อมูลที่ไม่มีความเป็นแบบที่แน่นอน ลักษณะของข้อมูลที่เกิดขึ้นส่วนใหญ่จะเป็นลักษณะของเหตุการณ์ที่ไม่ได้คาดการณ์เอาไว้ล่วงหน้า

## 2.3 วิธี Box และ Jenkins

ในปี ค.ศ. 1972 George E.P. Box และ Gwilym M. Jenkins ได้พัฒนาวิธีการวิเคราะห์ข้อมูลอนุกรมเวลาขึ้น โดยใช้ตัวแบบอาร์มาหรือวิธีบ็อกซ์-เจนกินส์ ภายใต้แนวคิดที่ว่าอนุกรมเวลาที่ พิจารณาสามารถอธิบายได้ด้วยตัวแบบสโตคาสติก (Stochastic model) ซึ่งมีลักษณะของสหสัมพันธ์ในตัวเอง (Autocorrelation) และสหสัมพันธ์ในตัวเองบางส่วน (Partial autocorrelation) ที่แตกต่างกันไปเมื่อตัวแบบมีความแตกต่างกัน จึงมีการเลือกตัวแบบที่ใช้ในการพยากรณ์ โดยพิจารณาจากลักษณะของสหสัมพันธ์ในตัวเองและสหสัมพันธ์ในตัวเองบางส่วน และกำหนดตัวแบบสโตคาสติกที่เป็นไปได้ ให้เป็นตัวแบบสโตคาสติกที่มีลักษณะของสหสัมพันธ์ในตัวเอง และสหสัมพันธ์ในตัวเองบางส่วนที่เหมือนกัน โดยตัวแบบที่เป็นไปได้อาจมีได้มากกว่า 1 ตัวแบบ ซึ่งต้องทำการประมาณค่าพารามิเตอร์ของตัวแบบที่กำหนดขึ้นและตรวจสอบ ความถูกต้องและเหมาะสมของตัวแบบนั้น ก่อนที่จะนำตัวแบบไปใช้ในการพยากรณ์ต่อไป จึงทำให้การพยากรณ์ โดยวิธีบ็อกซ์-เจนกินส์มีความถูกต้องแม่นยำ และได้รับความนิยมอย่างมาก

การกำหนดรูปแบบ ARMA(p,q) ที่เหมาะสมให้กับอนุกรมเวลา จะได้จากการพิจารณาว่าอนุกรมเวลาที่มีค่าวัดลักษณะบางค่าได้แก่ ค่าสัมประสิทธิ์สหสัมพันธ์ในตนเองและค่าสัมประสิทธิ์สหสัมพันธ์ในตนเองบางส่วนของตัวอย่าง ( $r_k$  และ  $r_{kk}$ ) สอดคล้องกับค่าวัดลักษณะ ได้แก่ ค่าสัมประสิทธิ์สหสัมพันธ์ในตนเอง และค่าสัมประสิทธิ์สหสัมพันธ์ในตนเองบางส่วนของประชากร ( $\rho_k$  และ  $\rho_{kk}$ ) ของรูปแบบ ARMA(p,q) ใด ขนาดของอนุกรมเวลาจะต้องใหญ่พอสมควร นั่นคืออนุกรมเวลาควรมีขนาดเกิน 30 เพราะจะต้องหาค่า  $r_k$  และ  $r_{kk}$  หลายค่าเพื่อจะได้นำมาเปรียบเทียบกับ  $\rho_k$  และ  $\rho_{kk}$

อนุกรมเวลาที่จะนำมาศึกษาเพื่อประโยชน์ในการพยากรณ์มีลักษณะต่าง ๆ กันหลายแบบ โดยวิธีของ Box และ Jenkins จะแบ่งอนุกรมเวลาออกเป็น 2 ประเภทดังนี้

1) อนุกรมเวลาที่เป็นสเตชันนารี (Stationary Series) เป็นอนุกรมเวลา  $Y_t$  ที่มีค่าเฉลี่ยและความแปรปรวน  $Y_t$  คงที่ นั่นคือค่าเฉลี่ย  $E(Y_t)$  และค่าความแปรปรวน  $V(Y_t)$  มีค่าคงที่ สำหรับแต่ละเวลา  $t$  ซึ่งอนุกรมเวลาที่มีแนวโน้ม และฤดูกาลจะมี  $E(Y_t)$  ไม่คงที่ ส่วนอนุกรม เวลาที่มีความผันแปรของ  $Y_t$  สูงจะเป็นลักษณะของอนุกรมเวลาที่  $V(Y_t)$  ไม่คงที่ จะเรียกว่าอนุกรมเวลาที่ไม่เป็นสเตชันนารี

นอกจากจะเป็นอนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่แล้วอนุกรมเวลาที่เป็นสเตชันนารีจะต้องมีค่าสัมประสิทธิ์สหสัมพันธ์ในตนเองที่ lag  $k$  ขึ้นอยู่กับค่า  $k$  อย่างเดียว อนุกรมเวลาที่จะกำหนดรูปแบบ ARMA(p,q) ให้จะต้องเป็นอนุกรมเวลาที่เป็นสเตชันนารีเท่านั้น

2) อนุกรมเวลาที่ไม่เป็นสเตชันนารี (Nonstationary Series) เป็นอนุกรมเวลาที่ไม่มีความสมบัติเป็นสเตชันนารี ARMA(p,q) ให้กับอนุกรมเวลาดังกล่าวไม่ได้ จะต้องแปลงอนุกรมเวลานั้นให้เป็นอนุกรมเวลาใหม่ที่มีความสมบัติสเตชันนารีเสียก่อน จึงจะหารูปแบบ ARMA(p,q) ให้กับอนุกรมเวลาใหม่ได้ การแปลงอนุกรมเวลาเดิมให้เป็นอนุกรมเวลาใหม่ที่เป็นสเตชันนารีจะทำได้ด้วยวิธีการต่าง ๆ ดังนี้

ก. หาผลต่าง (Regular Differencing) ของอนุกรมเวลา นั่นคือ ถ้าอนุกรมเวลา  $Y_t$  มีแนวโน้ม จะแปลงให้เป็นอนุกรมเวลาใหม่ที่ไม่มีแนวโน้ม  $Z_t$  โดย  $Z_t = \nabla^d Y_t$  และ  $d$  เป็นจำนวนครั้งของการหาผลต่าง เช่น

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

เมื่อ  $d=2 \quad Z_t = \nabla^2 Y_t = \nabla(Y_t - Y_{t-1}) = \nabla Y_t - \nabla Y_{t-1} = Y_t - Y_{t-1} - Y_{t-1} + Y_{t-2} = Y_t - 2Y_{t-1} + Y_{t-2}$  เป็นต้น จำนวนครั้งที่หาผลต่างจะขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็นสเตชันนารีหรือไม่ ถ้ายังไม่เป็นสเตชันนารีต้องหาผลต่างต่อไป โดยทั่วไปถ้าอนุกรมเวลามีแนวโน้มเป็นแบบเส้นตรง จะใช้  $d$  เป็น 1 ส่วนเมื่ออนุกรมเวลามีแนวโน้มเป็นแบบ Quadratic จะใช้  $d$  เป็น 2

ข. หาผลต่างฤดูกาล (Seasonal Differencing) ของอนุกรมเวลา ถ้าอนุกรมเวลามีอิทธิพลของฤดูกาลเข้ามาเกี่ยวข้อง จะแปลงอนุกรมเวลาเดิม  $Y_t$  ให้เป็นอนุกรมเวลาใหม่ที่ไม่มีฤดูกาล  $Z_t$  โดย  $Z_t = \nabla_L^D Y_t$  โดย  $D$  เป็นจำนวนครั้งของการหาผลต่างฤดูกาลและ  $L$  เป็นจำนวนฤดูกาลต่อปี เช่น สำหรับอนุกรมเวลารายเดือน ( $L=12$ )

$$\text{เมื่อ } D=1 \quad Z_t = \nabla_{12} Y_t = Y_t - Y_{t-12}$$

เมื่อ  $D=2 \quad Z_t = \nabla_{12}^2 Y_t = \nabla_{12}(Y_t - Y_{t-12}) = \nabla Y_t - \nabla Y_{t-12} = Y_t - Y_{t-12} - Y_{t-12} + Y_{t-24} = Y_t - 2Y_{t-12} + Y_{t-24}$  เป็นต้น ผลต่างนี้จะทำก็ครั้งขึ้นอยู่กับว่าเมื่อหาผลต่างแล้วอนุกรมเวลาใหม่เป็นสเตชันนารีแล้วหรือไม่ ถ้ายังไม่เป็นก็ต้องหาผลต่างต่อไป

ค. หาผลต่างและผลต่างฤดูกาล กรณีที่อนุกรมเวลามีทั้งแนวโน้มและฤดูกาล การปรับให้อนุกรมเวลาเป็นสเตชันนารีนั้นจะทำได้โดยหาผลต่างและผลต่างฤดูกาลควบคู่กันไป  $d$  และ  $D$  จะมีค่าเป็นเท่าไรนั้นขึ้นอยู่กับว่าอนุกรมเวลาใหม่เป็นสเตชันนารีแล้วหรือยัง เช่น อนุกรมรายเดือนที่มีทั้งแนวโน้มและฤดูกาล

เมื่อ  $d = 1$  และ  $D = 1$  จะแปลงอนุกรมเวลาเดิม  $Y_t$  ให้เป็นอนุกรมเวลาใหม่  $Z_t$  ซึ่ง  $Z_t = \nabla \nabla_{12} Y_t = \nabla(Y_t - Y_{t-12}) = \nabla Y_t - \nabla Y_{t-12} = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$  เป็นต้น

ง. การหาลอการิทึมของค่าสังเกตในอนุกรมเวลา นั่นคือแปลงอนุกรมเวลาเดิม  $Y_t$  ให้ เป็นอนุกรมเวลาใหม่  $Z_t$  ซึ่ง  $Z_t = \log(Y_t)$  การแปลงนี้จะทำเมื่อ ความผันแปรของอนุกรมเวลาไม่คงที่ นั่นคือ  $V(Y_t)$  ไม่คงที่สำหรับค่า  $t$  ต่างๆ

### 2.3.1 ขั้นตอนการพยากรณ์โดยวิธีของ Box และ Jenkins

มี 4 ขั้นตอนได้แก่

1) การกำหนดรูปแบบ (Identification) เป็นการหารูปแบบ ARMA(p,q) ที่คาดว่าจะเหมาะสมให้กับอนุกรมเวลาโดยพิจารณาเปรียบเทียบค่า  $r_k$  และ  $r_{kk}$  ของอนุกรมเวลากับค่า  $\rho_k$  และ  $\rho_{kk}$  ของแต่ละรูปแบบ เนื่องจากต้องพิจารณา  $r_k, r_{kk}, \rho_k$  และ  $\rho_{kk}$  พร้อมกันหลายๆค่า จึงมักจะพิจารณาจากรูปที่เรียกว่าคอเรลโรแกรม (Correlogram) ที่ได้จากการพล็อต  $r_k, r_{kk}, \rho_k$  และ  $\rho_{kk}$  กับ  $k$  ดังนั้นการพิจารณาเปรียบเทียบจะเป็นการเปรียบเทียบคอเรลโรแกรมของ  $r_k$  กับ  $\rho_k$  และคอเรลโรแกรมของ  $r_{kk}$  กับ  $\rho_{kk}$  สำหรับแต่ละรูปแบบมีคอเรลโรแกรมของ  $\rho_k$  และ  $\rho_{kk}$  ต่างกัน อนุกรมเวลาที่จะนำมากำหนดรูปแบบจะต้องเป็นอนุกรมเวลาที่สเตชันนารีเท่านั้น หากไม่เป็นสเตชันนารีจะต้องแปลงให้เป็นสเตชันนารีเสียก่อน

2) การประมาณค่าพารามิเตอร์ในรูปแบบ (Estimation) จะทำได้โดยการหาค่าประมาณแบบง่ายหรือค่าประมาณที่ได้จากการวิเคราะห์ตัวเลข (Numerical Analysis) สำหรับค่าประมาณแบบง่ายจะทำได้โดยการสร้างสมการที่มาจากความสัมพันธ์ระหว่าง  $\rho_k$  และพารามิเตอร์ โดยสมการที่สร้างขึ้นจะมีจำนวนเท่ากับจำนวนพารามิเตอร์ที่ต้องการประมาณ ส่วนค่าประมาณที่ได้จากการวิเคราะห์ตัวเลขจะได้รับการแก้สมการที่สร้างขึ้นจากวิธีกำลังสองน้อยที่สุด ขั้นตอนของการวิเคราะห์ตัวเลขจะต้องมีการหาค่าประมาณเริ่มต้น ซึ่งส่วนใหญ่จะใช้ค่าประมาณแบบง่ายเป็นค่าประมาณเริ่มต้น เมื่อการวิเคราะห์สิ้นสุดจะได้ค่าประมาณสุดท้ายที่จะนำไปใช้ประโยชน์ในการสร้างสมการพยากรณ์

3) การตรวจสอบรูปแบบ (Diagnostic Checking) เมื่อกำหนดรูปแบบ และประมาณค่าพารามิเตอร์ในรูปแบบแล้ว จะต้องตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ การตรวจสอบจะทำได้หลายวิธีได้แก่ การพิจารณาคอเรลโรแกรมของ  $r_k$  ค่าคลาดเคลื่อน การทดสอบค่าพารามิเตอร์ในรูปแบบโดยการทดสอบแบบ  $t$  และการทดสอบความเหมาะสมของรูปแบบโดยการทดสอบของ Box และ Pierce หรือการทดสอบของ Box และ Ljung หากตรวจสอบพบว่ารูปแบบที่กำหนดนั้นเหมาะสมแล้วจะใช้รูปแบบนั้นในการพยากรณ์ต่อไป แต่หากพบว่ารูปแบบที่กำหนดนั้นไม่

เหมาะสมจะต้องทำตามขั้นตอนที่ 1 เพื่อกำหนดรูปแบบใหม่

4) การพยากรณ์ จะทำได้ทั้งการพยากรณ์แบบจุด (Point Forecast) และการพยากรณ์แบบช่วง (Interval Forecast) โดยการพยากรณ์จะใช้สมการพยากรณ์ที่สร้างจากรูปแบบการพยากรณ์ที่กำหนด และผ่านการตรวจสอบในขั้นตอนที่ผ่านมาแล้ว

### 2.3.2 รูปแบบ ARIMA

รูปแบบ Autoregressive Intergrade Moving Average Model จากระยะเอียงนำแบบจำลอง Autoregressive แบบจำลอง Moving Average และกระบวนการ Integrated มาพิจารณา รวมกันสามารถนำมากำหนดเป็นรูปแบบทั่วไปของแบบจำลอง ARIMA(p,d,q) ที่ใช้ในการประเมินการตั้งสมการ

$$\nabla^d Y_t = \delta + \phi \nabla^d Y_{t-1} + \phi \nabla^d Y_{t-2} + \dots + \phi \nabla^d Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

โดยที่  $Y_t$  คือ ค่าสังเกตในอนุกรมเวลา ณ เวลา  $t$

$d$  คือ จำนวนครั้งของการหาผลต่างเพื่อให้อนุกรมเวลามีคุณสมบัติคงที่ (Stationary)

$p$  คือ อันดับของ Autoregressive

$q$  คือ อันดับของ Moving Average

$\delta$  คือ ค่าคงที่

$t$  คือ เวลา

$\nabla^d$  คือ ผลต่างอันดับที่  $d$

$\phi_1, \dots, \phi_q$  คือ พารามิเตอร์ของ Auto Regressive

$\theta_1, \dots, \theta_q$  คือ พารามิเตอร์ของ Moving Average

$\varepsilon_t$  คือ ค่าความคลาดเคลื่อน ณ เวลา  $t$

### 2.3.3 รูปแบบ ARIMAX

รูปแบบ Autoregressive Intergrade Moving Average with Model แบบจำลอง ARIMAX เป็นการอธิบายร่วมกันของแบบจำลอง ARIMA กับตัวแปรภายนอกที่เป็นปัจจัยที่เกี่ยวข้องหรือปัจจัยอื่น (X) ที่น่าจะมีอิทธิพลต่อ  $Y_t$  สามารถนำมากำหนดเป็นรูปแบบทั่วไปของแบบจำลอง ARIMAX(p,d,q) ที่ใช้ในการประเมินการตั้งสมการ

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \sum_{i=1}^r \beta_i X_{it}$$

### 2.3.4 สหสัมพันธ์ Autocorrelation Function

ฟังก์ชันของการวัดสหสัมพันธ์ Autocorrelation Function ระหว่างข้อมูล ณ เวลา  $t$  ( $y_t$ ) และข้อมูล ณ เวลา  $t-k$  ( $y_{t-k}$ ) ของช่วงเวลาห่างกัน  $k$  หน่วย ซึ่งแทนด้วยสัญลักษณ์  $\rho_k$  หรือ  $r_k$  ในกรณีสหสัมพันธ์ในตัวเองของตัวอย่าง ซึ่งสามารถคำนวณ ดังสมการ

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}$$

### 2.3.5 ตัวแบบของอนุกรมเวลา

ตัวแบบอนุกรมเวลาแบบคลาสสิก มีตัวแบบพื้นฐานอยู่ 2 ตัวแบบ คือ

1) ตัวแบบคูณ (Multiplicative Model)

$$Y_t = T_t \times S_t \times C_t \times I_t$$

2) ตัวแบบบวก (Additive Model)

$$Y_t = T_t + S_t + C_t + I_t$$

โดยที่  $Y_t$  คือ ค่าสังเกต ณ ช่วงเวลา  $t$

$T_t$  คือ ค่าแนวโน้ม ณ ช่วงเวลา  $t$

$S_t$  คือ ค่าอิทธิพลของฤดูกาล ณ ช่วงเวลา  $t$

$C_t$  คือ ค่าอิทธิพลของวัฏจักร ณ ช่วงเวลา  $t$

$I_t$  คือ ค่าเหตุการณ์ที่ผิดปกติ ณ ช่วงเวลา  $t$

อนุกรมเวลารายปี ค่าสังเกตรายปีทำให้องค์ประกอบของอิทธิพลของฤดูกาลหายไป จะได้รูปแบบของอนุกรมเวลาเป็น

$$\text{รูปแบบคูณ : } Y_t = T_t \times C_t \times I_t$$

$$\text{รูปแบบบวก : } Y_t = T_t + C_t + I_t$$

อนุกรมเวลารายไตรมาสหรือรายเดือน จะกำหนดรูปแบบของอนุกรมเวลาเป็น

$$\text{รูปแบบคูณ : } Y_t = T_t \times S_t \times C_t \times I_t$$

$$\text{รูปแบบบวก : } Y_t = T_t + S_t + C_t + I_t$$

### 2.3.6 การทดสอบความนิ่งของข้อมูล (Unit Root Test)

เนื่องจากข้อมูลที่นำมาใช้ในการศึกษานี้เป็นข้อมูลอนุกรมเวลา (Time series data) ซึ่งส่วนมากจะมีลักษณะเป็น Non-stationary กล่าวคือ ค่าเฉลี่ย (Mean) และค่าความแปรปรวน (Variances) ของข้อมูลจะมีค่าไม่คงที่เปลี่ยนแปลงไปตามกาลเวลา โดยอาจมีแนวโน้ม (Trend) ในระยะยาว และขณะเดียวกันก็มีการแกว่งตัวระยะสั้น (Cyclical Swing) ขึ้นอยู่กับสิ่งที่มากระทบ (Shock) ดังนั้นการใช้วิธีสมการถดถอยแบบกำลังสองน้อยที่สุด (Ordinary Least Square: OLS) ในการประมาณค่า อาจจะทำให้ผลการศึกษาความสัมพันธ์ออกมาผิดพลาด หรือเรียกว่าเป็น Spurious regression ดังนั้นจึงมีความจำเป็นที่จะต้องทดสอบ Unit Root ก่อน (ชนานุช จันทรา, 2552)

วิธีการที่ง่ายที่สุดในการพิจารณาคุณลักษณะของข้อมูลของตัวแปร นั่นคือการ Plot กราฟ เพื่อดูภาพรวมของความแปรผัน (Variation) ของข้อมูล เราทราบมาแล้วว่าคุณลักษณะ Stationary จะมีค่าพารามิเตอร์ ค่าเฉลี่ย และความแปรปรวนที่ไม่ปรากฏแนวโน้มหรือรูปแบบ แต่หากพบลักษณะดังกล่าว จำเป็นต้องตรวจสอบต่อไป (รักชนก นุชพงษ์, 2540) และผู้วิจัยได้นำข้อมูลมาทดสอบ Unit Root ด้วยวิธี Augmented Dickey-Fuller (ADF) Test โดยวิธี ADF Test จะมีรายละเอียดของสมการดังนี้

$$\Delta Y_t = \theta Y_{t-1} + \sum_{i=1}^p \phi_i \Delta Y_{t-i} + \varepsilon_t \quad \text{กรณีไม่มีค่าคงที่และแนวโน้ม}$$

$$\Delta Y_t = \alpha + \gamma Y_{t-1} + \sum_{i=1}^p \phi_i \Delta Y_{t-i} + \varepsilon_t \quad \text{กรณีมีเฉพาะค่าคงที่}$$

$$\Delta Y_t = \alpha + \beta t + \theta Y_{t-1} + \sum_{i=1}^p \phi_i \Delta Y_{t-i} + \varepsilon_t \quad \text{กรณีมีทั้งค่าคงที่และแนวโน้ม}$$

โดยที่  $Y_t, Y_{t-1}, Y_{t-i}$  คือ ข้อมูลอนุกรมเวลาของตัวแปรอิสระ ณ เวลา  $t, t-1, t-i$

$\alpha$  คือ ค่าคงที่

$\beta, \theta, \phi$  คือ ค่าพารามิเตอร์

$t$  คือ ค่าแนวโน้มเวลา

$i$  คือ ค่าความล่าช้า

$\varepsilon_t$  คือ ความคลาดเคลื่อนเชิงสุ่ม (Random Error)

สมมติฐานในการทดสอบความนิ่งของข้อมูลเป็นดังนี้

สมมติฐานหลัก  $H_0: \theta = 0$  มี Unit Root (มีลักษณะไม่นิ่ง)

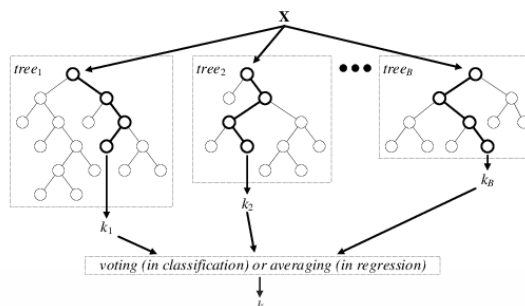
สมมติฐานรอง  $H_1: \theta \neq 0$  ไม่มี Unit Root (มีลักษณะนิ่ง)

เกณฑ์ในการตัดสินใจในการทดสอบ Unit Root ด้วยวิธี Augmented Dickey-Fuller (ADF) Test คือ การเปรียบเทียบค่าสถิติ Augmented Dickey - Fuller test Statistic กับค่าสถิติ MacKinnon Critical ณ ระดับนัยสำคัญร้อยละ 5 แล้วค่าสถิติ Augmented Dickey - Fuller test Statistic มีค่ามากกว่า ค่าสถิติ MacKinnon Critical แสดงว่ายอมรับสมมติฐานหลัก ( $H_1$ ) ข้อมูลมีลักษณะไม่นิ่ง แต่ถ้าค่าสถิติ Augmented Dickey - Fuller test Statistic มีค่าน้อยกว่า ค่าสถิติ MacKinnon Critical แสดงว่าปฏิเสธสมมติฐานหลัก ( $H_0$ ) ข้อมูลมีลักษณะนิ่ง (จริยภรณ์และธงชัย, 2561)

## 2.5 วิธีป่าสุ่ม

Random Forest หรือวิธีป่าสุ่มเป็นอีกการทำนายที่ใช้พื้นฐานจากต้นไม้ตัดสินใจ (Decision Tree) เป็นการทำนายแบบชุดของ Decision Tree หลายๆ ต้น (Ensemble of Decision Trees) โดยสร้างจากการสุ่มข้อมูลตัวอย่างแบบเลือกแล้วใส่กลับ (Random Sampling with Replacement) เพื่อนำมาสร้างเป็นแบบจำลองต้นไม้โดยแต่ละต้นมีลักษณะที่ไม่ซ้ำกัน โดยแต่ละแบบจำลองจะมีการทำนายผล ซึ่งผลจากการทำนายของต้นไม้แต่ละต้นจะทำการโหวตเลือกผลการทำนายที่ได้รับการโหวตมากที่สุด วิธีการนี้เรียกว่า Bagging หรือ Bootstrapping โดยงานวิจัยนี้ใช้ Random Forest ของ Scikit-Learn (Random Forest Regressor) จะได้ผลลัพธ์ออกมาเป็นค่าความน่าจะเป็นสำหรับพารามิเตอร์หลักที่ปรับได้ในการสร้างแบบจำลอง Random Forest ได้แก่ (ธนัท จรรย์สมบูรณ์, 2561)

- n estimator: จำนวนต้นไม้ทั้งหมดในแบบจำลองการทำนาย ซึ่งโดยทั่วไปมีค่ายิ่งมากทำให้การทำนายแม่นยำ แต่จะใช้เวลานานในการคำนวณ
- max features: จำนวนของฟีเจอร์ที่ถูกสุ่มมาสร้างต้นไม้ตัดสินใจแต่ละต้น
- min\_samples\_leaf: จำนวนข้อมูลอย่างน้อยที่สุ่มมาใช้เป็น leaf node



รูปที่ 2.2 แบบจำลองป่าสุ่ม

(ที่มา: <https://medium.com/mmp-li/random-forest-ทำนายว่าจักรยานจะขายได้กี่ต่อเมื่อ-c60080354380> )

## 2.6 วิธีโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Artificial Neural Network) เป็นศาสตร์แขนงหนึ่งของปัญญาประดิษฐ์ มีลักษณะเป็นโมเดลทางคณิตศาสตร์ ที่จำลองการทำงานของเซลล์ประสาทในสมองของมนุษย์ อาศัยการเรียนรู้และจดจำรูปแบบคล้ายกับสมองมนุษย์ จึงทำให้โครงข่ายประสาทเทียมถูกนำมาประยุกต์ใช้อย่างแพร่หลายเพื่อตัดสินใจตัดแยกหรือแบ่งกลุ่มข้อมูลออกจากกันในงานทางด้านวิทยาการคอมพิวเตอร์ และวิศวกรรมที่หลากหลาย เช่น การจดจำรูปแบบที่มีความไม่แน่นอน ลายมือ ลายเซ็น ตัวอักษร รูปหน้า รวมถึงการทำนายต่างๆ เช่น ทำนายระดับน้ำ โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Back propagation) มีลักษณะโครงสร้างการจัดเซลล์ประสาทเทียมเป็นชั้น (Layer) ดังรูปที่ 2.3 โดยชั้นที่รับข้อมูลเข้าเรียกว่าชั้นอินพุต (Input layer) ชั้นที่ผลิตผลของโครงข่ายเรียกว่าชั้นเอาต์พุต (Output layer) ส่วนชั้นอื่นๆ ที่มีส่วนในการช่วยประมวลผลภายในเรียกว่า ชั้นซ่อนเร้น (Hidden layer) ในระหว่างชั้นมีการเชื่อมต่อโดยกำหนดค่าน้ำหนัก (Weight) กำกับอยู่ที่เส้นเชื่อมทุกเส้น เมื่อข่ายงานเริ่มทำงานจะกำหนดค่าให้แก่ชั้นอินพุต ค่าอินพุตเหล่านี้อาจเป็นค่าที่มนุษย์เป็นผู้สร้างขึ้นหรือเป็นค่าที่วัดมาจากเซ็นเซอร์ หรือเป็นผลมาจากการทำงานของโปรแกรม ค่าจากชั้นอินพุตนี้จะถูกส่งต่อที่ได้รับไปตามเส้นที่เชื่อมของขาออกโดยนำไปคูณกับค่าถ่วงน้ำหนักของเส้นเชื่อม จากนั้นส่วนของในชั้นที่อยู่ถัดไปจะทำงานรับข้อมูลซึ่งเป็นผลรวมของชั้นต่างๆ แล้วคำนวณผลอย่างง่าย โดยทั่วไปจะใช้ฟังก์ชันซิกมอยด์ (Sigmoid function) แล้วส่งค่าไปยังชั้นถัดไป การคำนวณลักษณะนี้จะเกิดขึ้นไปเรื่อยๆ ทีละชั้นจนถึงชั้นเอาต์พุต (Livingstone, D. J., 2008)

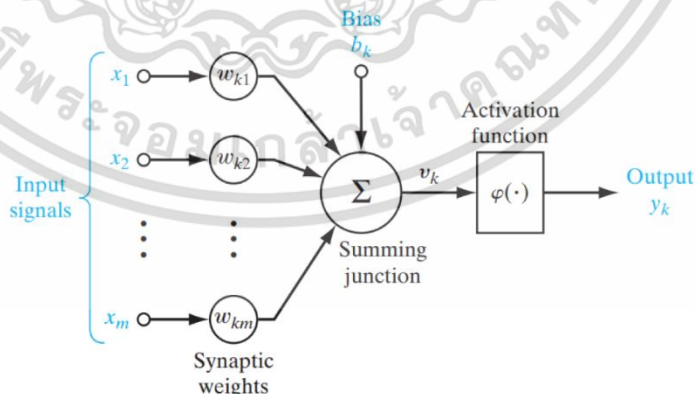
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กระบวนการนำโครงข่ายประสาทเทียมมาใช้งานแบ่งเป็น 2 ขั้นตอน ได้แก่ ขั้นตอนการฝึกหัด (Training) หรือการเรียนรู้ (Learning) และขั้นตอนการทดสอบ (Testing) หรือใช้งานจริง (Working) ซึ่งในขั้นตอนการเรียนรู้อย่างสามารถแบ่งออกเป็น 2 กลุ่มหลักคือ

1. การเรียนรู้แบบไม่มีครูสอน (Unsupervised Learning Algorithm) เป็นการเรียนรู้แบบไม่มีผู้แนะนำ ไม่มีการตรวจสอบคำตอบว่าถูกหรือผิด โครงสร้างของโครงข่ายประสาทเทียมชนิดนี้จะจัดเรียงโครงสร้างด้วยตัวเองตามลักษณะของข้อมูล ผลลัพธ์ที่ได้โครงข่ายประสาทเทียมจะสามารถจัดหมวดหมู่ของข้อมูลได้

2. การเรียนรู้แบบมีครูสอน (Supervised Learning Algorithm) เป็นการเรียนรู้แบบที่มีการตรวจสอบคำตอบเพื่อให้โครงข่ายประสาทเทียมปรับตัว ชุดข้อมูลที่ใช้สอนโครงข่ายประสาทเทียมจะมีคำตอบไว้คอยตรวจสอบว่าโครงข่ายประสาทเทียมให้คำตอบที่ถูกหรือไม่ ถ้าตอบไม่ถูก โครงข่ายประสาทเทียมก็จะปรับตัวเองเพื่อให้ได้คำตอบที่ดีขึ้น

โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ (Feed Forward Back Propagation) หรือของโครงข่ายประสาทเทียมชนิดเพอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron: MLP) เป็นโครงข่ายประสาทเทียมที่มีการเรียนรู้แบบมีครูสอน ที่มักใช้ทำงานกับงานคาดเดาสามารถช่วยกำหนดลำดับความสำคัญได้ สถาปัตยกรรมของโครงข่ายประสาทเทียมชนิดนี้เป็นที่นิยมสูงสุดและมีประสิทธิภาพสูง รวมถึงมีความง่ายสำหรับเป็นต้นแบบโครงข่ายประสาทเทียมที่มีความซับซ้อนมากขึ้นแบบหลายเลเยอร์ จุดเด่นที่สำคัญของโครงข่ายประสาทเทียมชนิดนี้คือ มีวิธีการทำงานแบบไม่เชิงเส้น (Non-Linear) (กิริติบุญโชติ, 2549)



รูปที่ 2.3 แบบจำลองโครงข่ายประสาทเทียม

(ที่มา: <https://towardsdatascience.com/building-a-simple-neural-network-from-scratch-a5c6b2eb0c34>)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 2.7 การวัดประสิทธิภาพของการพยากรณ์

- 1) ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error: MSE)

$$MSE = \frac{\sum e_t^2}{n}$$

- 2) ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE)

$$MAPE = \frac{\sum \left| \frac{e_t}{Y_t} \right|}{n} \times 100$$

## 2.8 ศูนย์บริการข้อมูลลูกค้า

Call Center หรือ ที่เรียกกันว่า ศูนย์บริการข้อมูลลูกค้า เป็นหน่วยงานที่มีเจ้าหน้าที่คอยให้บริการข้อมูล ข่าวสาร รวมถึงรับทำรายการทางธุรกิจต่างๆ โดยมีช่องทางการให้บริการทางโทรศัพท์ E-Mail Fax และ Internet การทำงานของศูนย์บริการข้อมูลลูกค้าจะทำงานบนพื้นฐานของ ระบบคอมพิวเตอร์และระบบโทรศัพท์ ที่สามารถเข้าถึงข้อมูลลูกค้าได้สะดวกและรวดเร็วเป็นส่วนหนึ่งที่จะช่วยเพิ่มประสิทธิภาพในการดำเนินธุรกิจ และใช้เป็นช่องทางสำหรับติดต่อสื่อสารและสร้างความสัมพันธ์ที่ดีกับลูกค้าผู้ใช้บริการ

- 2.8.1 การจัดตั้งศูนย์บริการข้อมูลลูกค้าสามารถทำได้ 2 แบบคือ

2.8.1.1 องค์กร หรือ บริษัทจัดตั้งศูนย์บริการข้อมูลลูกค้าโดยแยกเป็นแผนกหนึ่งในองค์กรขึ้นอยู่กับขนาดขององค์กร เงินทุน และจำนวนลูกค้าที่ติดต่อมายังศูนย์บริการข้อมูลลูกค้าของธุรกิจ สำหรับธุรกิจที่มีขนาดใหญ่นิยมจัดตั้งศูนย์บริการข้อมูลลูกค้าขึ้นมาเป็นของตนเอง ซึ่งต้องมีการลงทุนจัดตั้งระบบรวมทั้งการหาพนักงานการฝึกอบรมพนักงานและพัฒนาการบริการ ซึ่งจะต้องใช้เวลานาน สำหรับหน่วยงานราชการที่จัดตั้งศูนย์บริการข้อมูลลูกค้าเอง อาทิ กรมสรรพากร ซึ่งเจ้าหน้าที่ศูนย์บริการข้อมูลลูกค้าจะเป็นผู้ที่มีความรู้เฉพาะและสามารถแก้ไขปัญหาเกี่ยวกับกฎหมายเป็นส่วนใหญ่

2.8.1.2 ใช้บริการของบริษัทศูนย์บริการข้อมูลลูกค้าภายนอก (Outsource) โดยพนักงานของบริษัทภายนอกจะเป็นตัวแทนรับสายลูกค้าแทนเจ้าของธุรกิจ เพราะบริษัทภายนอกจะมีความพร้อมด้านบุคลากรและระบบที่ใช้ในการบริการศูนย์บริการข้อมูลลูกค้า เป็นการลดต้นทุนในการจัดตั้งระบบและลดความยุ่งยากในการบริหารงาน เพราะพนักงานของบริษัทภายนอกจะมีทักษะและประสบการณ์ด้านงานบริการพร้อมให้บริการอย่างมีประสิทธิภาพ ทั้งนี้ เจ้าของธุรกิจจะเป็นผู้อบรมความรู้ด้านผลิตภัณฑ์ให้แก่

## พนักงานของบริษัทภายนอก

### 2.8.2 ส่วนประกอบของศูนย์บริการข้อมูลลูกค้า

ศูนย์บริการข้อมูลลูกค้าจะประกอบด้วย 2 ส่วนใหญ่ๆ คือ ระบบศูนย์บริการข้อมูลลูกค้าและเจ้าหน้าที่ที่ทำงานในศูนย์บริการข้อมูลลูกค้า มีความรับผิดชอบและหน้าที่แตกต่างกัน เช่น

2.8.2.1 Agent (เจ้าหน้าที่รับสาย) ทำหน้าที่รับสายเข้าและ โทรออก เป็นผู้ให้บริการลูกค้า

โดยตรง เป็นผู้ที่มีความรู้ในสินค้าและบริการต่างๆ เป็นอย่างดี

2.8.2.2 Supervisor (หัวหน้างาน) ทำหน้าที่ปฏิบัติงานร่วมกับเจ้าหน้าที่รับสายอย่างใกล้ชิด หัวหน้างานต้องเป็นผู้ที่มีความรู้เกี่ยวกับตัวสินค้าหรือบริการต่างๆของลูกค้า และต้องเข้าใจในขั้นตอนการปฏิบัติงานเป็นอย่างดี สามารถแก้ปัญหาแทนเจ้าหน้าที่รับสายได้ในกรณี ที่การบริการเกิดข้อผิดพลาด รวมทั้งสามารถประเมินผลการปฏิบัติงานของเจ้าหน้าที่รับสายได้

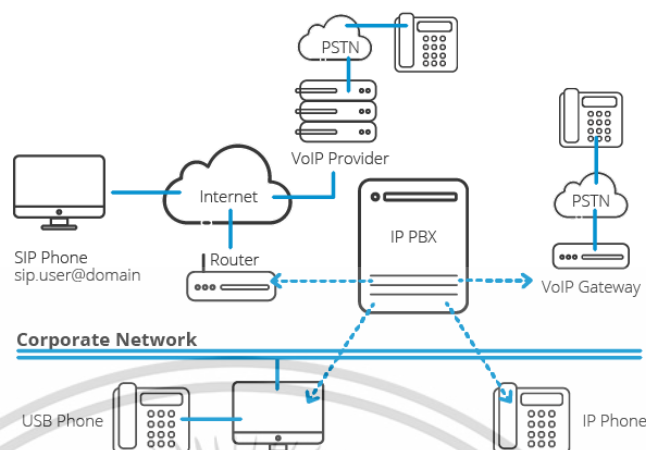
2.8.2.3 Trainer (เจ้าหน้าที่ฝึกอบรม) มีหน้าที่ฝึกอบรมเจ้าหน้าที่รับสายในเรื่องการให้บริการ และทุกเรื่องที่เจ้าหน้าที่รับสายพึงปฏิบัติก่อนการปฏิบัติงาน รวมทั้งการให้ความรู้เพิ่มเติมในเรื่องใหม่ๆ

2.8.2.4 Quality Analyst (เจ้าหน้าที่วิเคราะห์คุณภาพ) ทำหน้าที่ตรวจสอบคุณภาพการบริการ ให้เป็นไปตามมาตรฐานและตามความพึงพอใจของลูกค้า และมีหน้าที่ให้คำแนะนำแก่เจ้าหน้าที่รับสาย เกี่ยวกับเทคนิคการบริการลูกค้าให้มีคุณภาพ

2.8.2.5 IT Support (เจ้าหน้าที่สนับสนุนระบบคอมพิวเตอร์และระบบ Call center) มีหน้าที่ดูแลระบบให้การปฏิบัติงานของเจ้าหน้าที่รับสายไม่ติดขัดในขณะที่ให้บริการ รวมทั้งพัฒนาระบบต่างๆ เพื่อเพิ่มประสิทธิภาพของการปฏิบัติงานที่ถูกต้องและรวดเร็ว

### 2.8.3 เทคโนโลยีพื้นฐานสำหรับศูนย์บริการข้อมูลลูกค้า

ในปัจจุบันศูนย์บริการข้อมูลลูกค้าเข้ามามีบทบาทในธุรกิจมากขึ้นเทคโนโลยีที่ใช้สำหรับศูนย์บริการข้อมูลลูกค้ามีต้นทุนที่สูง สำหรับธุรกิจ หรือ SME ที่ต้องการจัดตั้งหน่วยงานศูนย์บริการข้อมูลลูกค้าและมีเงินลงทุนจำกัด ไม่จำเป็นต้องใช้ระบบที่ซับซ้อน สามารถนำเทคโนโลยีพื้นฐานมาใช้ในการดูแลลูกค้าได้ โดยการ ใช้ระบบตู้สาขา หรือ ที่เรียกสั้นๆว่า PBX (Private Branch Exchange) ที่มีอยู่แล้วในองค์กรจึงควรนำมาใช้ให้เกิดประโยชน์เพิ่มขึ้น โครงสร้างเบื้องต้น PBX Call Center ดังแสดงที่รูป 2.4 (พัชรินทร์, 2559)



รูป 2.4 โครงสร้างเบื้องต้น PBX Call Center  
(ที่มา: <https://www.3cx.de/voip-sip/pbx/>)

## 2.9 งานวิจัยที่เกี่ยวข้อง

Anggraeni et al. (2015) ศึกษาการพยากรณ์ปริมาณความต้องการชุดประกอบพีซีเด็กของชาวมุสลิมมีแนวโน้มเพิ่มขึ้นในช่วงเวลาโดยเฉพาะอย่างยิ่งใกล้วันฮีดซึ่งเป็นวันเฉลิมฉลองทางศาสนาอิสลาม ความต้องการในแต่ละปีมีความแตกต่างกันไปและความต้องการมักเพิ่มขึ้นเมื่อใกล้วันหยุดเทศกาลฮีด แต่จะลดลงในเดือนหน้าเข้าทุกปีและจะเกิดขึ้นในเวลาที่แตกต่างกันทุกปี หมายความว่าวันหยุดพักผ่อนในเทศกาลฮีดเกิดขึ้นในเวลาที่แตกต่างกันทุกปี ทำให้บริษัทมีความไม่แน่นอนของการผลิตสินค้า งานวิจัยนี้ได้เปรียบเทียบวิธีบ็อกซ์-เจนกินส์แบบไม่มีตัวแปรอิสระ  $X$  (Autoregressive Integrated Moving Average Model: ARIMA) และวิธีบ็อกซ์-เจนกินส์กับตัวแปรอิสระ  $X$  (Autoregressive Integrated Moving Average Model with  $X$ : ARIMAX) ในการเปรียบเทียบประสิทธิภาพใช้ค่าร้อยละของค่าสัมบูรณ์ของค่าความคลาดเคลื่อนโดยเฉลี่ย (MAPE) และเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด (RMPE) ผลจากการวิจัยพบว่า ตัวแบบของวิธีบ็อกซ์-เจนกินส์กับตัวแปรอิสระ  $X$  ให้ค่าความแม่นยำกว่าวิธีบ็อกซ์-เจนกินส์แบบไม่มีตัวแปรอิสระ  $X$

Sen et al. (2016) ศึกษาการพยากรณ์ปริมาณการใช้พลังงานและก๊าซเรือนกระจกการปล่อยก๊าซเรือนกระจกขององค์กรการผลิตเหล็กของประเทศอินเดีย ผู้ศึกษาทำการพยากรณ์วิธีบ็อกซ์-เจนกินส์ (Autoregressive Integrated Moving Average Model: ARIMA) และใช้ค่าร้อยละของค่าสัมบูรณ์ของค่าความคลาดเคลื่อนโดยเฉลี่ย (MAPE) และเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำ

ที่สุด (RMSE) ผลการศึกษาพบว่า การพยากรณ์วิธีบอซ-เจนกินส์ (Autoregressive Integrated Moving Average Model: ARIMA) ให้ค่าร้อยละของค่าสัมบูรณ์ของค่าความคลาดเคลื่อนโดยเฉลี่ย (MAPE) และเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยที่ต่ำที่สุด (RMPE) ที่น้อยมาก

Shamshad, B. et al. (2019) ศึกษาการสร้างแบบจำลองและการพยากรณ์สภาพอากาศโดยใช้ ANN-MLP, ARIMA และแบบจำลอง ETS: กรณีศึกษาสำหรับเมืองละฮอร์ ปากีสถาน เปรียบเทียบวิธีโครงข่ายประสาทเทียมแบบหลายชั้น (ANN-MLP), Exponential Smoothing Algorithm (ETS) และวิธีการของบอซ-เจนกินส์ (ARIMA) โดยใช้ข้อมูลอุตุนิยมวิทยาที่ได้จากเว็บไซต์ ([www.http://sdwebx.worldbank.org](http://sdwebx.worldbank.org)) ซึ่งเป็นข้อมูลสภาพอากาศในเมืองละฮอร์ของปากีสถาน ซึ่งเป็นข้อมูลค่าเฉลี่ยรายเดือนของอุณหภูมิต่ำสุด อุณหภูมิสูงสุด ความเร็วลม ความชื้นสัมพัทธ์และปริมาณฝน ตั้งแต่ปี 1987 ถึง 2014 รวม 30 ปี แบ่งเป็นชุดฝึกอบรม (1987 ถึง 2016) และชุดทดสอบ (2017 ถึง 2018) ผลจากการวิจัยพบว่า ตัวแบบของ ANN-MLP ให้ค่าความแม่นยำกว่าตัวแบบของ ARIMA และ ETS

Albrecht, T. et al. (2021) ศึกษาวิธีการและการนำไปประยุกต์ใช้ไปปฏิบัติจริงในการพยากรณ์สายโทรเข้าของคอลเซ็นเตอร์ โดยเปรียบเทียบความแม่นยำในการพยากรณ์ระหว่างวิธี ARIMA, Random forest, Gradient boosting, K-nearest neighbor และ Support vector regression ใช้ข้อมูลคอลเซ็นเตอร์ของผู้ค้าปลีกออนไลน์ชั้นนำของประเทศเยอรมัน ซึ่งเก็บรวบรวมข้อมูลจำแนกเป็นรายครึ่งชั่วโมง ตั้งแต่วันที่ 2 เดือนมกราคม ค.ศ. 2016 ถึงวันที่ 7 เดือนพฤษภาคม ค.ศ. 2019 ผลการศึกษาพบว่า วิธีที่มีความเหมาะสม คือวิธีการพยากรณ์โดย Random Forest

วรารพร งามสุข (2555) ศึกษาการพยากรณ์เพื่อเปรียบเทียบการพยากรณ์อนุกรมเวลาระหว่างวิธีแบบฉบับและวิธีบอซ-เจนกินส์ โดยใช้ข้อมูลสถิติจำนวนการเกิดอุบัติเหตุในประเทศไทย จากการศึกษาพบว่าข้อมูลดังกล่าวเป็นข้อมูลที่มีฤดูกาล การพยากรณ์โดยวิธีแบบฉบับได้สมการพยากรณ์คือ  $\hat{Y} = \hat{T} \times \hat{S}$  คำนวณค่ารากที่สองค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 229.5816 ครั้ง และตัวแบบพยากรณ์ที่เหมาะสมสำหรับการพยากรณ์โดยวิธีบอซ-เจนกินส์คือ  $ARIMA(0,1,3),(2,1,0)_{12}$  คำนวณค่ารากที่สองค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เท่ากับ 51.5996 ครั้ง ดังนั้นการพยากรณ์ข้อมูลอนุกรมเวลาโดยวิธีบอซ-เจนกินส์ จึงมีความเหมาะสมกับข้อมูลชุดดังกล่าว เนื่องจากให้ค่ารากที่สองค่าคลาดเคลื่อนกำลังสองเฉลี่ยน้อยกว่าการพยากรณ์โดยวิธีแบบฉบับ

ศิริเทพ จันทร์บุญแก้ว (2560) ศึกษาการพยากรณ์ปริมาณสายโทรเข้าสำหรับศูนย์บริการลูกค้า ธนาคารพาณิชย์ สำหรับการพยากรณ์รายเดือนโดยทำการเปรียบเทียบระหว่าง วิธีเฉลี่ยเคลื่อนที่อย่างง่าย (Simple Moving Average) วิธีปรับให้เรียบเอ็กซ์โปเนนเชียล (Exponential Smoothing Methods) วิธีการของบอซซ์-เจนกินส์ (ARIMA) และวิธีปัจจุบันของธนาคาร ผลการวิจัยพบว่าการพยากรณ์รายเดือน สำหรับกลุ่มบริการบัตรเครดิต กลุ่มบริการอิเล็กทรอนิกส์และบัตรเครดิตเงินสด กลุ่มบริการข้อมูลบัญชีเงินฝาก กลุ่มบริการอายุดับบัญชี และกลุ่มบริการสำหรับกลุ่มลูกค้าพิเศษผลการวิจัยพบว่า บอซซ์-เจนกินส์ให้ค่าพยากรณ์ที่ดีที่สุดเกือบทุกกรณี เมื่อวัดด้วยค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE)

กานต์สินี เจริญกิจวัชรชัย (2561) ศึกษาปัจจัยที่มีอิทธิพลในการกำหนดค่าจ้างแรงงานไทย โดยศึกษาจากข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน (SES) และข้อมูลการสำรวจภาวะการทำงานของประชากร (LFS) และสร้างสมการแบบจำลองพยากรณ์ค่าจ้างแรงงาน โดยใช้เทคนิควิธีการเชิงสถิติ และเทคนิควิธีการทางปัญญาประดิษฐ์อัลกอริทึมโครงข่ายประสาทเทียม และอัลกอริทึม Random Forest เพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์ ผลการศึกษาพบว่า ปัจจัยที่มีอิทธิพลต่อการกำหนดค่าจ้างแรงงานในช่วง 10 ปีที่ผ่านมา ประกอบไปด้วย 3 ปัจจัยหลัก คือปัจจัยด้านคุณลักษณะทั่วไปของแรงงาน ปัจจัยที่เกี่ยวข้องกับอาชีพและกิจการ และปัจจัยทางด้านภูมิศาสตร์ โดยผลจากการพยากรณ์ด้วยแบบจำลองพบว่า แบบจำลองอัลกอริทึม Random Forest มีความแม่นยำในการพยากรณ์มากกว่าแบบจำลองทางสถิติ และสามารถบอกถึงปัจจัยที่มีความสำคัญต่อการกำหนดค่าจ้างแรงงานได้เช่นเดียวกับแบบจำลองทางสถิติ ในขณะที่แบบจำลองอัลกอริทึมโครงข่ายประสาทเทียมนั้น ถึงแม้ว่ามีความแม่นยำสูงกว่าแบบจำลองทางสถิติ แต่ยังมีข้อจำกัดด้านความสามารถในการอธิบายผลของแบบจำลอง และพบว่ายังมีข้อจำกัดในการประมวลผลข้อมูลที่ประกอบด้วยปัจจัยที่อยู่ในรูปแบบข้อมูลเชิงตัวแปรหุ่น

บุญกอง ทะกลโยธิน และยุพาทรณ์ อารีพงษ์ (2561) ได้ศึกษาการเปรียบเทียบตัวแบบการพยากรณ์ราคาหุ้นโดยใช้แบบจำลองอาร์มาและอาร์แม็กซ์ โดยทำการเปรียบเทียบจากแบบจำลองอาร์มา (Autoregressive Integrated Moving Average: ARIMA) และแบบจำลองอาร์แม็กซ์ (Autoregressive Integrated Moving Average with Exogenous Variable: ARIMAX) ข้อมูลที่นำมาใช้เป็นข้อมูลราคาปิดของหลักทรัพย์ธนาคารกรุงเทพ จำกัด (มหาชน) รายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2555 ถึงเดือนธันวาคม พ.ศ. 2559 และมีตัวแปรอิสระที่นำมาศึกษาคือ อัตราการแลกเปลี่ยน ผลจากการวิจัยพบว่า

แบบจำลอง ARIMAX เหมาะสมกับข้อมูลชุดนี้มากที่สุด เนื่องจากให้ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำที่สุด

สำราญ วานนท์ และคณะ (2561) ศึกษาเทคนิคการพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้เทคนิคเหมืองข้อมูลโดยได้เปรียบเทียบผลการพยากรณ์ด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอมฟอเรส (Random Forest) และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิง (Bagging) ผลการวิจัยพบว่าเทคนิคแรนดอมฟอเรสให้ความถูกต้องในการจำแนกประเภทข้อมูลสูงที่สุด

ทัศนีย์ อัครพิณฑ์และจิรวุฒิ เขียวเกียรติประดับ (2563) ศึกษาการพยากรณ์ปริมาณการชำระเงินผ่านระบบอิเล็กทรอนิกส์แอปพลิเคชันโมบายแบงก์กิ้ง โดยใช้การรวบรวมข้อมูลปริมาณรายการการทางธุรกรรมการชำระเงินทางอิเล็กทรอนิกส์ผ่านแอปพลิเคชันโมบายแบงก์กิ้งของประเทศไทย ผู้วิจัยได้ทำการเปรียบเทียบ 2 วิธีการคือวิธีการพยากรณ์แบบปรับเรียบเอ็กโพเนนเชียล (Holt's linear trend.) กับวิธีการพยากรณ์บอซ-เจนกินส์ ( $ARIMA(2,1,0)(2,1,0)_{12}$ ) ผลการวิจัยพบว่า วิธีการพยากรณ์ด้วยวิธีบอซ-เจนกินส์ ตัวแบบ  $ARIMA(2,1,0)(2,1,0)_{12}$  คือตัวแบบที่เหมาะสมกับชุดข้อมูลมากที่สุด โดยมีค่าเฉลี่ยของร้อยละความผิดพลาดสัมบูรณ์ (Mean Absolute Percentage Error) หรือ MAPE มีค่าคิดเป็น 4.22 เปอร์เซ็นต์ที่ระดับความเชื่อมั่น 95 เปอร์เซ็นต์

วิชญ์วีร์ สุริยอมร และนุชนาถ คงช่วย (2563) ทำการศึกษาเรื่องการเปรียบเทียบตัวแบบทำนายเงินเดือนของบัณฑิตสาขาวิชาที่เกี่ยวข้องกับสถิติจากมหาวิทยาลัยในประเทศไทย โดยทำการเปรียบเทียบความแม่นยำในการทำนายของตัวแบบ Linear regression, Decision tree, Random forest และ K-Nearest neighbors โดยใช้ข้อมูลการมีงานทำของบัณฑิตในประเทศไทย ซึ่งมีการเก็บรวบรวมอย่างต่อเนื่องโดยกระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัย และนวัตกรรม จากสถาบันอุดมศึกษาทั่วประเทศไทยระหว่างปีการศึกษา 2556 ถึง 2560 และคัดกรองข้อมูลบัณฑิตที่จบจากสาขาวิชาที่เกี่ยวข้องกับสถิติ เช่น สถิติศาสตร์ สถิติประยุกต์ คณิตศาสตร์ประยุกต์ ผลการศึกษาพบว่าตัวแบบที่มีค่า MAPE ที่น้อยที่สุดคือตัวแบบ Random forest ดังนั้นจึงมีความแม่นยำในการทำนายมากที่สุด

## บทที่ 3

### วิธีดำเนินงานวิจัย

งานวิจัยนี้เป็นเปรียบเทียบประสิทธิภาพการพยากรณ์ระหว่างวิธีบอกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม กรณีศึกษาปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า เพื่อเป็นทางเลือก ในการนำไปประยุกต์ใช้กับการคาดการณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าในอนาคต โดยในบทนี้จะกล่าวถึงขั้นตอนการดำเนินงานวิจัย ข้อมูลที่นำมาใช้ การวิเคราะห์ข้อมูล และการวัดประสิทธิภาพของตัวแบบ ซึ่งมีรายละเอียดดังนี้

#### 3.1 ขั้นตอนการดำเนินงานวิจัย

ขั้นตอนการดำเนินงานวิจัยมีดังนี้

- 1) รวบรวมข้อมูลทุติยภูมิ
- 2) ศึกษาและวิเคราะห์หาตัวแบบ
- 3) ตรวจสอบความแม่นยำของตัวแบบ
- 4) สรุปผลการทดลอง

#### 3.2 ข้อมูลที่ใช้ในงานวิจัย

ข้อมูลที่นำมาใช้ในงานวิจัยนี้เป็นข้อมูลจริงของบริษัทแห่งหนึ่งซึ่งเป็นข้อมูลของลูกค้าที่โทรเข้าศูนย์บริการข้อมูลลูกค้าที่ถูกเก็บเป็นรายวันตั้งแต่วันที่ 1 เดือนกันยายน พ.ศ. 2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563 โดยข้อมูลแต่ละแถวคือ ปริมาณสายที่ลูกค้าโทรเข้ามาแบบรายวันประกอบด้วยข้อมูล 91 แถว และตัวแปรทั้งหมด 6 ตัว แบ่งเป็นตัวแปรเชิงคุณภาพ 2 ตัว คือวันที่และวันทำงานและตัวแปรเชิงปริมาณ 4 ตัว คือ ปริมาณสายที่โทรเข้า ปริมาณสายโทรเข้าที่พนักงานรับ เวลาสูงสุดที่การโทรอยู่ในคิวก่อนที่จะถูกวางสาย และความเร็วในการรับสายเฉลี่ย

โดยในชุดข้อมูลจะแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ชุดข้อมูลที่จะนำมาสอนระบบเพื่อสร้างตัวแบบจำลอง การทำนาย หรือ ข้อมูลฝึกสอน โดยใช้ข้อมูลในระหว่างวันที่ 1 เดือนกันยายน พ.ศ. 2563 ถึงวันที่ 2 พฤศจิกายน พ.ศ. 2563 และแบ่งข้อมูลที่เหลือในวันที่ 3 เดือนพฤศจิกายน พ.ศ. 2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563 สำหรับใช้เป็นข้อมูลในการทดสอบ หรือ ข้อมูลทดสอบแบบจำลองการพยากรณ์

### 3.3 สถิติที่ใช้ในการวิเคราะห์ข้อมูล

#### 3.3.1 สถิติสำหรับวิเคราะห์

ในการศึกษาครั้งนี้ได้ใช้แบบจำลอง วิธีบอกซ์-เจนกินส์, วิธี Random forestและวิธีโครงข่ายประสาทเทียม ในการหาตัวแบบพยากรณ์ที่เหมาะสมของแต่ละวิธี ซึ่งมีขั้นตอนการวิเคราะห์ดังนี้

##### 1) วิธีบอกซ์-เจนกินส์ มีขั้นตอนดังนี้

ขั้นตอนที่ 1 ตรวจสอบข้อมูลอนุกรมเวลา เพื่อพิจารณาการเป็นอนุกรมเวลาที่คงที่ด้วยการสร้างกราฟของข้อมูลอนุกรมเวลา

ขั้นตอนที่ 2 ในกรณีที่อนุกรมเวลาเป็นแบบไม่คงที่ มีแนวโน้ม หรือมีฤดูกาล ให้นำข้อมูลมาทำการกำจัดแนวโน้มและฤดูกาล (Differencing) เพื่อให้ข้อมูลอนุกรมเวลาเป็นแบบคงที่

ขั้นตอนที่ 3 ตรวจสอบฟังก์ชันความสัมพันธ์อัตโนมัติ (Autocorrelation Function: ACF) และฟังก์ชันความสัมพันธ์อัตโนมัติบางส่วน (Partial Autocorrelation Function: PACF) เพื่อดูว่าข้อมูลอนุกรมเวลานั้นมีลักษณะของ Autoregressive (AR) หรือ Moving Average (MA)

ในงานวิจัยครั้งนี้ การสร้างกราฟของ ACF และ PACF สำหรับอนุกรมเวลาที่ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า ประยุกต์ใช้ คำสั่ง `graphics.tsaplots` ในไลบรารี (Library) `statsmodels` ของโปรแกรม Python

ขั้นตอนที่ 4 กำหนดตัวแบบ ARIMAX(p,d,q) สามารถกำหนดได้จากกราฟ ACF และ PACF

ขั้นตอนที่ 5 ตรวจสอบความเหมาะสมของรูปแบบ

ก. การทดสอบว่าพารามิเตอร์แต่ละตัวในรูปแบบเท่ากับ 0 หรือไม่ โดยการใช้การทดสอบ t-statistic ภายใต้สมมติฐาน  $H_0: \theta = 0$  และ  $H_1: \theta \neq 0$  เมื่อกำหนดให้เป็นค่าพารามิเตอร์ ตัวทดสอบที่ใช้คือ

$$t = \frac{\theta}{S_{\hat{\theta}}}$$

โดยกำหนดให้

$\theta$  = ค่าประมาณของพารามิเตอร์

$S_{\hat{\theta}}$  = ค่าความคลาดเคลื่อนมาตรฐานของ  $\theta$

ข. การทดสอบว่าค่าความคลาดเคลื่อนของการพยากรณ์ที่อยู่ห่างกัน  $1, 2, \dots, m$  ช่วงเวลาเป็นอิสระกันหรือไม่ วิธีตรวจสอบความเหมาะสมของรูปแบบภายใต้สมมติฐาน  $H_0: \rho_1(e_t) = \dots = \rho_k(e_t) = 0$  โดยใช้ตัวทดสอบสถิติ Box - Ljung Chi -Square ( $Q$ ) เพื่อ

ตรวจสอบว่าความคลาดเคลื่อน ณ เวลา  $t$  คือ  $e_t$  โดย  $t=1,2,\dots,n$  มีความเป็นอิสระต่อกันหรือไม่ โดยเปรียบเทียบผลรวมของค่าสหสัมพันธ์ของ  $e_t$  ณ เวลาต่าง ๆ

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

โดยที่

$n$  = จำนวนค่าสังเกตในอนุกรมเวลา

$d$  = อันดับผลต่างของอนุกรมเวลาที่ทำให้อนุกรมเวลาเป็น Stationary

$r_k^2(e_t)$  = พิงก์ชันสหสัมพันธ์ระหว่างค่าความคลาดเคลื่อนที่อยู่ห่างกัน  $j$  ช่วงเวลา

$a$  = จำนวนพารามิเตอร์ทั้งหมดในรูปแบบ ซึ่งรวม  $\theta_0$  ด้วย

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

ซึ่ง  $Q$  มีการแจกแจงแบบ Chi-Square โดยประมาณการที่ระดับองศาแห่งความเป็นอิสระ (Degree of Freedom) เท่ากับ  $k - n_p$  ซึ่งจะยอมรับสมมติฐานหลัก เมื่อ  $Q$  มากกว่าหรือเท่ากับ  $\chi_{a,m-a}^2$  แสดงว่าไม่เป็นอิสระต่อกัน ดังนั้นรูปแบบที่ใช้ยังไม่เหมาะสมจึงต้องกลับไปพิจารณาหารูปแบบที่เหมาะสมใหม่ต่อไป

กรณีมีหลายตัวแบบผ่านการตรวจสอบความเหมาะสม จะใช้เกณฑ์การคัดเลือกตัวแบบ Akaike Information Criterion (AIC) โดยเลือกตัวแบบที่มีค่า AIC ต่ำที่สุด

ขั้นตอนที่ 6 นำแบบจำลอง ARIMAX ที่เลือกมาทดสอบความแม่นยำของแบบจำลอง (Testing model)

## 2) วิธีปาสุ่ม มีขั้นตอนดังนี้

ขั้นตอนที่ 1 ระบุตัวแปรตาม ( $y$ ) และตัวแปรอิสระ ( $x$ ) โดยตัวแปรตามคือ ปริมาณสายที่โทรเข้ามา และตัวแปรอิสระคือปริมาณสายโทรเข้าที่พนักงานรับ เวลาสูงสุดที่การโทรอยู่ในคิวก่อนที่จะถูกวางสาย ความเร็วในการรับสายเฉลี่ย และวันทำงาน

ขั้นตอนที่ 2 ทำการแบ่งชุดข้อมูลออกเป็นข้อมูลฝึกสอนและข้อมูลทดสอบ

ขั้นตอนที่ 3 กำหนดจำนวนต้นไม้ตัดสินใจที่ต้องการ

ขั้นตอนที่ 4 หาค่าพยากรณ์ โดยค่าพยากรณ์ที่ได้จะเป็นการให้ต้นไม้ตัดสินใจแต่ละต้น หาค่าพยากรณ์ของตัวเอง ในกรณีที่ปัญหาวิเคราะห์การถดถอย จะใช้วิธีคำนวณ หาค่าเฉลี่ยโดยนำเอาค่าพยากรณ์ของทุกต้นไม้ตัดสินใจมาคำนวณหาค่าเฉลี่ยเพื่อแสดง เป็นค่าพยากรณ์ของปัญหาวิธีป่าสุ่ม

โดยที่ตัวแบบป่าสุ่มสร้างโดยใช้คำสั่ง RandomForestRegressor ในไลบรารีsklearn.ensemble ของโปรแกรม Python จากนั้นจะทำการกำหนดค่าไฮเปอร์พารามิเตอร์ของตัวแบบป่าสุ่ม ที่จะใช้ทดลองสร้างตัวแบบดังแสดงในตาราง ซึ่งประกอบด้วยไฮเปอร์พารามิเตอร์ดังนี้

ตารางที่ 3.1 กำหนดค่าไฮเปอร์พารามิเตอร์สำหรับตัวแบบป่าสุ่ม

Hyperparameter	ค่ากำหนดที่ใช้
n estimator	50, 100, 200
min_samples_leaf	1, 3, 5
max features	0.5

1. n estimators คือ จำนวนต้นไม้ใน Random Forest
2. min\_samples\_leaf คือ ระบุจำนวนข้อมูลขั้นต่ำใน leaf node ของแต่ละต้นไม้ตัดสินใจ
3. max features คือ ระบุว่าแต่ละต้นไม้ตัดสินใจใน Random Forest จะสามารถสุ่มหยิบ feature ไปได้มากที่สุดกี่เปอร์เซ็นต์

3) วิธีโครงข่ายประสาทเทียม มีขั้นตอนดังนี้

ขั้นตอนที่ 1 ระบุตัวแปรตาม (y) และตัวแปรอิสระ (x) โดยตัวแปรตามคือ ปริมาณสายที่โทรเข้ามา และตัวแปรอิสระคือปริมาณสายโทรเข้าที่พนักงานรับ เวลาสูงสุดที่การโทรอยู่ในคิวก่อนที่จะถูกวางสาย ความเร็วในการรับสายเฉลี่ย และวันทำงาน

ขั้นตอนที่ 2 แบ่งชุดข้อมูลออกเป็นข้อมูลฝึกสอนและข้อมูลทดสอบ

ขั้นตอนที่ 3 กำหนดค่าพารามิเตอร์ของ จำนวนชั้นซ่อน จำนวนโหนดของชั้นข้อมูลเข้า จำนวนโหนดในชั้นซ่อน จำนวนโหนดในชั้นข้อมูลออก อัตราการเรียนรู้ และจำนวนรอบทำซ้ำ เป็นต้น

ขั้นตอนที่ 4 สร้างตัวแบบพยากรณ์จากข้อมูลฝึกสอน

ขั้นตอนที่ 5 พยากรณ์ข้อมูลทดสอบโดยใช้ตัวแบบพยากรณ์ที่สร้างขึ้น

โดยที่ตัวแบบโครงข่ายประสาทเทียม สร้างโดยใช้ คำสั่ง MLPRegressor ในไลบรารี sklearn.neural\_network ของโปรแกรม Python จากนั้นจะทำการกำหนดค่าไฮเปอร์พารามิเตอร์ของ ตัวแบบ MLP ที่จะใช้ทดลองสร้างตัวแบบดังแสดงในตาราง ซึ่งประกอบด้วยไฮเปอร์พารามิเตอร์ดังนี้ ตารางที่ 3.2 กำหนดค่าไฮเปอร์พารามิเตอร์สำหรับตัวแบบโครงข่ายประสาทเทียม

Hyperparameter	ค่ากำหนดที่ใช้
hidden_layer_sizes	10, 30, 50
activation	'relu'
learning_rate_init	0.01, 0.02
max_iter	1000, 3000, 5000

1. hidden\_layer\_sizes คือ จำนวนเซลล์ประสาทที่ซ่อนอยู่
2. activation คือ ฟังก์ชันกระตุ้นที่ใช้งานในชั้นซ่อนและชั้นข้อมูลออก
3. learning\_rate\_init คือ อัตราการเรียนรู้ของอัลกอริทึมที่ใช้ในการเรียนรู้
4. max\_iter คือ จำนวนรอบการทำซ้ำสูงสุด

### 3.3.2 สถิติสำหรับวัดประสิทธิภาพของตัวแบบพยากรณ์

- 1) ค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Mean Square Error : MSE)

$$MSE = \frac{\sum e_t^2}{n}$$

- 2) ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percent Error : MAPE)

$$MAPE = \frac{\sum \left| \frac{e_t}{Y_t} \right|}{n} \times 100$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.4 ขั้นตอนการทำ Dashbord

ในการสร้างแดชบอร์ดจะสร้างโดยใช้โปรแกรม Power bi

ขั้นตอนที่ 1 เตรียมข้อมูลเพื่อเตรียมนำเข้าโปรแกรม Power bi โดยใช้เมนู Get Data

ขั้นตอนที่ 2 เลือก Visual ที่ต้องการสร้าง โดยข้อมูลปริมาณสายโทรเข้าที่พนักงานรับ เวลาสูงสุด ที่การโทรอยู่ในคิวก่อนที่จะถูกวางสาย ความเร็วในการรับสายเฉลี่ย จะแสดงโดยใช้ Visual ชื่อ Card เพื่อแสดงเฉพาะตัวเลข จากนั้นเลือกข้อมูลโดยนำชื่อตัวแปรวางที่ช่อง Fields ส่วนตัวแปร Workdays ใช้ Pie Chart เนื่องจากต้องการแสดงว่าวันหยุดและวันทำงานมีปริมาณสายแตกต่างกันอย่างไร และตัวแปรปริมาณสายที่โทรเข้ามาแบ่งตามรายวันโดยใช้ Line Chart ในการแสดง

ขั้นตอนที่ 3 เมื่อทำการพยากรณ์ปริมาณสายโทรเข้าด้วยวิธีต่างๆแล้ว นำเข้าข้อมูลโดยใช้เมนู Enter Data เพื่อนำค่าพยากรณ์เข้าสู่โปรแกรม Power bi

ขั้นตอนที่ 4 นำค่าจริงของปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้ากับค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้ามาสร้างกราฟโดยใช้ Line Chart เพื่อเปรียบเทียบค่าจริงและค่าพยากรณ์เป็นรายวัน

## บทที่ 4

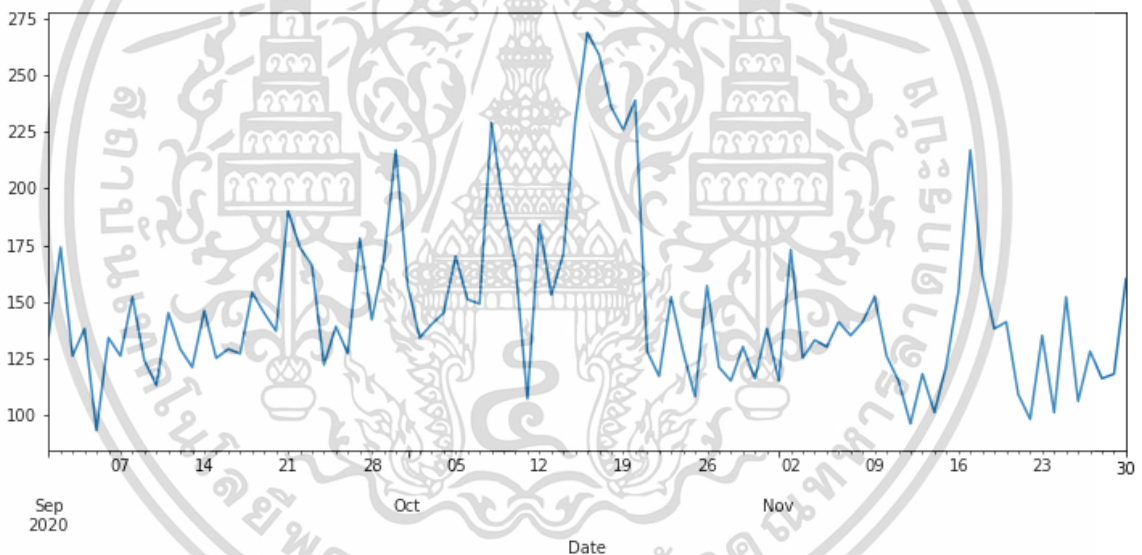
### ผลการวิจัย

ผลการวิจัยพบว่า การวัดประสิทธิภาพการพยากรณ์เพื่อใช้ในการทำนายปริมาณสายโทรเข้า ศูนย์บริการข้อมูลลูกค้า โดยใช้วิธีบอซซ์-เจนกินส์ วิธีป่าสุ่มและวิธีโครงข่ายประสาทเทียมกับการทดสอบ ชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ ตามขั้นตอนวิธีวิจัยสามารถแบ่งออกเป็น 6 ส่วนดังนี้

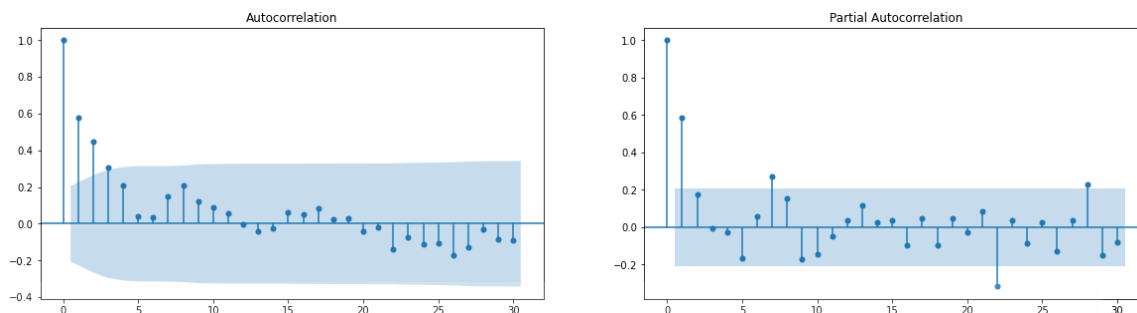
#### 4.1 วิธีบอซซ์-เจนกินส์

##### การตรวจสอบการเป็นอนุกรมเวลาที่เป็นสเตชันนารี

เมื่อนำข้อมูลปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้ารายวัน ในช่วงวันที่ 1 เดือนกันยายน พ.ศ. 2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563 มาสร้างกราฟอนุกรมเวลาแสดงดังรูปที่ 4.1



รูปที่ 4.1 การเคลื่อนไหวของอนุกรมเวลา



รูปที่ 4.2 คอเรลโรแกรมของ ACF และ PACF ของปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า

จากผลการทดสอบความนิ่งของข้อมูลโดยใช้ Unit Root Test พบว่า ค่าสถิติ ADF = -4.815 ซึ่งน้อยกว่าค่าวิกฤตที่ระดับนัยสำคัญ 0.05 และ p-value มีค่าน้อยกว่า 0.001 แสดงว่า ข้อมูลที่นำมาพิจารณาเป็นข้อมูลอนุกรมเวลาที่สเตชันนารี ซึ่งสามารถนำไปสร้างตัวแบบ ARIMAX ได้

#### การเปรียบเทียบประสิทธิภาพตัวแบบที่เหมาะสม

จากการทดลองสร้างตัวแบบ ARIMAX จากข้อมูลชุดฝึกสอนโดยเปลี่ยนค่าอันดับ p, q และใช้ MSE เป็นเกณฑ์ในการประเมินตัวแบบ ผลการทดลองใช้ตัวแบบภายใต้เงื่อนไขดังกล่าวพยากรณ์แสดงดังตารางที่ 4.1 เปรียบเทียบประสิทธิภาพตัวแบบบ็อกซ์-เจนกินส์

Model ARIMAX	MSE	AIC
ARIMAX(0,0,0)	47.044	304.587
ARIMAX(0,0,1)	49.216	305.749
ARIMAX(1,0,0)	49.195	305.912
ARIMAX(1,0,1)	50.772	307.745

ผลการทดลองพบว่าการใช้ข้อมูลเข้าเป็นชุดข้อมูลย้อนหลัง 1 วัน และกำหนดค่าอันดับ  $p = 0$ ,  $q = 0$  จะให้ค่า MSE ต่ำที่สุด คือ 47.044

#### การประมาณค่าพารามิเตอร์ตัวแบบ ARIMAX

จากการตรวจสอบค่า ACF ของอนุกรมเวลาของปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า ในรูปที่ 4.2 จะพบว่าอนุกรมเวลาเป็นแบบ Stationary และจากการเปรียบเทียบรูปแบบของ ACF และ PACF ของอนุกรมเวลากับลักษณะของ ACF และ PACF ตามทฤษฎี Box-Jenkins และจากการเปรียบเทียบประสิทธิภาพตัวแบบในตารางที่ 4.1 พบว่า ตัวแบบที่น่าจะเหมาะสมคือตัวแบบ ARIMAX(0,0,0)

และนำข้อมูลไปทำการหา Akaike's Information Criterion ของแต่ละอันดับ p,q พบว่า  $p = 0$  ,  $q = 0$  ให้ค่าที่ต่ำที่สุดคือ 304.587

ดังนั้นตัวแบบที่นำไปใช้ คือ ARIMAX(0,0,0) เมื่อนำตัวแบบไปประมาณค่าสัมประสิทธิ์ได้ผลลัพธ์ ดังนี้

ตารางที่ 4.2 ผลการทดสอบค่าพารามิเตอร์ของตัวแบบ ARIMAX(0,0,0)

ตัวแปร	Coef	SE Coef	t-value	p-value
Answered Calls ( $X_1$ )	0.9916	0.011	89.676	<0.001***
Average of Max Queue Time ( $X_2$ )	-0.1555	0.035	-4.427	<0.001***
Average Speed of Answer ( $X_3$ )	0.1535	0.012	12.873	<0.001***
Workdays ( $X_4$ )	0.2671	0.93	0.287	0.774
Lag Total Offered Calls ( $X_5$ )	-0.1902	0.17	-1.12	0.263
Lag Answered Calls ( $X_6$ )	0.2116	0.168	1.259	0.208
Lag Average of Max Queue Time ( $X_7$ )	-0.0579	0.061	-0.943	0.346
Lag Average Speed of Answer ( $X_8$ )	-0.0263	0.034	-0.78	0.436
Lag Workdays ( $X_9$ )	-1.7496	0.714	-2.45	0.014**

- หมายเหตุ: 1) \*\*\* หมายถึง ข้อมูลมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 99%  
 2) \*\* หมายถึง ข้อมูลมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 95%  
 3) \* หมายถึง ข้อมูลมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น 90%

จากตารางที่ 4.2 ผลการประมาณค่าพารามิเตอร์ของตัวแบบ ARIMAX(0,0,0) สามารถนำมาเขียนในรูปสมการได้ ดังนี้

$$\hat{Y}_t = 0.9916X_1 - 0.1555X_2 + 0.1535X_3 + 0.2671X_4 - 0.1902X_5 + 0.2116X_6 - 0.0579X_7 - 0.0263X_8 - 1.7496X_9 + e_t$$

### การตรวจสอบความเหมาะสมตัวแบบบอช-เจนกินส์

ก. การทดสอบว่าพารามิเตอร์แต่ละตัวในรูปแบบเท่ากับ 0 หรือไม่

สมมติฐานการทดสอบ

$$H_0 : X_1 = 0$$

$$H_1 : X_1 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_1$  น้อยกว่า  $\alpha = 0.05$  จึงปฏิเสธ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_1$  ในตัวแบบไม่เท่ากับ 0 นั่นคือ พารามิเตอร์  $X_1$  ควรมีในตัวแบบ

สมมติฐานการทดสอบ

$$H_0 : X_2 = 0$$

$$H_1 : X_2 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_2$  น้อยกว่า  $\alpha = 0.05$  จึงปฏิเสธ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_2$  ในตัวแบบไม่เท่ากับ 0 นั่นคือ พารามิเตอร์  $X_2$  ควรมีในตัวแบบ

สมมติฐานการทดสอบ

$$H_0 : X_3 = 0$$

$$H_1 : X_3 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_3$  น้อยกว่า  $\alpha = 0.05$  จึงปฏิเสธ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_3$  ในตัวแบบไม่เท่ากับ 0 นั่นคือ พารามิเตอร์  $X_3$  ควรมีในตัวแบบ

สมมติฐานการทดสอบ

$$H_0 : X_4 = 0$$

$$H_1 : X_4 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_4 = 0.774$  ซึ่งมากกว่า  $\alpha = 0.05$  จึงยอมรับ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_4$  ในตัวแบบเท่ากับ 0 นั่นคือ พารามิเตอร์  $X_4$  ไม่ควรมีในตัวแบบ

สมมติฐานการทดสอบ

$$H_0 : X_5 = 0$$

$$H_1 : X_5 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_5 = 0.263$  ซึ่งมากกว่า  $\alpha = 0.05$  จึงยอมรับ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_5$  ในตัวแบบเท่ากับ 0 นั่นคือ พารามิเตอร์  $X_5$  ไม่ควรมีในตัวแบบ

สมมติฐานการทดสอบ

$$H_0 : X_6 = 0$$

$$H_1 : X_6 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_6 = 0.208$  ซึ่งมากกว่า  $\alpha = 0.05$  จึงยอมรับ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_6$  ในตัวแบบเท่ากับ 0 นั่นคือ พารามิเตอร์  $X_6$  ไม่ควรมีในตัวแบบ

สมมติฐานการทดสอบ

$$H_0 : X_7 = 0$$

$$H_1 : X_7 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_7 = 0.346$  ซึ่งมากกว่า  $\alpha = 0.05$  จึงยอมรับ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_7$  ในตัวแบบเท่ากับ 0 นั่นคือ พารามิเตอร์  $X_7$  ไม่ควรมีในตัวแบบ

สมมติฐานการทดสอบ

$$H_0 : X_8 = 0$$

$$H_1 : X_8 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_8 = 0.436$  ซึ่งมากกว่า  $\alpha = 0.05$  จึงยอมรับ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_8$  ในตัวแบบเท่ากับ 0 นั่นคือ พารามิเตอร์  $X_8$  ไม่ควรมีในตัวแบบ

สมมติฐานการทดสอบ

$$H_0 : X_9 = 0$$

$$H_1 : X_9 \neq 0$$

จากตารางที่ 4.2 พบว่า p-value ของค่าพารามิเตอร์  $X_9$  น้อยกว่า  $\alpha = 0.05$  จึงปฏิเสธ  $H_0$  ที่ระดับนัยสำคัญ 0.05 สรุปว่าค่าพารามิเตอร์  $X_9$  ในตัวแบบไม่เท่ากับ 0 นั่นคือ พารามิเตอร์  $X_9$  ควรมีในตัวแบบ

ตารางที่ 4.3 ตรวจสอบตัวแบบ ARIMAX (0,0,0)

ตัวแบบที่เหมาะสม	ค่าสถิติ	
	Ljung-Box (Q)	Prob (Q)
ARIMAX(0,0,0)	0.14	0.71

ข. การทดสอบว่าค่าความคลาดเคลื่อนของการพยากรณ์ที่อยู่ห่างกัน  $1, 2, \dots, m$  ช่วงเวลาเป็นอิสระกันหรือไม่

สมมติฐานการทดสอบ

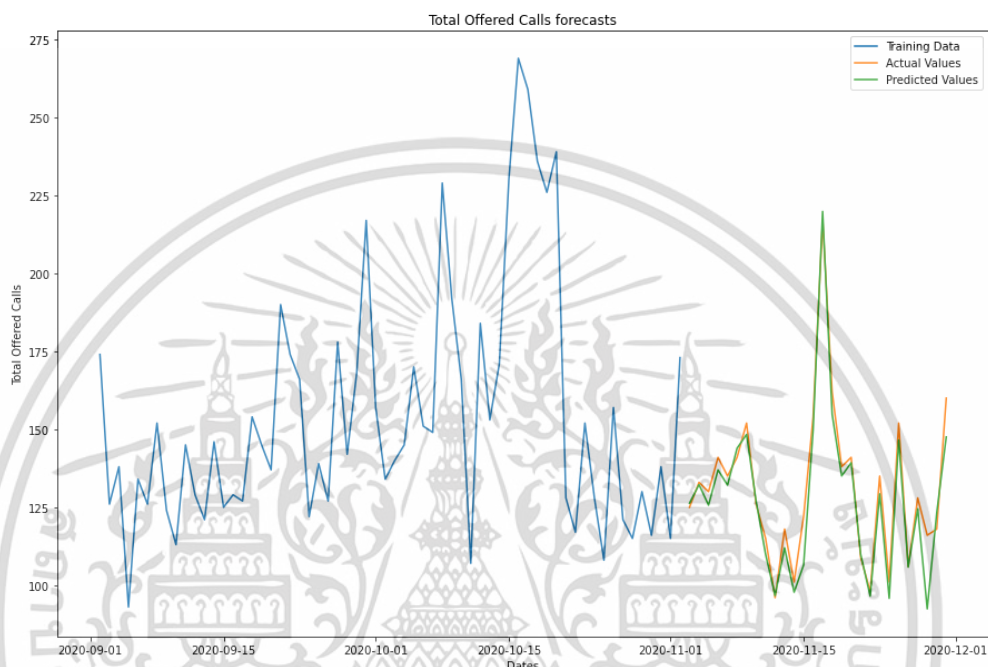
$$H_0 : \rho_1(e_t) = \dots = \rho_{63}(e_t) = 0$$

$$H_1 : \rho_k(e_t) \text{ อย่างน้อย 1 ค่า ไม่เท่ากับ 0 สำหรับ } k = 1, 2, 3, \dots, 63$$

จากตารางที่ 4.3 พิจารณาว่า p-value ของค่าสถิติ Ljung-Box พบว่า p-value = 0.72 ซึ่งน้อยกว่า  $\alpha = 0.05$  จึงปฏิเสธ  $H_0$  ที่ระดับนัยสำคัญ 0.05 นั่นคือ ค่าความคลาดเคลื่อนของการพยากรณ์ที่ห่างกัน  $1, 2, \dots, 63$  ช่วงเวลา มีการเคลื่อนไหวที่เป็นอิสระกัน สรุปได้ว่า ค่าความคลาดเคลื่อนของตัวแบบ ARIMAX(0,0,0) มีการเคลื่อนไหวที่เป็นอิสระกันและเป็นตัวแบบที่เหมาะสมกับอนุกรมเวลา

## การพยากรณ์ด้วยตัวแบบบ็อกซ์-เจนกินส์

เมื่อทำการพยากรณ์ 28 วัน เทียบกับค่าจริง ตั้งแต่วันที่ 3 เดือนพฤศจิกายน พ.ศ. 2563 ถึง วันที่ 30 เดือนพฤศจิกายน พ.ศ. 2563 ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 4.3



รูปที่ 4.3 กราฟแสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีบ็อกซ์-เจนกินส์

จากรูปที่ 4.3 พบว่าลักษณะจุดของค่าพยากรณ์มีลักษณะเกาะกันใกล้เคียงกับลักษณะข้อมูลปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า จากนั้นทำการคำนวณหาค่า MAPE ซึ่งจะได้  $MAPE = 3.681$  และได้ผลพยากรณ์ทั้ง 28 ค่า ดังนี้

ตารางที่ 4.4 ตารางแสดงค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีบอกซ์และเงินกินส์

วันที่	วิธีบอกซ์และเงินกินส์	วันที่	วิธีบอกซ์และเงินกินส์
03-11-20	126.3	17-11-20	219.9
04-11-20	132.3	18-11-20	154.8
05-11-20	125.7	19-11-20	135.0
06-11-20	137.0	20-11-20	139.3
07-11-20	132.1	21-11-20	109.9
08-11-20	143.9	22-11-20	96.5
09-11-20	148.3	23-11-20	129.3
10-11-20	127.0	24-11-20	95.8
11-11-20	109.9	25-11-20	146.6
12-11-20	96.7	26-11-20	105.8
13-11-20	112.0	27-11-20	124.5
14-11-20	97.7	28-11-20	92.4
15-11-20	106.5	29-11-20	123.3
16-11-20	149.8	30-11-20	147.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4.2 วิธีป่าสุ่ม

### การเปรียบเทียบประสิทธิภาพตัวแบบของวิธีป่าสุ่ม

จากการทดลองสร้างตัวแบบ Random forest จากข้อมูลชุดฝึกสอน โดยเปลี่ยนค่า Hyperparameter และใช้ MSE เป็นเกณฑ์ในการประเมินตัวแบบ ผลการทดลองใช้ตัวแบบภายใต้เงื่อนไขดังกล่าวพยากรณ์แสดงดังตารางที่ 4.5

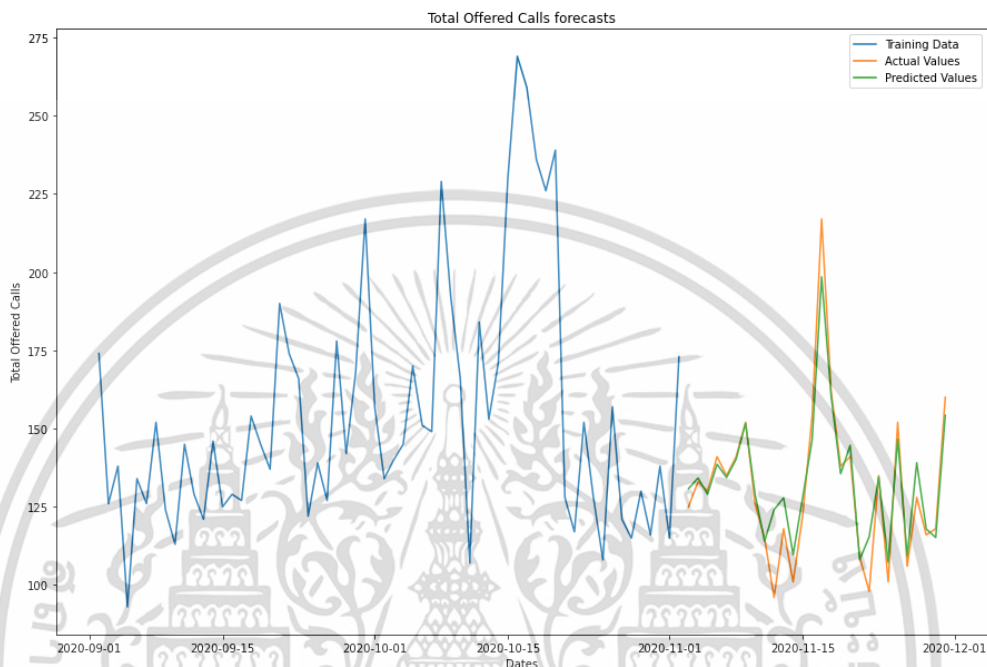
ตารางที่ 4.5 เปรียบเทียบประสิทธิภาพตัวแบบป่าสุ่ม

Hyperparameter			MSE
n estimators	min samples leaf	max features	
50	1	0.5	86.623
50	3	0.5	150.249
50	5	0.5	254.493
100	1	0.5	84.023
100	3	0.5	165.484
100	5	0.5	175.814
200	1	0.5	72.152*
200	3	0.5	142.162
200	5	0.5	247.173

ผลการทดลองพบว่าการใช้ข้อมูลเข้าเป็นชุดข้อมูลย้อนหลัง 1 วัน และกำหนดจำนวนต้นไม้เท่ากับ 200 ค่า min samples leaf เท่ากับ 1 และค่า max feature เท่ากับ 0.5 จะให้ค่า MSE ต่ำที่สุด คือ 72.152

## การพยากรณ์ด้วยตัวแบบป่าสุ่ม

เมื่อทำการพยากรณ์ 28 วัน เทียบกับค่าจริง ตั้งแต่วันที่ 3 เดือนพฤศจิกายน พ.ศ. 2563 ถึง วันที่ 30 เดือนพฤศจิกายน พ.ศ. 2563 ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 4.4



รูปที่ 4.4 กราฟแสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีป่าสุ่ม

จากรูปที่ 4.4 พบว่าลักษณะจุดของค่าพยากรณ์มีลักษณะเกาะกันใกล้เคียงกับลักษณะข้อมูลปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า จากนั้นทำการคำนวณหาค่า MAPE ซึ่งจะได้  $MAPE = 4.653$  และได้ผลพยากรณ์ทั้ง 28 ค่า ดังนี้

ตารางที่ 4.6 ตารางแสดงค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีป่าสุ่ม

วันที่	วิธีป่าสุ่ม	วันที่	วิธีป่าสุ่ม
03-11-20	130.8	17-11-20	198.5
04-11-20	134.2	18-11-20	160.3
05-11-20	129	19-11-20	135.6
06-11-20	138.6	20-11-20	144.7
07-11-20	134.4	21-11-20	108.1
08-11-20	140.2	22-11-20	115.4
09-11-20	151.7	23-11-20	134.5
10-11-20	128.9	24-11-20	107.2
11-11-20	114	25-11-20	146.6
12-11-20	124.1	26-11-20	109.3
13-11-20	127.9	27-11-20	139.1
14-11-20	109.7	28-11-20	117.9
15-11-20	127.8	29-11-20	115.2
16-11-20	146.4	30-11-20	154.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 4.3 วิธีโครงข่ายประสาทเทียม

#### การเปรียบเทียบประสิทธิภาพตัวแบบของวิธีโครงข่ายประสาทเทียม

จากการทดลองสร้างตัวแบบ MLP จากข้อมูลชุดฝึกสอนโดยเปลี่ยนค่า Hyperparameter และใช้ MSE เป็นเกณฑ์ในการประเมินตัวแบบ ผลการทดลองใช้ตัวแบบภายใต้เงื่อนไขดังกล่าวพยากรณ์แสดงดังตารางที่ 4.7

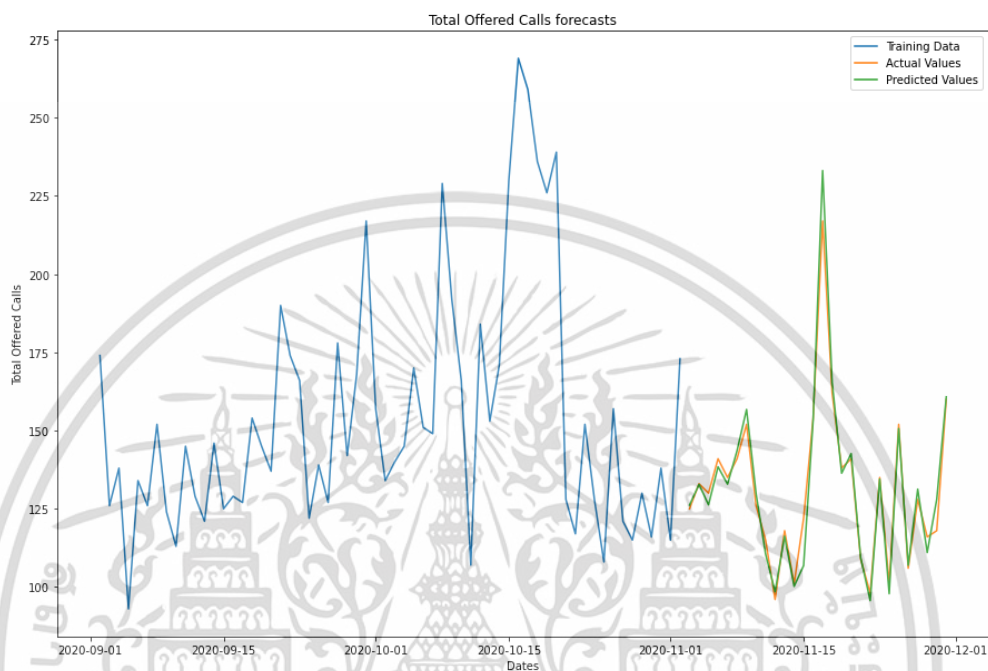
ตารางที่ 4.7 เปรียบเทียบประสิทธิภาพตัวแบบโครงข่ายประสาทเทียม

Hyperparameter				MSE
hidden layer	activation	learning rate	max iter	
10	relu	0.01	1000	5494.445
10	relu	0.01	3000	6331.795
10	relu	0.01	5000	5703.493
10	relu	0.02	1000	7473.528
10	relu	0.02	3000	7146.689
10	relu	0.02	5000	7811.383
50	relu	0.01	1000	6351.731
50	relu	0.01	3000	6052.290
50	relu	0.01	5000	5961.765
50	relu	0.02	1000	95.430
50	relu	0.02	3000	27.270*
50	relu	0.02	5000	57.157

ผลการทดลองพบว่าการใช้ข้อมูลเข้าเป็นชุดข้อมูลย้อนหลัง 1 วัน และกำหนดจำนวนชั้นซ่อน 50 โหนด อัตราการเรียนรู้เท่ากับ 0.02 และค่ามากที่สุด 3000 จะให้ค่า MSE ต่ำที่สุด คือ 27.270

### การพยากรณ์ด้วยตัวแบบโครงข่ายประสาทเทียม

เมื่อทำการพยากรณ์ 28 วัน เทียบกับค่าจริง ตั้งแต่วันที่ 3 เดือนพฤศจิกายน พ.ศ. 2563 ถึง วันที่ 30 เดือนพฤศจิกายน พ.ศ. 2563 ผลลัพธ์ของการพยากรณ์แสดงในรูปที่ 4.5



รูปที่ 4.5 กราฟแสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีโครงข่ายประสาทเทียม

จากรูปที่ 4.5 พบว่าลักษณะจุดของค่าพยากรณ์มีลักษณะเกาะกันใกล้เคียงกับลักษณะข้อมูลปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า จากนั้นทำการคำนวณหาค่า MAPE ซึ่งจะได้  $MAPE = 2.604$  และได้ผลพยากรณ์ทั้ง 28 ค่า ดังนี้

ตารางที่ 4.8 ตารางแสดงค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีโครงข่ายประสาทเทียม

วันที่	วิธีโครงข่ายประสาทเทียม	วันที่	วิธีโครงข่ายประสาทเทียม
03-11-20	126.2	17-11-20	233.1
04-11-20	132.8	18-11-20	165.5
05-11-20	126.2	19-11-20	136.4
06-11-20	138.4	20-11-20	142.6
07-11-20	132.9	21-11-20	109.4
08-11-20	143.6	22-11-20	95.6
09-11-20	156.8	23-11-20	134.5
10-11-20	129.9	24-11-20	97.8
11-11-20	110.6	25-11-20	150.7
12-11-20	98.5	26-11-20	106.9
13-11-20	116.2	27-11-20	131.3
14-11-20	100.2	28-11-20	111
15-11-20	106.9	29-11-20	128.1
16-11-20	153.4	30-11-20	160.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

#### 4.4 การเปรียบเทียบค่าจริงกับค่าพยากรณ์แบบรายวัน

ตารางที่ 4.9 ตารางแสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลกับค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลด้วยวิธีบอกซ์และเจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม

วันที่	ปริมาณสายโทรเข้า	ค่าพยากรณ์ปริมาณสายโทรเข้า		
		วิธีบอกซ์และเจนกินส์	วิธีป่าสุ่ม	วิธีโครงข่ายประสาทเทียม
03-11-20	125	126.3	130.8	126.2
04-11-20	133	132.3	134.2	132.8
05-11-20	130	125.7	129.0	126.2
06-11-20	141	137.0	138.6	138.4
07-11-20	135	132.1	134.4	132.9
08-11-20	141	143.9	140.2	143.6
09-11-20	152	148.3	151.7	156.8
10-11-20	126	127.0	128.9	129.9
11-11-20	115	109.9	114.0	110.6
12-11-20	96	96.7	124.1	98.5
13-11-20	118	112.0	127.9	116.2
14-11-20	101	97.7	109.7	100.2
15-11-20	122	106.5	127.8	106.9
16-11-20	154	149.8	146.4	153.4
17-11-20	217	219.9	198.5	233.1
18-11-20	162	154.8	160.3	165.5
19-11-20	138	135.0	135.6	136.4
20-11-20	141	139.3	144.7	142.6
21-11-20	109	109.9	108.1	109.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วันที่	ปริมาณสายโทรเข้า	ค่าพยากรณ์ปริมาณสายโทรเข้า		
		วิธีบอกซ์และเงินกินส์	วิธีป่าสุ่ม	วิธีโครงข่ายประสาทเทียม
22-11-20	98	96.5	115.4	95.6
23-11-20	135	129.3	134.5	134.5
24-11-20	101	95.8	107.2	97.8
25-11-20	152	146.6	146.6	150.7
26-11-20	106	105.8	109.3	106.9
27-11-20	128	124.5	139.1	131.3
28-11-20	116	92.4	117.9	111.0
29-11-20	118	123.3	115.2	128.1
30-11-20	160	147.6	154.3	160.8

#### 4.5 การเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์

ตารางที่ 4.10 เปรียบเทียบความแม่นยำระหว่างตัวแบบพยากรณ์

ตัวแบบ	MAPE
วิธีบอกซ์และเงินกินส์	3.721
วิธีป่าสุ่ม	4.653
วิธีโครงข่ายประสาทเทียม	2.604

จากตารางที่ 4.10 พบว่า ค่าพยากรณ์ที่ได้จากวิธีโครงข่ายประสาทเทียม ให้ค่า MAPE เท่ากับ 2.604 ซึ่งมีค่าความคลาดเคลื่อนน้อยที่สุด

#### 4.6 การเปรียบเทียบประสิทธิภาพของการพยากรณ์กับชุดข้อมูล

จากชุดข้อมูลที่ผู้วิจัยเลือกปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า ซึ่งเป็นชุดข้อมูลสำหรับการสร้างแบบจำลองการพยากรณ์ ผู้วิจัยนำข้อมูลในปี พ.ศ.2563 เป็นชุดข้อมูลทดสอบแบ่งเป็น ชุดข้อมูลย้อนหลัง (Lagged) ดังนี้ 1) 1 วัน 2) 2 วัน 3) 3 วัน 4) 4 วัน 5) 5 วัน 6) 6 วัน และ 7) 7 วัน นำมาสร้างตัวแบบในการพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า เพื่อวิเคราะห์ประสิทธิภาพที่ได้จาก

เทคนิควิธีทั้ง 3 วิธี โดยการเปรียบเทียบความคลาดเคลื่อนด้วยวิธีการ RMSE และ MAPE จากผลการทดลองสามารถแสดงได้ดังตารางที่ 4.11

ตารางที่ 4.11 การเปรียบเทียบประสิทธิภาพแบบจำลองการทำนายโดยใช้ชุดข้อมูลย้อนหลัง (Lagged)

Lagged (วัน)	วิธีบ็อกซ์และเจนกินส์		วิธีป่าสุ่ม		วิธีโครงข่ายประสาทเทียม	
	MSE	MAPE	MSE	MAPE	MSE	MAPE
1 วัน	44.502	3.687	97.756*	5.613*	27.270*	2.601*
2 วัน	29.818*	3.073*	149.997	7.739	87.577	3.763
3 วัน	32.570	3.321	181.290	8.784	64.527	3.802
4 วัน	30.976*	3.074	189.711	9.012	49.095	3.889
5 วัน	35.972	3.507	310.992	12.243	63.349	4.243
6 วัน	35.270	3.3465	350.037	12.973	48.435	4.059
7 วัน	34.498	3.375	347.643	13.028	86.194	4.058

จากตารางที่ 4.11 ผลการทำนายของแต่ละวิธีกับการใช้ชุดข้อมูล โดยมีจำนวนวันย้อนหลังที่แตกต่างกัน โดย MSE และ MAPE เป็นดัชนีชี้วัดประสิทธิภาพจากการทดลองแสดงให้เห็นว่า เมื่อสร้างตัวแบบด้วยวิธีบ็อกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม ผู้วิจัยพิจารณาแล้วพบว่า เมื่อสร้างข้อมูลย้อนหลังในเทคนิควิธีป่าสุ่มและวิธีโครงข่ายประสาทเทียม ด้วยชุดข้อมูลย้อนหลัง 1 วัน มีประสิทธิภาพสูงที่สุด ดังนี้ วิธีป่าสุ่ม 1 วัน มีค่า MSE เท่ากับ 97.756 ค่า MAPE เท่ากับ 5.613 และวิธีโครงข่ายประสาทเทียม 1 วัน มีค่า MSE เท่ากับ 27.270 ค่า MAPE เท่ากับ 2.601 ส่วนวิธีบ็อกซ์-เจนกินส์เมื่อสร้างด้วยข้อมูลย้อนหลัง 2 วัน มีประสิทธิภาพสูงที่สุด คือ มีค่า MSE เท่ากับ 29.818 ค่า MAPE เท่ากับ 3.073

จากผลการทดลอง ผู้วิจัยจึงเลือกชุดข้อมูลย้อนหลังที่ทำให้ได้ค่าประสิทธิภาพดีที่สุดคือ 1 วัน มาใช้สร้างตัวแบบเพื่อการพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าบนชุดข้อมูลทดสอบ ซึ่งเป็นปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าในช่วงวันที่ 1 เดือนกันยายน พ.ศ. 2563 ถึงวันที่ 2 เดือนพฤศจิกายน พ.ศ.2563

## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

การศึกษาเปรียบเทียบประสิทธิภาพการพยากรณ์ระหว่างวิธีบอกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม โดยมีวัตถุประสงค์เพื่อค้นหาตัวแบบที่มีความแม่นยำมากที่สุด สำหรับตัวแบบการพยากรณ์ หลังจากการศึกษาการสร้างตัวแบบและการพยากรณ์ด้วยตัวแบบที่สร้างขึ้นแล้วผู้วิจัยได้คำนวณค่า MAPE เพื่อทำการเปรียบเทียบผลการพยากรณ์ดังตารางที่ 4.9 และการเปรียบเทียบความแม่นยำของตัวแบบดังตารางที่ 4.10 โดยสรุปผลการวิจัยได้ดังนี้

#### 5.1 สรุปผลการวิจัย

การศึกษาเปรียบเทียบประสิทธิภาพการพยากรณ์ระหว่างวิธีบอกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม โดยรวบรวมข้อมูลปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้า ตั้งแต่วันที่ 1 เดือนกันยายน พ.ศ. 2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563 รวม 91 วัน นำผลการพยากรณ์ความต้องการมาเปรียบเทียบกัน พบว่า ค่าพยากรณ์ที่ได้จากวิธีโครงข่ายประสาทเทียม ให้ค่า MAPE เท่ากับ 2.604 ซึ่งมีค่าความคลาดเคลื่อนน้อยที่สุด

#### 5.2 อภิปรายผล

การพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีโครงข่ายประสาทเทียม ได้ผลลัพธ์การพยากรณ์ที่ใกล้เคียงกับค่าจริงมากกว่าการพยากรณ์รูปแบบอื่น ซึ่งสอดคล้องกับ B. Shamshad et al. (2019) ศึกษาการสร้างแบบจำลองและการพยากรณ์สภาพอากาศโดยใช้ ANN-MLP, ARIMA และแบบจำลอง ETS พบว่า ตัวแบบของ ANN-MLP ให้ค่าความแม่นยำกว่าตัวแบบอื่น ค่าพยากรณ์ที่ได้จากวิธีโครงข่ายประสาทเทียมให้ค่า MAPE ต่ำสุด แสดงว่าวิธีการพยากรณ์นี้เหมาะสมที่สุดกับข้อมูลปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าของกรณีศึกษาและสามารถนำมาใช้ได้จริงเพื่อช่วยให้การวางแผนจัดพนักงานรับสายให้มีประสิทธิภาพมากขึ้นและเป็นการลดค่าใช้จ่ายในการจ้างพนักงานรับสาย

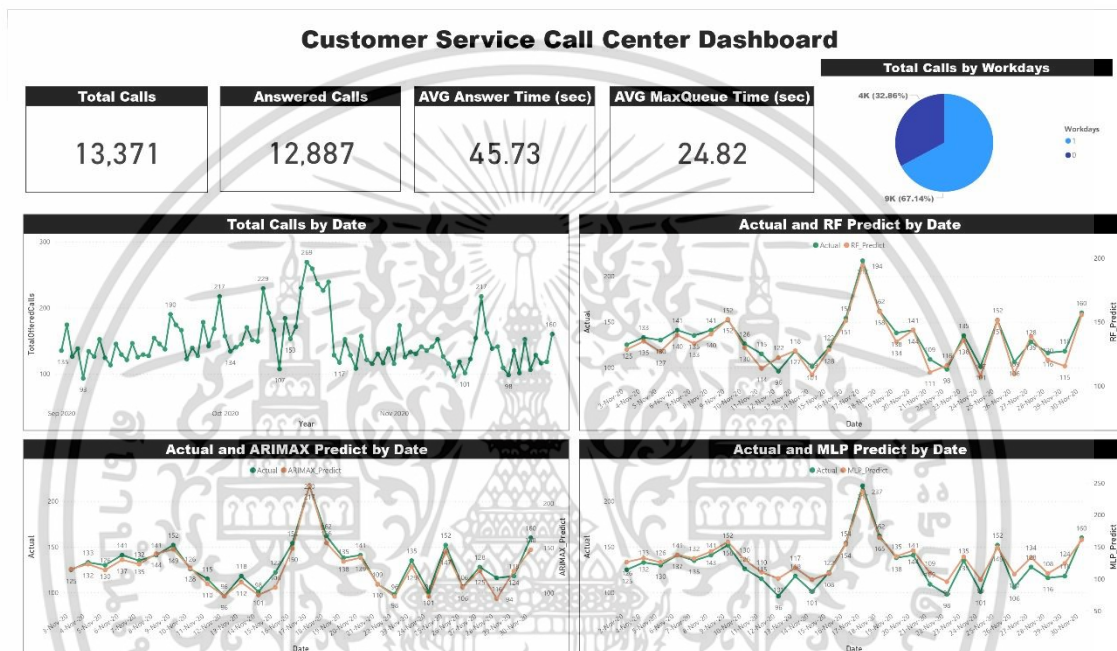
#### 5.3 ข้อเสนอแนะสำหรับการศึกษาครั้งต่อไป

1) การประยุกต์ใช้วิธีบอกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียม ในการพยากรณ์ปริมาณสายโทรเข้าอาจเปรียบเทียบกับการวิเคราะห์ทางเทคนิคอื่นๆ หรือหาตัวแปรภายนอกอื่นที่มีความสัมพันธ์กับปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้ามาคำนวณ เช่น Average Handling Time, Repeat Calls, Agent Utilization Rate เป็นต้น

2) ในวิธีบ็อกซ์-เจนกินส์ หากตัดตัวแปรที่มีค่า p-value น้อยกว่าระดับนัยสำคัญ 0.05 ออกจะทำให้ตัวแบบในวิธีนี้พยากรณ์ได้ดีขึ้น แต่เนื่องจากในงานวิจัยนี้ต้องการเปรียบเทียบตัวแบบพยากรณ์ระหว่าง 3 วิธีจึงไม่ตัดตัวแปรใดออกจากตัวแบบ

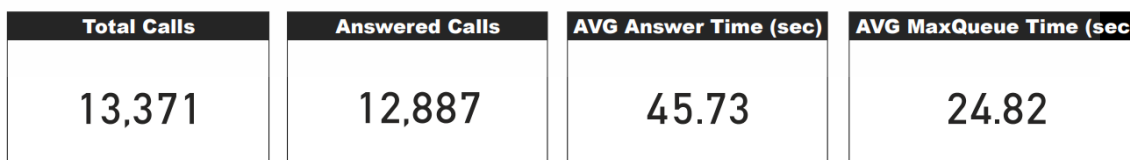
3) การประยุกต์ใช้วิธีป่าสุ่มอาจเพิ่มจำนวนต้นไม้ตัดสินใจ เพื่อให้ค่าพยากรณ์ที่ดีขึ้น

#### 5.4 Dashboard



รูปที่ 5.1 Dashboard แสดงภาพรวมของข้อมูลและการพยากรณ์

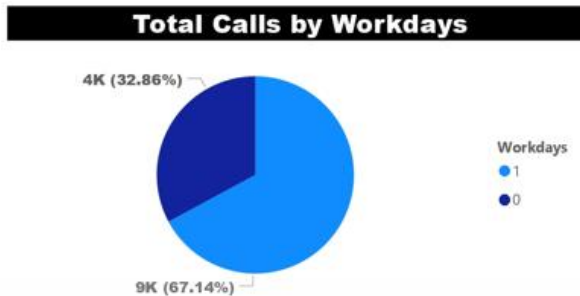
จากรูปที่ 5.1 แสดงจำนวนปริมาณสายที่โทรเข้ามาทั้งหมด ปริมาณสายที่พนักงานรับ เวลาเฉลี่ยในการรับสายและระยะเวลาที่ลูกค้ารอสายนานสุด กราฟเส้นแสดงปริมาณสายที่โทรเข้ามาแบบรายวัน รวมถึงค่าที่ได้จากการพยากรณ์ปริมาณสายที่โทรเข้าระหว่างวิธีบ็อกซ์-เจนกินส์ วิธีป่าสุ่ม และวิธีโครงข่ายประสาทเทียมแบบรายวัน



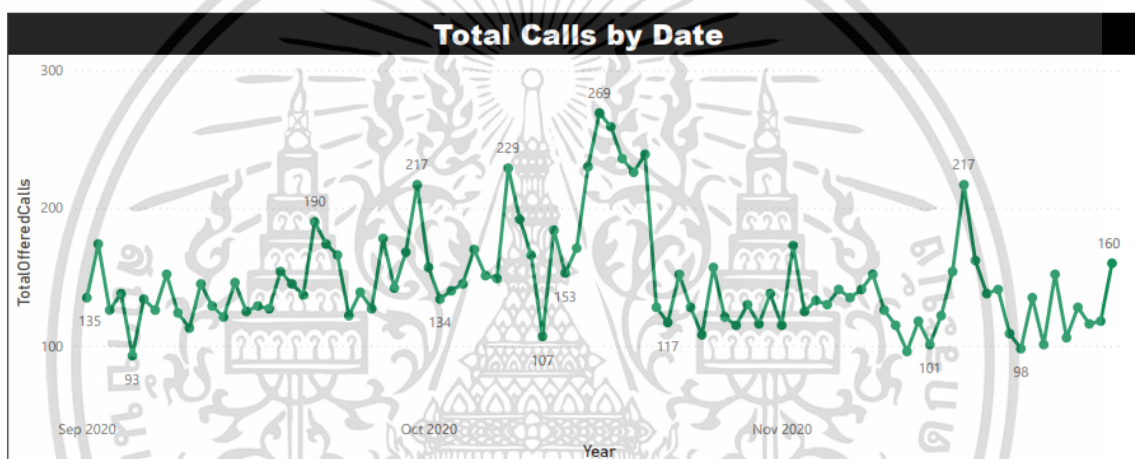
รูปที่ 5.2 Dashboard แสดงปริมาณสายโทรเข้าและระยะเวลา

จากรูปที่ 5.2 แสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าทั้งหมด 13,371 สาย และสายที่รับทัน 12,887 สาย ระยะเวลาที่ตอบรับสาย 45.73 วินาที ระยะเวลารอสายสูงสุด 24.82 วินาที

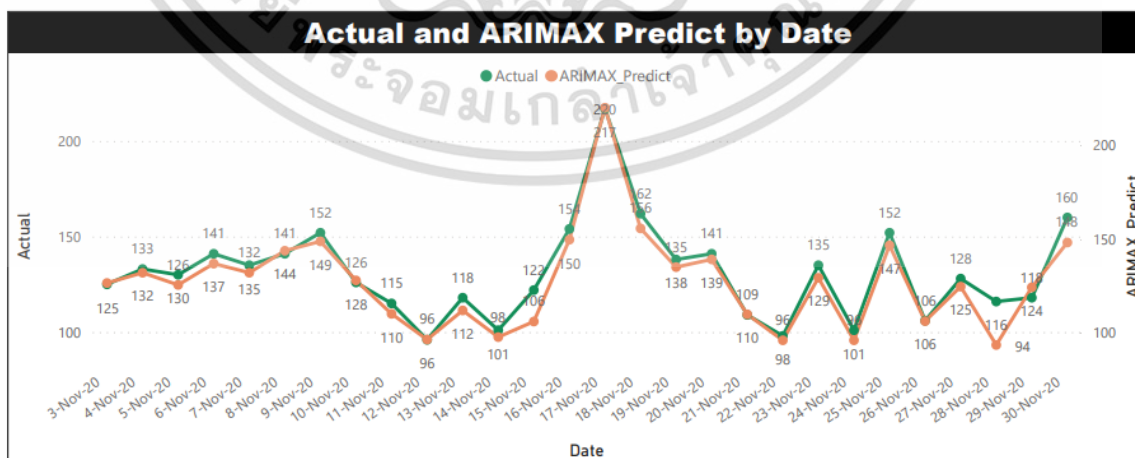
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.3 Dashboard แสดงปริมาณสายโทรเข้าแบ่งตามวันทำงาน  
จากรูปที่ 5.3 แสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าแบ่งตามวันทำงานและวันหยุด  
พบว่าวันทำงานมีปริมาณสายโทรเข้ามามากกว่าวันหยุด



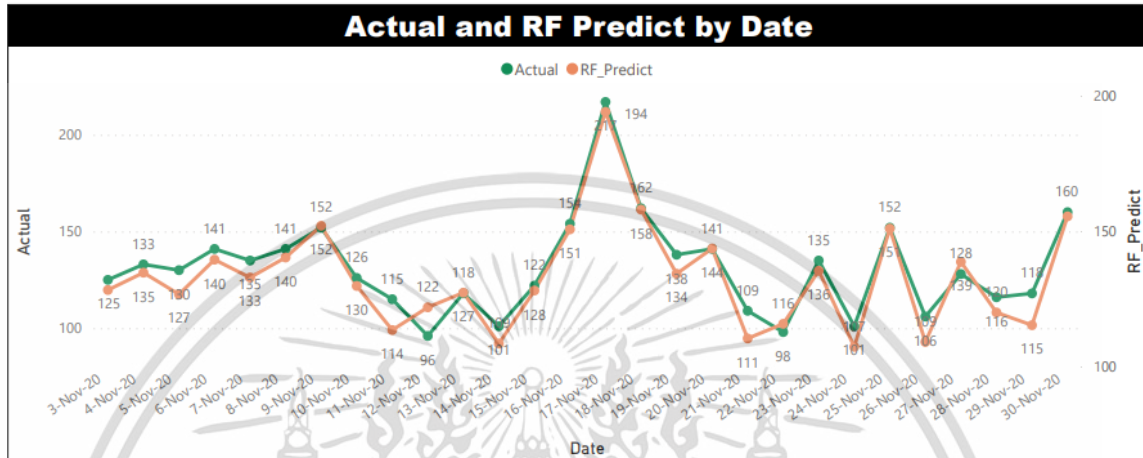
รูปที่ 5.4 Dashboard แสดงปริมาณสายโทรเข้า  
จากรูปที่ 5.4 แสดงจำนวนสายที่โทรเข้ามาแบ่งเป็นรายวันตั้งแต่วันที่ 1 เดือนกันยายน พ.ศ.  
2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563 รวม 61 วัน



รูปที่ 5.5 Dashboard แสดงปริมาณสายโทรเข้าและค่าพยากรณ์ด้วยวิธีบอกซ์-เจนกินส์

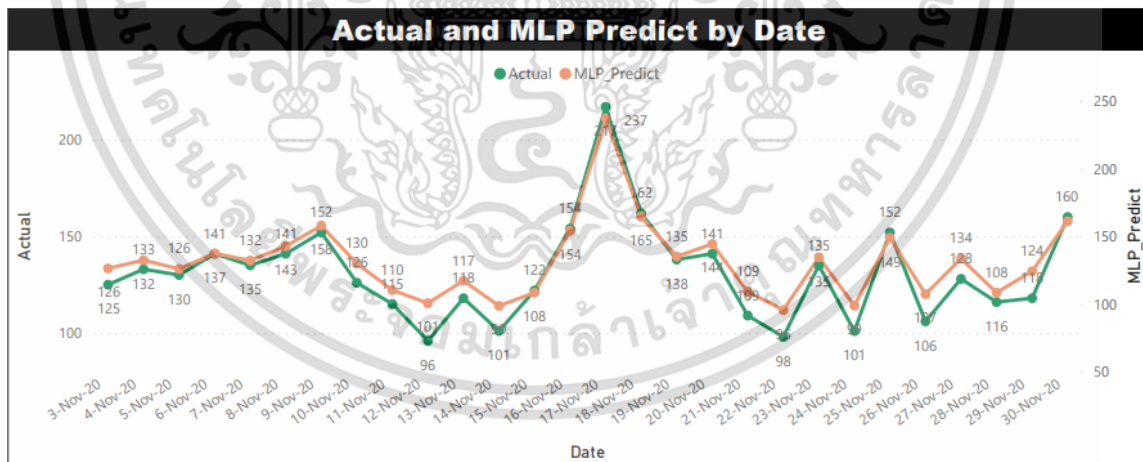
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 5.5 แสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีบ็อกซ์-เจนกินส์ ตั้งแต่วันที่ 3 เดือนพฤศจิกายน พ.ศ.2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563



รูปที่ 5.6 Dashboard แสดงปริมาณสายโทรเข้าและค่าพยากรณ์ด้วยวิธีป่าสุ่ม

จากรูปที่ 5.6 แสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีป่าสุ่ม ตั้งแต่วันที่ 3 เดือนพฤศจิกายน พ.ศ.2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563



รูปที่ 5.7 Dashboard แสดงปริมาณสายโทรเข้าและค่าพยากรณ์ด้วยวิธีโครงข่ายประสาทเทียม

จากรูปที่ 5.7 แสดงปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าและค่าพยากรณ์ปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าด้วยวิธีโครงข่ายประสาทเทียม ตั้งแต่วันที่ 3 เดือนพฤศจิกายน พ.ศ.2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง

- กานต์สินี เจริญกิจวัชรชัย. 2561. “ปัญญาประดิษฐ์กับการพยากรณ์ค่าจ้างแรงงานไทย.” วิทยานิพนธ์  
เศรษฐศาสตรมหาบัณฑิต มหาวิทยาลัยธรรมศาสตร์.
- กীরติ บุญโชติ. 2549. “การสกัดวลีสำคัญภาษาไทยด้วยโครงข่ายประสาทเทียม.”  
วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต มหาวิทยาลัยสงขลานครินทร์.
- จริยภรณ์ ไช้สิงห์ทอง และธงชัย อินตรา. 2561. “การประยุกต์ใช้แบบจำลองไบบารีเอทการช  
การศึกษาความสัมพันธ์ของความผันผวนระหว่างอัตราแลกเปลี่ยนกับดัชนีหลักทรัพย์กลุ่ม  
พลังงานและกลุ่มธนาคารในตลาดหลักทรัพย์แห่งประเทศไทย.” วิทยานิพนธ์เศรษฐศาสตร  
บัณฑิต มหาวิทยาลัยนเรศวร.
- ชนานุช จันทร์. 2552. “การทดสอบความสัมพันธ์ระหว่างอัตราแลกเปลี่ยนกับอัตราผลตอบแทนของ  
ดัชนีตลาดหลักทรัพย์แห่งประเทศไทย โดยวิธีไบบารีเอทการช.”  
วิทยานิพนธ์เศรษฐศาสตรมหาบัณฑิต มหาวิทยาลัยเชียงใหม่.
- ทัศนีย์ อัครพินท์และ จีรวุฒิ เขียวเกียรติประดับ. 2563. “การประยุกต์ใช้เทคนิคการวิเคราะห์อนุกรม  
เวลาเพื่อพยากรณ์ปริมาณการชำระเงินผ่านระบบอิเล็กทรอนิกส์ แอปพลิเคชันโมบาย  
แบงก์กิ้ง.” วารสารวิทยาการจัดการวไลยอลงกรณ์ปริทัศน์. 2563: 36-54.
- ธันท์ จรณะสมบุรณ์. 2561. “การทำนายการซื้อซ้ำของผู้ซื้อโดยใช้เทคนิคการเรียนรู้ของเครื่องจักร”  
วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒ.
- บุญกอง ทะกลโยธิน และยุพภรณ์อารีพงษ์. 2561. “การเปรียบเทียบตัวแบบการพยากรณ์ราคาหุ้นโดย  
ใช้แบบจำลองอาร์มาและอาร์แม็กซ์.” วารสารวิทยาศาสตร์และเทคโนโลยีหัวเฉียว  
เฉลิมพระเกียรติ. 2561: 44-55.
- ปฎิมาพร อุดม. 2556. “การพยากรณ์ยอดขายสำหรับผู้แทนจำหน่ายเม็ดพลาสติก.” วิทยานิพนธ์  
วิศวกรรมศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- พัชรินทร์ พันธุ์พานิชย์. 2559. “การศึกษาการส่งมอบงานจากบันทึกเหตุการณ์ในศูนย์บริการข้อมูลลูกค้า  
โดยใช้ Process Mining รูปแบบ Social Network Miner (Handover of Work)”  
วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต มหาวิทยาลัยสยาม.

- รักษนก นุชพงษ์. 2540. “ความสัมพันธ์ระหว่างอัตราแลกเปลี่ยนและส่วนต่างของอัตราดอกเบี้ยที่แท้จริง.” วิทยานิพนธ์เศรษฐศาสตร์มหาบัณฑิต มหาวิทยาลัยธรรมศาสตร์.
- วราพร งามสุข. 2555. “การพยากรณ์อนุกรมเวลาโดยการเปรียบเทียบแบบฉบับและวิธีบอกซ์-เจนกินส์กรณีศึกษาจำนวนการเกิดอุบัติเหตุของประเทศไทย.” วิทยานิพนธ์วิทยาศาสตรบัณฑิต สาขาวิชาสถิติ มหาวิทยาลัยบูรพา.
- วิชญ์วีร์ สุริยมร และนุชนาด คงช่วย. 2563. “การเปรียบเทียบตัวแบบทำนายเงินเดือน ของบัณฑิตสาขาวิชาที่เกี่ยวข้องกับสถิติจากมหาวิทยาลัยในประเทศไทย.” **การประชุมนำเสนอผลงานวิจัยระดับบัณฑิตศึกษา(ครั้งที่ 15).** 2563 (15): 2338-2349.
- ศิริเทพ จันทน์บุญแก้ว. 2560. “การพยากรณ์ปริมาณสายโทรศัพท์เข้าสำหรับศูนย์บริการลูกค้าธนาคารพาณิชย์.” วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- สำราญ วานนท์ ธีรัช อารีราษฎร์ และจรัญ แสนราช. 2561. “การศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์ โดยใช้เทคนิคเหมืองข้อมูล.” *วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม.* 2561(1): 164-171.
- Albrecht, T., Rausch, M. T. & Derra, D. N. 2021. “Call Me Maybe: Methods and Practical Implementation of Artificial Intelligence in Call Center Arrivals’ Forecasting.” *Journal of Business Research*, 120, 267-278.
- Anggraeni, W., Vinarti, R. A. & Kurniawati, Y. D. 2015. “Performance Comparisons between Arima and Arimax Method in Moslem Kids Clothes Demand Forecasting: Case Study.” *Procedia Computer Science*, 72, 630-637.
- B. Shamshad, M. Z Khan & Z. Omar. 2019 “Modeling and forecasting weather Parameters using ANN-MLP, ARIMA and ETS model: a case study for Lahore, Pakistan”, *Journal of Applied Statistics* 5(2019). 388.
- Dickey, D.A., Fuller, W.A. 1981: “Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root.” *Econometrica* 49: 1057–1072.
- Hamilton, J. 1994. “Time Series Analysis Princeton University Press Princeton.” Princeton University Press: Princeton, NJ, USA.

- Jing, L., and G. Min. 2010. "Predicting Call Center Service Grade with Improved Neural Network Algorithm." *Intelligent Systems and Applications*. 1-4.
- Kubat, M. 1999. "Neural Networks: A Comprehensive Foundation by Simon Haykin." *Macmillan*. 1994 : 409-412.
- Livingstone, D.J. 2008. "Artificial neural networks methods and applications." Humana Press.
- L.Breiman and R. Forests. 2001. Machine Learning. [Online]. Retrieved from <http://dx.doi.org/10.1023%2FA%3A1010933404324>
- Noureen, S., Atique, S., Roy, V. and Bayne, S. 2019. "A Comparative Forecasting Analysis Of Arima Model Vs Random Forest Algorithm for a Case Study of Small-Scale Industrial Load," *International Research Journal of Engineering and Technology*. 1812–1821.
- Sen, P., Roy, M. & Pal, P. 2016. "Application of ARIMA for Forecasting Energy Consumption and GHG Emission: A Case Study of an Indian Pig Iron Manufacturing Organization." *Energy*, 116, 1031-1038.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### ภาคผนวก ก

ข้อมูลปริมาณสายโทรเข้าศูนย์บริการข้อมูลลูกค้าของบริษัทแห่งหนึ่ง ตั้งแต่วันที่ 1 เดือนกันยายน พ.ศ. 2563 ถึงวันที่ 30 เดือนพฤศจิกายน พ.ศ.2563

Date	Total Offered Calls	Answered Calls	Average of Max Queue Time	Average Speed of Answer	Workdays
1-Sep-20	135	127	30.409	74.955	1
2-Sep-20	174	171	42.138	48.103	1
3-Sep-20	126	125	23.517	24.586	1
4-Sep-20	138	131	62.069	110.345	0
5-Sep-20	93	91	13.552	21.828	0
6-Sep-20	134	130	12.207	20.690	0
7-Sep-20	126	119	20.481	27.037	0
8-Sep-20	152	148	14.172	44.586	1
9-Sep-20	124	119	22.483	25.552	1
10-Sep-20	113	111	10.138	13.621	1
11-Sep-20	145	144	21.172	37.793	1
12-Sep-20	129	124	13.310	27.483	0
13-Sep-20	121	119	3.586	6.966	0
14-Sep-20	146	143	13.379	26.345	1
15-Sep-20	125	125	25.483	51.414	1
16-Sep-20	129	124	10.483	17.828	1
17-Sep-20	127	127	9.138	14.103	1
18-Sep-20	154	153	9.931	17.276	1
19-Sep-20	145	142	8.345	13.103	0
20-Sep-20	137	136	6.586	12.483	0
21-Sep-20	190	188	2.000	7.690	1
22-Sep-20	174	168	34.931	60.586	1
23-Sep-20	166	163	16.000	25.241	1
24-Sep-20	122	120	2.207	5.552	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Date	Total Offered Calls	Answered Calls	Average of Max Queue Time	Average Speed of Answer	Workdays
25-Sep-20	139	137	9.931	17.103	1
26-Sep-20	127	125	6.655	14.414	0
27-Sep-20	178	178	8.862	17.448	0
28-Sep-20	142	142	13.000	35.655	1
29-Sep-20	168	168	15.483	31.138	1
30-Sep-20	217	214	12.552	23.517	1
1-Oct-20	157	154	9.138	15.483	1
2-Oct-20	134	132	11.966	16.828	1
3-Oct-20	140	138	23.655	40.862	0
4-Oct-20	145	142	26.034	39.621	0
5-Oct-20	170	163	15.862	30.034	1
6-Oct-20	151	148	17.000	40.241	1
7-Oct-20	149	147	7.690	14.862	1
8-Oct-20	229	218	22.276	119.138	1
9-Oct-20	192	190	23.966	41.241	1
10-Oct-20	166	165	9.069	23.207	0
11-Oct-20	107	105	1.862	4.586	0
12-Oct-20	184	184	14.621	23.724	1
13-Oct-20	153	152	3.828	8.897	0
14-Oct-20	171	168	21.828	32.138	1
15-Oct-20	230	223	7.655	20.483	1
16-Oct-20	269	261	16.621	59.552	1
17-Oct-20	259	254	35.069	68.586	0
18-Oct-20	236	205	23.345	142.517	0
19-Oct-20	226	219	28.828	112.759	1
20-Oct-20	239	224	30.241	141.241	1
21-Oct-20	128	126	15.069	35.379	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Date	Total Offered Calls	Answered Calls	Average of Max Queue Time	Average Speed of Answer	Workdays
22-Oct-20	117	117	17.000	26.724	1
23-Oct-20	152	149	14.552	34.000	0
24-Oct-20	128	126	16.310	22.448	0
25-Oct-20	108	104	8.690	13.448	0
26-Oct-20	157	152	24.034	39.138	1
27-Oct-20	121	120	21.828	32.414	1
28-Oct-20	115	112	7.759	12.414	1
29-Oct-20	130	128	9.000	18.552	1
30-Oct-20	116	112	3.966	5.690	1
31-Oct-20	138	133	31.069	84.586	0
1-Nov-20	115	114	12.207	18.207	0
2-Nov-20	173	163	43.621	70.345	1
3-Nov-20	125	123	8.517	20.517	1
4-Nov-20	133	131	8.621	17.241	1
5-Nov-20	130	125	8.345	11.034	1
6-Nov-20	141	137	16.138	18.759	1
7-Nov-20	135	130	19.724	30.897	0
8-Nov-20	141	138	18.138	45.379	0
9-Nov-20	152	143	81.517	104.345	1
10-Nov-20	126	120	47.172	67.172	1
11-Nov-20	115	108	16.897	17.621	1
12-Nov-20	96	94	28.897	47.517	1
13-Nov-20	118	108	54.276	75.759	1
14-Nov-20	101	97	17.655	15.310	0
15-Nov-20	122	98	46.690	90.655	0
16-Nov-20	154	142	46.103	105.483	1
17-Nov-20	217	189	145.586	352.759	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Date	Total Offered Calls	Answered Calls	Average of Max Queue Time	Average Speed of Answer	Workdays
18-Nov-20	162	150	33.966	59.966	1
19-Nov-20	138	133	21.793	36.759	0
20-Nov-20	141	127	65.655	133.586	0
21-Nov-20	109	106	17.103	29.586	0
22-Nov-20	98	94	19.172	23.966	0
23-Nov-20	135	124	55.690	79.759	1
24-Nov-20	101	94	19.621	20.862	1
25-Nov-20	152	143	55.483	83.621	1
26-Nov-20	106	104	20.345	17.966	1
27-Nov-20	128	116	78.310	129.069	1
28-Nov-20	116	96	112.310	78.069	0
29-Nov-20	118	115	17.586	31.379	0
30-Nov-20	160	142	107.172	132.793	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้