The Journal of Applied Science **Vol. 21 No. 1 [2022]: 244662**
วารสารวิทยาศาสตร์ประยุกต์ doi: 10.14416/j.appsci.2022.01.016
ISSN 1513-7805 Printed in Thailand ISSN 2586-9663 (Online)

Research Article

การเปรียบเทียบตัวแบบพยากรณ์อนุกรมเวลาระหว่างตัวแบบผสมกับตัวแบบเดี่ยวเพื่อการพยากรณ์จำนวนสายโทรเข้ารายวัน

A Comparison of Time Series Forecasting Models between Hybrid Model and Individual Model for Forecasting Daily Incoming Call Volume

พรพิมล ชัยวุฒิศักดิ์[1]*

Pornpimol Chaiwuttisak [1]*

[1]ภาควิชาสถิติ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพ 10520 ประเทศไทย

[1]Departmen t of Statistics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

***E-mail**: pornpimol.ch@kmitl.ac.th

Received: 27/06/2021; Revised: 19/12/2021; Accepted: 28/03/2022

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความถูกต้องของวิธีการพยากรณ์จำนวนสายโทรเข้ารายวัน โดยวิธีการพยากรณ์ที่ทำการศึกษามี 3 วิธี ได้แก่ 1) วิธีบ็อกซ์-เจนกินส์ประกอบด้วยตัวแบบ SARIMA และตัวแบบ SARIMAX 2) วิธีโครงข่ายประสาทเทียม และ 3) วิธีผสมระหว่างตัวแบบ SARIMAX และโครงข่ายประสาทเทียม ข้อมูลที่นำมาใช้ในการศึกษาเป็นข้อมูลอนุกรมเวลารายวันของจำนวนสายโทรเข้าไปที่ศูนย์บริการลูกค้าของบริษัทแห่งหนึ่ง ซึ่งแบ่งข้อมูลเป็น 2 ชุด คือ ข้อมูลในอดีตตั้งแต่เดือนมกราคม พ.ศ. 2559 ถึง เดือนธันวาคม พ.ศ. 2561 เพื่อคัดเลือกตัวแบบที่เหมาะสมที่สุด และ ข้อมูลในอดีตตั้งแต่เดือนมกราคม พ.ศ. 2562 ถึง เดือนธันวาคม พ.ศ. 2562 เพื่อเปรียบเทียบความแม่นยำของค่าพยากรณ์ที่ได้ โดยเกณฑ์ที่ใช้ในการเปรียบเทียบความคลาดเคลื่อนของการพยากรณ์คือ ค่าเฉลี่ยของค่าสัมบูรณ์เปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) ผลการศึกษาพบว่า ตัวแบบที่มีค่า MAPE ต่ำสุดคือตัวแบบผสมระหว่าง SARIMAX และ โครงข่ายประสามเทียม ซึ่งเท่ากับ 24.02% ซึ่งน้อยกว่าตัวแบบ SARIMAX ตัวแบบโครงข่ายประสาทเทียม และตัวแบบ SARIMA ซึ่งมีค่า MAPE เท่ากับ 24.06%, 43.70%,

44.97% ตามลำดับ จะเห็นได้ว่าตัวแบบผสมมีความแม่นยำในการพยากรณ์มากกว่าตัวแบบเดี่ยว โดยตัวแบบผสมที่ได้สามารถนำไปพยากรณ์จำนวนสายโทรเข้ารายวันล่วงหน้าสำหรับใช้เป็นข้อมูลประกอบการวางแผนการกำลังคนทำงานของศูนย์บริการข้อมูลลูกค้าที่เหมาะสมในอนาคต

คำสำคัญ การพยากรณ์, จำนวนสายโทรเข้ารายวัน, วิธีบ็อกซ์-เจนกินส์, โครงข่ายประสาทเทียม, ตัวแบบผสม

Abstract

The research aimed to compare the accuracy of forecasting methods for daily incoming call volume. There were three forecasting methods that were used in the study: Box-Jenkins method with SRIMA model and SARIMAX model, Artificial Neural Network (ANN) model and the hybrid model combining SARIMAX and Artificial Neural Network (SARIMAX-ANN) model. The data used in this study is time series of daily incoming call volume to call center which can be divided into 2 data sets. The first data set which was the past data from January 2016 to December 2018 were used for selecting of the most suitable model and the second data set was the past data from January 2019 to December 2019 for the comparison of the accuracy of forecasting model by using Mean Absolute Percentage Error (MAPE). The results showed model with the lowest MAPE is hybrid model of SARIMAX-ANN (MAPE = 24.02%), while the MAPE values for SARIMAX, ANN and SARIMA were 24.06%, 43.70%, and 44.97% respectively. It indicates that the hybrid model is more accurate in forecasting than individual model. The hybrid model can be used to forecast daily incoming call volume which is supporting information for the suitable workforce planning of customer service center in the future.

Keywords: Forecasting, Daily Incoming Call Volume, Box-Jenkins, Artificial Neural Network, Hybrid model.

บทนำ

ปัจจุบันการดำเนินธุรกิจต่างๆ มักเน้นการสร้างความพึงพอใจให้แก่ลูกค้า โดยแนวทางหนึ่งในการสร้างความพึงพอใจคือการที่ลูกค้าสามารถติดต่อกับหน่วยงานหรือองค์กรได้ทุกที่ ทุกเวลา ดังนั้นการจัดตั้งศูนย์บริการข้อมูลลูกค้า หรือศูนย์ลูกค้าสัมพันธ์ ที่เรียกกันโดยทั่วไปว่า Call Center จะถูกใช้เป็นหนึ่งในช่องทางการตอบสนองความต้องการของลูกค้าในการติดต่อเข้ามาสอบถามข้อมูล ไม่ว่าจะเป็นเรื่องของผลิตภัณฑ์ หรือเมื่อต้องการความช่วยเหลือในเรื่องต่างๆ เพื่อทำให้ลูกค้าเกิดความพึงพอใจสูงสุดในบริการที่ได้รับ

บริษัทแห่งหนึ่งในธุรกิจประกันชีวิตมีศูนย์บริการข้อมูลลูกค้า (Call Center) ที่ถูกจัดตั้งขึ้นเพื่อเป็นการให้ข้อมูลในด้านต่างๆ แก่ลูกค้า เช่น ด้านผลิตภัณฑ์ หรือ ด้านการใช้งานผลิตภัณฑ์ แนะนำสินค้าและบริการ ประสานงานและช่วยเหลือในการแก้ไขปัญหาที่เกิดขึ้นของผู้ใช้บริการ เพื่อให้ตรงตามความต้องการและความพึงพอใจของลูกค้า โดยมีการให้บริการตลอด 24 ชั่วโมง ทั้งนี้ทางศูนย์บริการข้อมูลลูกค้ามีความประสงค์ที่จะพัฒนาและเพิ่มประสิทธิภาพในการให้บริการแก่ลูกค้า โดยอาศัยตัวแบบพยากรณ์จำนวนสายโทรเข้ารายวันสำหรับเป็นเครื่องมือหนึ่งที่ช่วยในการวางแผนการปฏิบัติงานในด้านต่างๆ ในอนาคต รวมทั้งลดระยะเวลารอคอยในการรับบริการของลูกค้าได้ ซึ่งเป็นการส่งเสริมภาพลักษณ์และความเชื่อมั่นของลูกค้าที่มีต่อบริษัท (Ibrahim et al., 2016)

Taylor (2008) ศึกษาการเปรียบเทียบการพยากรณ์อนุกรมเวลาของปริมาณสายที่โทรเข้า ของศูนย์บริการลูกค้า โดยศึกษาจากกลุ่มการให้บริการ 5 กลุ่ม ของธนาคารพาณิชย์รายย่อยแห่งหนึ่ง ของสหราชอาณาจักร โดยทำการเปรียบเทียบ 2 วิธี คือ วิธีการของบ็อกซ์-เจนกินส์ และ วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลแบบโฮลท์-วินเทอร์ ผลการวิจัยพบว่าวิธีการของบ็อกซ์-เจนกินส์ ให้ความแม่นยำในการพยากรณ์ดีกว่าเนื่องให้ค่า MAPE ต่ำที่สุด ในขณะที่ Njeru (2018) ศึกษาการพยากรณ์ปริมาณการโทรเข้าของลูกค้าในอนาคต โดยทำการพยากรณ์ปริมาณสายโทรเข้ามาของลูกค้ารายวันของศูนย์บริการทั้งหมดในประเทศและศูนย์บริการอุตสาหกรรมโทรคมนาคมทั่วไป ประเทศเคนย่า 2 รูปแบบ คือ การโทรรายวันที่รับบนระบบเสียงแบบโต้ตอบ (IVR Hits) และการโทรที่ย้ายจากระบบ IVR ไปเพื่อรอการบริการ (Offered) ด้วยตัวแบบบ็อกซ์-เจนกินส์ โดยลักษณะข้อมูลมีความแปรผันของฤดูกาลเข้ามาเกี่ยวข้อง จึงเลือกใช้วิธี SARIMA โดยทำการเลือกตัวแบบ SARIMA ที่เหมาะสม และทำการเปรียบเทียบค่าพยากรณ์ใช้เกณฑ์ค่าเฉลี่ยความคลาดเคลื่อนกาลังสอง (MSE) ผลวิจัยพบว่าตัวแบบที่เหมาะสมที่สุดสำหรับ IVR คือ ARIMA(0,0,0)xSARIMA$(5,1,3)_7$ ส่วนตัวแบบที่เหมาะสมที่สุดสำหรับ Offered คือ ARIMA(0,0,0)xSARIMA$(6,3,1)_7$ จากนั้นนำตัวแบบที่ดีที่สุดไปพยากรณ์ 30 วันข้างหน้า เพื่อนำมาใช้ในการให้ข้อมูลเชิงลึกสำหรับการจัดการแรงงานที่เหมาะสม

การศึกษาวิจัยครั้งนี้ได้ทำการพยากรณ์จำนวนสายโทรเข้ารายวัน เนื่องจากการพยากรณ์รายวันจะเป็นประโยชน์ต่อการตัดสินใจ การวางแผน การเตรียมตัวจากการคาดการณ์ในอนาคต ผู้วิจัยได้สุ่มข้อมูลจำนวนสายโทรเข้ารายวันจากฐานข้อมูลของบริษัท มาทำการศึกษาเพื่อสร้างตัวแบบสำหรับการพยากรณ์จำนวนสายโทรเข้า โดยทำการวิเคราะห์ลักษณะของข้อมูล พร้อมทั้งจัดการกับข้อมูลเพื่อให้ข้อมูลมีความเหมาะสมที่จะนำไปใช้ทำตัวแบบพยากรณ์ทางสถิติเพื่อหาตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์ข้อมูลจำนวนสายโทรเข้าล่วงหน้าของบริษัทประกันชีวิตแห่งนี้

วิธีการดำเนินงานวิจัย

1. ลักษณะและแหล่งที่มาของข้อมูล

ข้อมูลที่ใช้ในงานวิจัยครั้งนี้ คือ ข้อมูลอนุกรมเวลาจำนวนปริมาณสายโทรเข้าแบบสุ่มของบริษัทประกันแห่งหนึ่ง ซึ่งข้อมูลนี้ถูกเก็บเป็นรายการ (Transaction) ของแต่ละสายที่โทรเข้ามาทาง Call Center โดยบริษัทมีการเก็บข้อมูลดังกล่าวตั้งแต่ พ.ศ.2559-2562 ในระบบฐานข้อมูลด้วยโปรแกรมจัดการฐานข้อมูล Oracle

2. ขั้นตอนการดำเนินงาน

การพิจารณาและคัดเลือกตัวแบบการพยากรณ์จำนวนสายเรียกเข้ารายวันที่เหมาะสมจำเป็นต้องจัดการกับข้อมูลอย่างเหมาะสมและวิเคราะห์ข้อมูลเพื่อทำให้เกิดความเข้าใจถึงลักษณะของข้อมูล และมีการจัดการกับข้อมูลอย่างเหมาะสม จากนั้นจึงจะสามารถเลือกวิธีการสร้างตัวแบบการพยากรณ์ที่เหมาะสมกับลักษณะข้อมูลของงานวิจัยนี้ โดยขั้นตอนการวิจัยสามารถแบ่งออกเป็น 3 ขั้นตอน

1) การจัดการข้อมูล

ผู้วิจัยต้องจัดการกับข้อมูลให้เหมาะสมก่อนที่จะนำข้อมูลไปวิเคราะห์ เพื่อใช้ในการสร้างตัวแบบการพยากรณ์ เช่น การสำรวจข้อมูล การหาความผิดปกติของข้อมูล (Outlier)

2) การวิเคราะห์ข้อมูล

การสร้างตัวแบบพยากรณ์รายวันนั้นจะต้องพิจารณาว่าข้อมูลที่จะพยากรณ์มีลักษณะแบบใด ดังนั้นควรตรวจสอบลักษณะข้อมูลก่อนเป็นอันดับแรกเพื่อที่จะได้เลือกวิธีการพยากรณ์ได้ถูกต้องเหมาะสม เช่นทำการตรวจสอบข้อมูลนั้นว่ามีส่วนประกอบของแนวโน้มและความแปรผันของฤดูกาลหรือไม่ หรือข้อมูลมีลักษณะเป็นฟังก์ชันเชิงเส้นตรง (Linear Model) หรือไม่ ซึ่งสามารถตรวจสอบได้จากการวาดกราฟ หรือการทดสอบทางสถิติและนำไปพยากรณ์ด้วยตัวแบบที่เหมาะสม

3) การสร้างตัวแบบการพยากรณ์

หลังจากข้อมูลได้ถูกเตรียมพร้อมสำหรับการสร้างตัวแบบการพยากรณ์จากข้อ 2. แล้ว ในงานวิจัยนี้จะทำการสร้างตัวแบบการพยากรณ์ด้วย 3 วิธีการ ตามข้อกำหนดขอบเขตงาน ดังนี้

1. การพยากรณ์ตัวแบบด้วยวิธีของบ็อกซ์-เจนกินส์

เป็นวิธีที่สร้างสมการตัวแบบจากความสัมพันธ์ของข้อมูลในเชิงเส้น ซึ่งข้อมูลที่นำมาใช้ในการสร้างตัวแบบต้องมีลักษณะที่คงที่ซึ่งสามารถทดสอบด้วย Augmented Dickey-Fuller Test หากข้อมูลยังไม่คงที่จะต้องทำการปรับข้อมูลให้ข้อมูลมีลักษณะคงที่ด้วยวิธีการหาผลต่าง ขั้นตอนต่อไปคือทำการหาพารามิเตอร์ในตัวแบบจากรูป

ฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และรูปฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) และทำการทดสอบว่าตัวแบบที่ได้มีความเหมาะสมหรือไม่ด้วยการทดสอบ Ljung-Box (Ljung & Box, 1978) นอกจากนี้สามารถเพิ่มความแม่นยำของตัวแบบด้วยการเพิ่มปัจจัยภายนอกเข้าไปในตัวแบบ

2. การพยากรณ์ตัวแบบด้วยวิธีโครงข่ายประสาทเทียม

หากข้อมูลมีลักษณะไม่เป็นฟังก์ชันเชิงเส้นตรง วิธีที่เหมาะสำหรับข้อมูลประเภทนี้คือ โครงข่ายประสาทเทียมวิธีเพอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron Neural Networks) ซึ่งวิธีนี้สามารถแก้ปัญหาพยากรณ์ที่มีความซับซ้อนได้ดี มีความแม่นยำอยู่ในระดับสูง โดยปกติจะใช้แก้ปัญหาที่ไม่เป็นเชิงเส้นตรง วิธีการพยากรณ์คือ ใช้ข้อมูลเก่าในอดีตของข้อมูลชุดนั้นเพียงอย่างเดียว มาพยากรณ์ค่าในอนาคต โดยไม่ใช้ตัวแปรอื่นๆ มาวิเคราะห์

3. การพยากรณ์ตัวแบบด้วยวิธีตัวแบบผสม

เป็นการนำตัวแบบที่ใช้พยากรณ์ข้อมูลอนุกรมเวลาส่วนที่เป็นฟังก์ชันเชิงเส้นตรง คือ ตัวแบบบ็อกซ์-เจนกินส์ มารวมกับตัวแบบที่ใช้พยากรณ์ข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรง คือ ตัวแบบเครือข่ายประสาทเทียม เพื่อทำให้ค่าพยากรณ์ที่ได้มีความถูกต้องและมีความแม่นยำมากยิ่งขึ้น

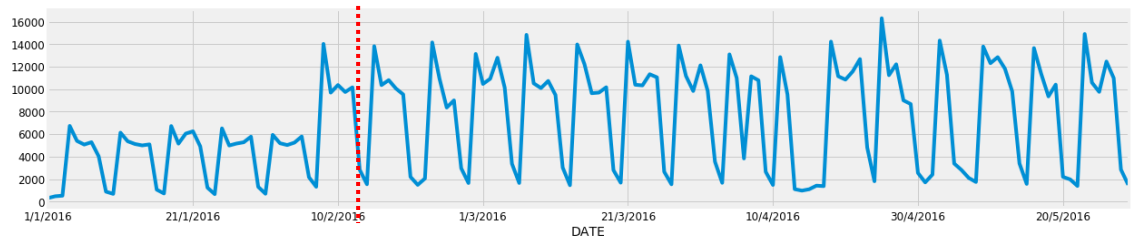
3. การวิเคราะห์ข้อมูล

เนื่องจากการสร้างตัวแบบพยากรณ์ จำเป็นต้องมีความเข้าใจในข้อมูลที่จะใช้ จึงต้องทำการวิเคราะห์ข้อมูลก่อนนำไปสร้างตัวแบบดังนี้

1) วิเคราะห์ข้อมูลอนุกรมเวลาว่ามีส่วนประกอบของแนวโน้มหรือความแปรผันของฤดูกาลเข้ามาเกี่ยวข้องหรือไม่

2) ถ้าข้อมูลอนุกรมเวลาเคลื่อนไหวรอบค่าเฉลี่ยหรือความแปรปรวนอยู่ในสถานะไม่คงที่จำเป็นต้องมีการแปลงข้อมูลนั้นให้มีสถานะคงที่ก่อนนำไปใช้งาน

ในการวิเคราะห์ข้อมูลได้อาศัยการเขียนโปรแกรมภาษา Python บน Jupyter Notebook



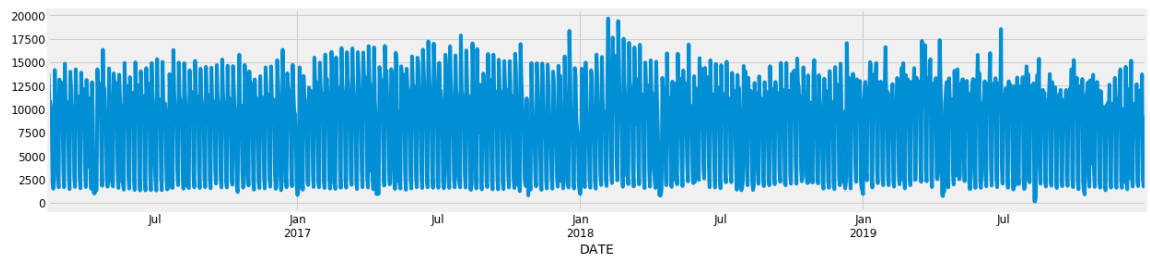
รูปที่ 1 ข้อมูลปริมาณสายโทรเข้ารายวันในช่วง 1/1/2016 - 29/5/2016

จากรูปที่ 1 จะสังเกตได้ว่าในช่วงต้นของกราฟตั้งแต่วันที่ 1/1/2016 ถึงวันที่ 14/2/2016 มีลักษณะของข้อมูลแตกต่างจากช่วงอื่น ดังนั้นจึงได้ทำการพล็อตกราฟในช่วงต้น เพื่อหาความแตกต่างระหว่างข้อมูลช่วงต้นกับข้อมูลที่เหลือ พบว่าข้อมูลมีความผิดปกติจริง เนื่องจากทางบริษัทประกันเปลี่ยนระบบในการจัดเก็บข้อมูล จึงทำให้ข้อมูลในช่วงต้นนี้ได้รับผลกระทบจากการเปลี่ยนระบบในการเก็บข้อมูล

ดังนั้น เพื่อลดความผิดพลาดและเพิ่มประสิทธิภาพของตัวแบบการพยากรณ์ ข้อมูลช่วงต้นนี้จะถูกตัดออกจากข้อมูลที่ใช้ในการวิเคราะห์ ทำให้ข้อมูลที่จะใช้ในการวิเคราะห์คือข้อมูลตั้งแต่วันที่ 15/2/2016 - 31/12/2019 (จำนวน 1,416 วัน หรือ 1,416 จุดเวลา) โดยนำมาทำการแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่

1) ข้อมูลฝึกสอนตั้งแต่วันที่ 15/2/2016 - 31/12/2018 (จำนวน 1,051 จุดเวลา) สำหรับสร้างตัวแบบ

2) ชุดข้อมูลทดสอบ 1/1/2019 - 31/12/2019 (จำนวน 365 จุดเวลา) สำหรับประเมินความถูกต้องของตัวแบบ

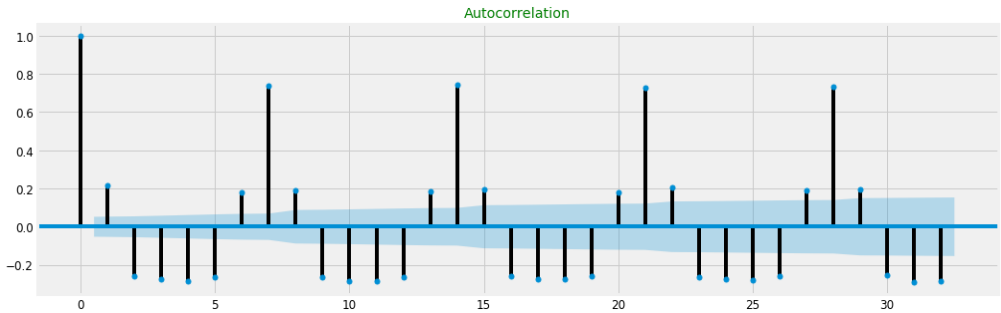


รูปที่ 2 ข้อมูลปริมาณสายโทรเข้ารายวันในช่วง 15/2/2016 - 31/12/2018

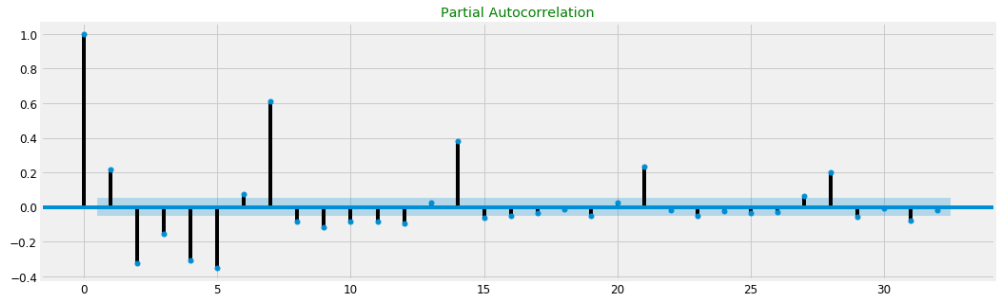
จากการพล็อตกราฟระหว่างวันที่และปริมาณสายโทรเข้าในระหว่างวันที่ 15/2/2016 ถึง วันที่ 31/12/2018 เพื่อวิเคราะห์ลักษณะข้อมูล ดังรูปที่ 2 จะพบว่าข้อมูลค่อนข้างมีลักษณะคงที่ กล่าวคือข้อมูลนี้มีค่าเฉลี่ยและความแปรปรวนคงที่ และเมื่อทำการพล็อตกราฟระหว่างวันที่และปริมาณสายโทรเข้าในระหว่างวันที่ 4/6/2016 ถึงวันที่ 25/5/2016 เพื่อวิเคราะห์ความแปรผันของฤดูกาล ซึ่งเป็นช่วงเวลาที่เกิดขึ้นการเปลี่ยนแปลงซ้ำ ๆ กันในช่วงเวลาหนึ่ง ดังรูปที่ 3 จะเห็นได้ว่าปริมาณสายโทรเข้ามีความแปรผันของฤดูกาลอย่างชัดเจน หากอนุกรมเวลาใดๆ ที่พิจารณามี

ความแปรผันทางฤดูกาลอยู่ด้วยจะต้องทำการกำจัดความแปรผันทางฤดูกาลออกไปก่อน ซึ่งวิธีกำจัดความผันแปรทางฤดูกาลนั้นมีหลายวิธี ในวิจัยนี้จะใช้วิธีหาผลต่างของฤดูกาล

รูปที่ 3 ข้อมูลปริมาณสายโทรเข้าในช่วง 4/6/2016-25/6/2016



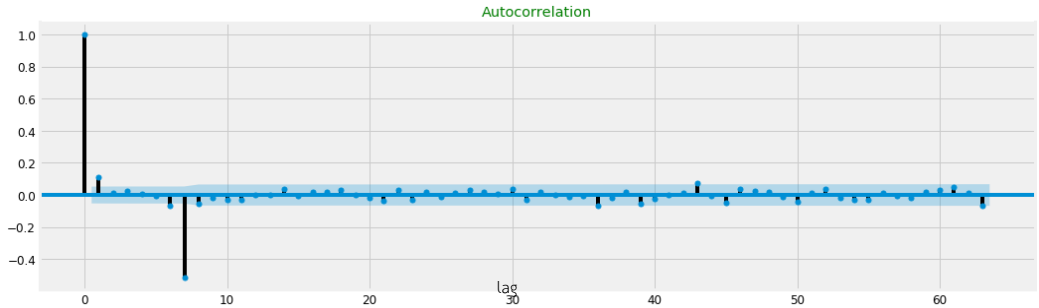
รูปที่ 4 ค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของปริมาณสายโทรเข้ารายวัน



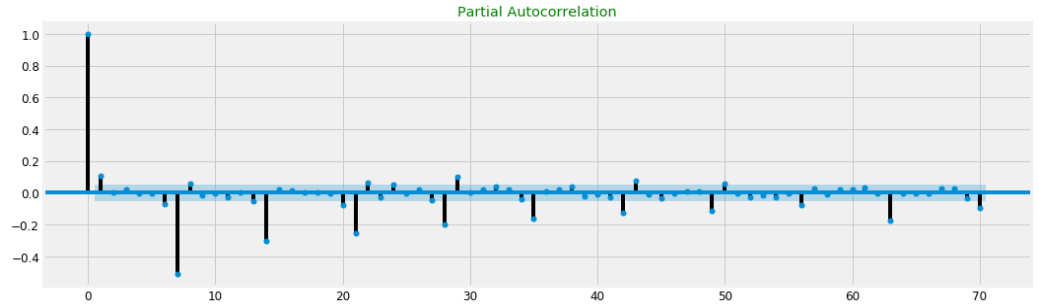
รูปที่ 5 ค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของปริมาณสายโทรเข้ารายวัน

เมื่อพิจารณารูปที่ 4 แสดงค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของปริมาณสายโทรเข้ารายวัน พบว่าฟังก์ชันสหสัมพันธ์ในตัวเอง ลดลงเข้าสู่ศูนย์อย่างช้า ๆ และรูปที่ 5 พบว่ากราฟฟังก์ชันสหสัมพันธ์ในตัวเอง

บางส่วน (PACF) ลดลงเข้าสู่ศูนย์อย่างช้าๆ แสดงว่า อนุกรมเวลาของปริมาณสายโทรเข้ารายวันไม่คงที่ จึงทำการหาผลต่างครั้งที่ 1 ของความแปรผันเกี่ยวกับฤดูกาลของปริมาณสายโทรเข้ารายวัน



รูปที่ 6 ค่าฟังก์ชันสหสัมพันธ์ในตัวเอง (ACF) ของผลต่างฤดูกาลครั้งที่ 1



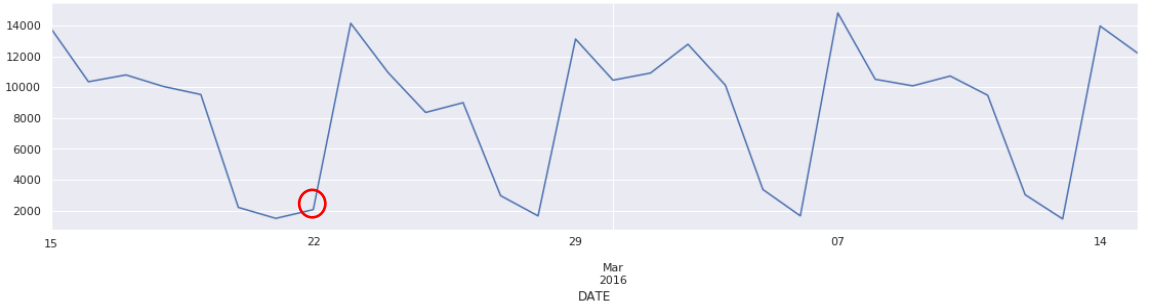
รูปที่ 7 ค่าฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (PACF) ของผลต่างฤดูกาลครั้งที่ 1

จากรูปที่ 6 และ 7 เมื่อทำการหาผลต่างครั้งที่ 1 ของความแปรผันเกี่ยวกับฤดูกาล พบว่าค่า ACF และ PACF ลดลงเข้าสู่ศูนย์อย่างรวดเร็ว แสดงว่าอนุกรมเวลาคงที่

เมื่อทำการพิจารณาวันหยุดต่างๆ ในปี 2016 – 2018 จากรูปที่ 8 พบว่าในวันที่ 5 ธันวาคม วันที่ 10 ธันวาคม 31 ธันวาคม และ 1 มกราคมของทุกปี มีปริมาณสายที่โทรเข้ามาน้อยกว่าปกติ เนื่องจากเป็นวันหยุดทางราชการ

รูปที่ 8 ปริมาณสายโทรเข้ามาในแต่ละวันตั้งแต่วันที่ 1/12/2016-10/1/2017

และจากรูปที่ 9 พบว่า ในที่วันที่ 22 กุมภาพันธ์ 2016 มีปริมาณสายที่โทรเข้ามาน้อยกว่าปกติ เนื่องจากเป็นวันชดเชยมาฆบูชา



รูปที่ 9 ปริมาณสายโทรเข้ามาในแต่ละวันตั้งแต่วันที่ 15/2/2016-15/3/2016

เนื่องจากสมบัติความคงที่ของข้อมูลเป็นสิ่งจำเป็นที่ข้อมูลพึงมีก่อนที่จะทำการพยากรณ์ตัวแบบวิธีบ็อกซ์-เจนกินส์ ดังนั้นจึงทำการทดสอบด้วยตัวทดสอบทางสถิติ Augmented Dickey-Fuller Test (ADF) ที่ระดับนัยสำคัญ 0.05 พบว่าค่า p-value = 0.0000002 < α=0.05 สรุปได้ว่า ข้อมูลชุดนี้มีลักษณะคงที่ ที่ระดับนัยสำคัญ 0.05

4. ตัวแบบพยากรณ์ที่ใช้ในการวิจัย

1) การพยากรณ์วิธีการบ็อก-เจนกินส์

วิธีการของบ็อกซ์-เจนกินส์ เหมาะกับข้อมูลที่มีลักษณะคงที่และไม่มีส่วนประกอบของแนวโน้มหรือความแปรผันเกี่ยวกับฤดูกาล หากข้อมูลยังมีลักษณะไม่คงที่หรือมีส่วนประกอบของแนวโน้มหรือความแปรผันเกี่ยวกับฤดูกาลอย่างชัดเจน จะต้องทำการหาผลต่างครั้งที่ 1 (1^{st} differencing) ซึ่งแบบจำลองมีข้อตกลงเบื้องต้นคือค่าสังเกตปัจจุบันมีความสัมพันธ์เป็นกระบวนเชิงเส้นระหว่างค่าสังเกตและค่าคลาดเคลื่อนในอดีต (Chatfield, 2016) เนื่องจากพบว่าข้อมูลมีความแปรผันทางฤดูกาลอย่างชัดเจน ทางผู้วิจัยจึงใช้การพยากรณ์วิธีบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMA ซึ่งเป็นตัวแบบที่เหมาะกับข้อมูลที่มีลักษณะนี้

2) ตัวแบบโครงข่ายประสาทเทียมวิธีเพอร์เซ็ปตรอนหลายชั้น

ตัวแบบ ANN สามารถประยุกต์ใช้ในการพยากรณ์อนุกรมเวลาได้ ส่วนใหญ่นิยมใช้พยากรณ์ข้อมูลอนุกรมเวลาชุดเดียวซึ่งเรียกว่า “Univariate Time Series” โดยกำหนดชั้นอินพุตแสดงข้อมูลในอดีต จำนวน n ข้อมูล และชั้นเอาท์พุตแสดงการพยากรณ์วันข้างหน้า 1 วัน และมีชั้นซ่อนเร้น ที่สามารถกำหนดความซับซ้อนของตัวแบบเพื่อให้ได้ค่าพยากรณ์ที่แม่นยำขึ้น

การพยากรณ์ข้อมูลอนุกรมเวลามีเป้าหมาย คือการพยากรณ์ข้อมูลล่วงหน้า โดยใช้ข้อมูลเก่าในอดีตของข้อมูลชุดนั้นเพียงอย่างเดียวมาพยากรณ์ค่าในอนาคต โดยไม่ใช่ตัวแปรอื่นมาวิเคราะห์ โดยส่วนมากการพยากรณ์อนุกรมเวลาจะเกี่ยวข้องกับตัวเลข (Anansapsuk, 2017)

Zhang (2012) กล่าวไว้ดังนี้ สมมติว่ามีข้อมูลอนุกรมเวลาอยู่ N ค่า $Y_1, Y_2, ..., Y_N$ ในชุดข้อมูลฝึกสอน (Training set) และต้องการพยากรณ์ 1 ช่วงเวลาไปข้างหน้า (1-step-ahead) คือ Y_{t+1} เมื่อกำหนดจำนวน lag หรือจุดเวลาย้อนหลัง เช่น lag = n = 1 คือทำการ Y_{t+1} ด้วย Y_t, lag = n = 2 คือทำการพยากรณ์ Y_{t+1} ด้วย Y_t และ Y_{t-1} lag = n = 3 คือทำการพยากรณ์ Y_{t+1} ด้วย Y_t, Y_{t-1} และ Y_{t-2}

3) **การพยากรณ์วิธีตัวแบบผสม**

Zhang (2003) ได้สรุปหลักในการสร้างตัวแบบผสมไว้ดังนี้ กล่าวคือการสร้างตัวแบบผสมจะกำหนดว่าข้อมูลที่นำมาวิเคราะห์นั้นมี 2 องค์ประกอบ คือ องค์ประกอบที่เป็นเส้นตรง (Linear Component) และองค์ประกอบที่ไม่เป็นเส้นตรง (Nonlinear Component) เพื่อทำให้ค่าพยากรณ์ที่ได้มีความถูกต้องและแม่นยำมากขึ้น วิธีการสร้างตัวแบบผสมมีขั้นตอนดังนี้

1) พยากรณ์ข้อมูลในส่วนฟังก์ชันเชิงเส้นตรง โดยวิเคราะห์ข้อมูลอนุกรมเวลาในส่วนฟังก์ชันเชิงเส้นตรงจากตัวแบบ SARIMAX

2) คำนวณค่าส่วนเหลือ หรือ ความคลาดเคลื่อนจากการพยากรณ์ (Residuals) ที่ได้จากการพยากรณ์ด้วยตัวแบบ SARIMAX ไปพยากรณ์ในส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรง

3) พยากรณ์ข้อมูลในส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรง โดยนำส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงไปพยากรณ์ด้วยตัวแบบเครือข่ายประสาทเทียม ด้วยการกำหนดค่าไฮเปอร์พารามิเตอร์หรือกระบวนการในการสร้างตัวแบบ

4) คำนวณค่าพยากรณ์รวม (Total Forecasting) โดยนำตัวแบบมารวมกันที่ได้จากการพยากรณ์ข้อมูลในส่วนฟังก์ชันเชิงเส้นตรง และ การพยากรณ์ข้อมูลในส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรง

หลังจากนั้นทำการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ โดยพิจารณาจากค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ (Mean Absolute Percentage Error : MAPE) เป็นเกณฑ์ในการเปรียบเทียบตัวแบบ ซึ่งตัวแบบที่ให้ค่า MAPE ที่ต่ำที่สุด จะเป็นตัวแบบที่ดีที่สุด และนำตัวแบบที่ดีที่สุดไปใช้ในการพยากรณ์ต่อไป

ผลการวิจัย

จากการตรวจสอบข้อมูล พบว่าข้อมูลมีความแปรผันทางฤดูกาลทุกๆ 7 วัน จึงกำหนดค่า s =7 ในการพิจารณาตัวแบบที่เหมาะสม จะทำการพิจารณาตัวแบบที่มีค่า AIC น้อยที่สุด

ตารางที่ 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)xSARIMA(0, 0, 0)$_7$	30,037.485
ARIMA(0, 0, 0)xSARIMA(0, 0, 1)$_7$	29,115.961
ARIMA(0, 0, 0)xSARIMA(0, 1, 0)$_7$	26,953.149
ARIMA(0, 0, 0)xSARIMA(0, 1, 1)$_7$	25,988.764
ARIMA(1, 1, 1)xSARIMA(0, 1, 0)$_7$	26,935.634
ARIMA(1, 1, 1)xSARIMA(0, 1, 1)$_7$	**25,900.738**
…	…
ARIMA(1, 1, 2)xSARIMA(0, 0, 2)$_7$	27,000.077
ARIMA(1, 1, 2)xSARIMA(0, 2, 0)$_7$	28,211.030
ARIMA(1, 1, 2)xSARIMA(0, 2, 2)$_7$	26,983.504
ARIMA(1, 1, 2)xSARIMA(0, 2, 2)$_7$	26,983.504
ARIMA(1, 2, 0)xSARIMA(0, 0, 2)$_7$	28,622.106
ARIMA(2, 0, 1)xSARIMA(3, 1, 1)$_7$	27,134.785
ARIMA(2, 0, 2)xSARIMA(3, 1, 2)$_7$	26,105.861
ARIMA(2, 1, 1)xSARIMA(3, 1, 1)$_7$	26,953.248
ARIMA(2, 1, 2)xSARIMA(3, 1, 1)$_7$	26,044.364
ARIMA(2, 1, 2)xSARIMA(3, 1, 2)$_7$	25,935.634

เมื่อพิจารณาตารางที่ 1 ค่า AIC ที่น้อยที่สุด พบว่าอันดับที่ดีที่สุด ได้แก่ p=2 , d=0 , q=2 และ P=3 , D=1 ,Q=1 และตัวแบบที่เหมาะสมคือ ตัวแบบ ARIMA(2,0,2)xSARIMA(3,1,1)$_7$ ซึ่งมีรูปแบบทั่วไปดังสมการที่ 1

$$(1- f_1 B)(1- B)(1- B^7)\,Y_t = (1- q_1 B)(1- Q_1 B^7)e_t \quad (1)$$

โดยกำหนด

Y_t แทน ค่าพยากรณ์ ณ เวลา t

 doi: 10.14416/j.appsci.2022.01.016

e_t แทน ค่าความคลาดเคลื่อน ณ เวลาที่ t

B แทน ตัวดำเนินการย้อนหลัง

f_1 แทน การถดถอยในตัวเองแบบไม่มีฤดูกาลอันดับ 1

q_1 แทน ค่าเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับ 1

Q_1 แทน ค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอับดับ 1

หลังจากได้ตัวแบบที่เหมาะสมจากสมการที่ 4.1 แล้วจึงนำตัวแบบดังกล่าวมาประมาณค่าพารามิเตอร์ดังตารางที่ 2

ตารางที่ 2 การทดสอบค่าพารามิเตอร์ ARIMA(1,1,1)xSARIMA$(0,1,1)_7$

Statisics	Coef	SE Coef	t	p-value
f_1	0.1185	0.017	6.929	0.000
q_1	-0.9999	0.022	-46.418	0.000
Q_1	-1.0782	0.009	-115.309	0.000

จากตารางที่ 4.2 พบว่าค่า p-value = 0.000 < α =0.05 จึงปฏิเสธ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปคือค่าพารามิเตอร์ f_1, q_1, Q_1 ในตัวแบบมีค่าไม่เท่ากับ 0 ดังนั้นตัวแบบ ARIMA(1,1,1) x SARIMA$(0,1,1)_7$ เป็นตัวแบบที่เหมาะสม

จากนั้นทำการตรวจสอบค่าความคลาดเคลื่อนว่าจะต้องไม่มีสหสัมพันธ์ในตัวเองด้วย Box-Pierce and Ljung-Box Tests พบว่า p-value = 0.8 > α =0.05 จึงยอมรับ H_0 ที่ระดับนัยสำคัญ 0.05 สรุปว่า ค่าความคลาดเคลื่อนของตัวแบบที่พิจารณาไม่มีสหสัมพันธ์ในตัวเองทุกช่วงเวลา ดังนั้นแสดงว่าตัวแบบ ARIMA(1,1,1)xSARIMA$(0,1,1)_7$ เป็นตัวแบบที่เหมาะสม

ดังนั้นสมพยากรณ์ด้วยตัวแบบ SARIMA คือ

$$Y_t = (1+f_1)Y_{t-1} - f_1Y_{t-2} + Y_{t-7} - (1+f_1)Y_{t-8} + f_1Y_{t-9} + e_t - Q_1e_{t-7} - q_1e_{t-1} + q_1Q_1e_{t-8}$$

สำหรับการพยากรณ์วิธีการบ็อกซ์-เจนกินส์ด้วยตัวแบบ SARIMAX สำหรับปริมาณสายโทรเข้ามาในแต่ละวันตั้งแต่วันที่ 1/12/2016-10/1/2017 พบว่าปริมาณสายที่โทรเข้ามาน้อยในวันหยุดทางราชการ และหาก

วันหยุดราชการตรงกับ เสาร์ - อาทิตย์ จะมีการชดเชยในวันจันทร์ถัดไป โดยประกอบไปด้วย วันสงกรานต์, วันปีใหม่, วันแม่, วันปิยมหาราช, วันพ่อ, วันจักรี, วันรัฐธรรมนูญ และวันแรงงาน รวมถึงวันหยุดทางศาสนา และหากวันหยุดทางศาสนาตรงกับ เสาร์ - อาทิตย์ จะมีการชดเชยในวันจันทร์ถัดไป โดยประกอบไปด้วยวันมาฆบูชา, วันวิสาขบูชา และวันอาสาฬหบูชา ทั้งนี้ได้เพิ่มตัวแปรอิสระหรือตัวแปรภายนอก (Exogenous) ขึ้นมาโดยที่

X1 แทน วันหยุดทางราชการ

X2 แทน วันหยุดชดเชยทางราชการ

X3 แทน วันหยุดพิเศษ

X4 แทน วันหยุดชดเชยพิเศษ

X5 แทน วันหยุดสงกรานต์-ปีใหม่

X6 แทน วันหยุดสงกรานต์-ปีใหม่ที่ตรงกับวันเสาร์-อาทิตย์

X7 แทน วันหยุดชดเชยสงกรานต์-ปีใหม่

โดยตัวแบบที่เหมาะสม คือ ARIMA (2,0,2)xSARIMA(3,1,1) with Exogenous ดังนั้นสมการพยากรณ์ด้วยตัวแบบ SARIMAX คือ

$$Y_t = (1+f_1)Y_{t-1} - f_1 Y_{t-2} + Y_{t-7} - (1+f_1)Y_{t-8} + f_1 Y_{t-9} + e_t - Q_1 e_{t-7} - q_1 e_{t-1} + q_1 Q_1 e_{t-8}$$
$$+ 9,663.5124X1 - 12,690X2 - 8,982.4182X3 - 12,420X4 - 5,856.0354X5$$
$$+ 1,480.2860X6 - 12,410X7$$

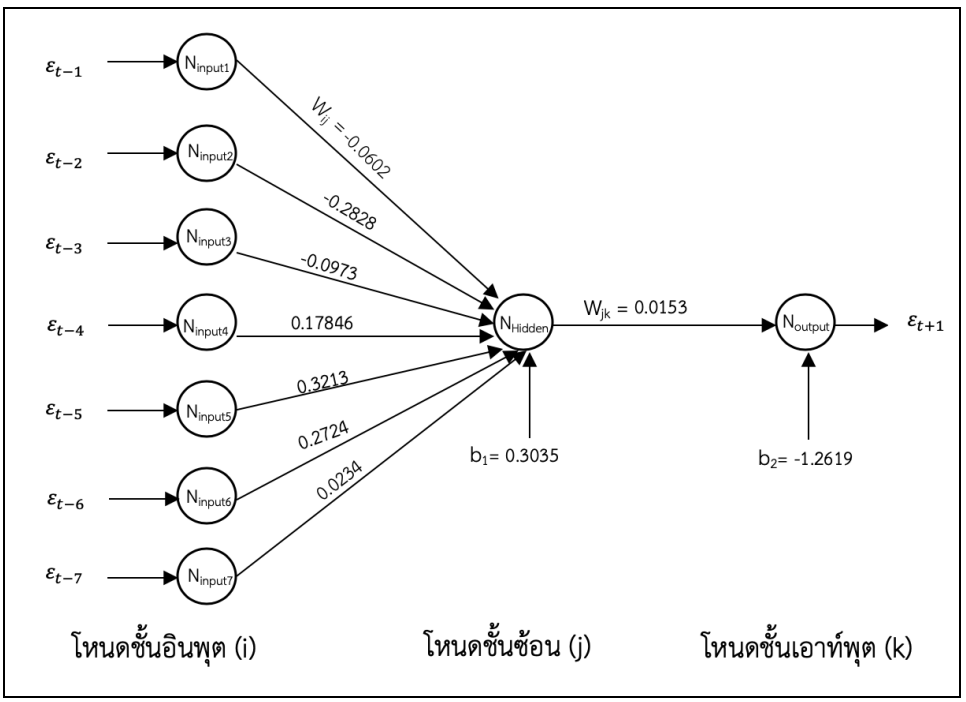
หลังจากนั้นคำนวณค่าส่วนเหลือ หรือ ความคลาดเคลื่อนจากการพยากรณ์ (Residuals) โดยส่วนเหลือนั้นจะเหลือเพียงฟังก์ชันส่วนไม่เป็นเส้นตรงด้วยตัวแบบ SARIMAX นำเข้าสู่ตัวแบบโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้นด้วยการเรียนรู้แบบไปข้างหน้า (Feed forward) โดยเลือกกำหนดค่าไฮเปอร์พารามิเตอร์ดังตารางที่ 3

ตารางที่ 3 การกำหนดค่าไฮเปอร์พารามิเตอร์หรือกระบวนการในการสร้างตัวแบบ

ข้อกำหนดต่าง ๆ ในการสร้างตัวแบบ	ค่าไฮเปอร์พารามิเตอร์
อัลกอริทึม (Algorithm)	Neural Network
จำนวนโหนดในชั้นอินพุต (Input Node)	7
จำนวนชั้นซ้อน (Hidden Layer)	1

ข้อกำหนดต่าง ๆ ในการสร้างตัวแบบ	ค่าไฮเปอร์พารามิเตอร์
จำนวนโหนดในชั้นซ้อน (Hidden Node)	1
จำนวนโหนดในชั้นเอาท์พุต (Output Node)	1
ฟังก์ชันกระตุ้น (Activation Function)	Linear Function
Optimizer	Adam
โมเมนตัม (Momentum)	0.9
อัตราการเรียนรู้ (Learning Rate)	0.001

จากตารางที่ 3 แสดงการกำหนดค่าไฮเปอร์พารามิเตอร์สำหรับการสร้างตัวแบบโครงข่ายประสาทเทียม โดยสมการพยากรณ์วิธีผสมเป็นการรวมข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรงที่ได้จากการพยากรณ์ด้วยตัวแบบ SARIMAX และข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงที่ได้จากการพยากรณ์ค่าส่วนเหลือด้วยตัวแบบโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้นที่พยากรณ์ 1 วันข้างหน้าด้วยค่าส่วนเหลือ 7 วันก่อนหน้า หรืออยู่ในรูปแบบ 7-1-1 เข้าด้วยกัน มีรูปแบบโครงสร้างดังรูปที่ 10



รูปที่ 10 รูปแบบโครงสร้างในการพยากรณ์ส่วนไม่เป็นฟังก์ชันเชิงเส้นตรง

 doi: 10.14416/j.appsci.2022.01.016

โดยกำหนดให้

e_{t-1} แทน ค่าส่วนเหลือก่อนหน้า 1 วัน

e_{t-2} แทน ค่าส่วนเหลือก่อนหน้า 2 วัน

e_{t-3} แทน ค่าส่วนเหลือก่อนหน้า 3 วัน

e_{t-4} แทน ค่าส่วนเหลือก่อนหน้า 4 วัน

e_{t-5} แทน ค่าส่วนเหลือก่อนหน้า 5 วัน

e_{t-6} แทน ค่าส่วนเหลือก่อนหน้า 6 วัน

e_{t-7} แทน ค่าส่วนเหลือก่อนหน้า 7 วัน

e_{t+1} แทน ค่าส่วนเหลือ 1 วันข้างหน้า

b_1 แทน ค่าไบเอส (Bias) ในชั้นซ้อน

b_2 แทน ค่าไบเอส (Bias) ในชั้นเอาต์พุท

w_{ij} แทน ค่าถ่วงน้ำหนักของเส้นเชื่อมโยงกับข้อมูลนำเข้าที่ i ไปยังโหนดที่ j หรือโหนดชั้นซ้อน

w_{jk} แทน ค่าถ่วงน้ำหนักของเส้นเชื่อมโยงกับโหนดชั้นซ้อนที่ j ไปยังโหนดที่ k หรือโหนดชั้นเอาท์พุต

จากการทดลองได้ตัวแบบที่เหมาะสมในการพยากรณ์อนุกรมเวลาในส่วนไม่เป็นฟังก์ชันเชิงเส้นตรง ดังตารางที่ 4

ตารางที่ 4 ค่าพารามิเตอร์ของตัวแบบที่เหมาะสมที่สุดส่วนไม่เป็นฟังก์ชันเชิงเส้นตรง

ตัวแบบ	**พารามิเตอร์**		
ANN (MLP)	**Lag**	**Hidden neurons**	**Output neurons**
	7	1	1

การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

หลังจากทดลองหาค่าพารามิเตอร์ที่ดีที่สุดของตัวแบบทั้ง 4 วิธี ได้แก่ ตัวแบบ SARIMA, ตัวแบบ SARIMAX, ตัวแบบโครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น และตัวแบบผสม Hybrid จากนั้นทำการเปรียบเทียบความแม่นยำของค่าพยากรณ์โดยใช้ค่าเฉลี่ยเปอร์เซ็นต์ความคลาดเคลื่อนสมบูรณ์ดังตารางที่ 5

ตารางที่ 5 ค่า MAPE ของแต่ละตัวแบบ

ตัวแบบ	MAPE
SARIMA	44.97%
SARIMAX	24.06%
ANN (MLP)	43.70%
Hybrid	**24.02%**

เมื่อพิจารณาค่า MAPE ที่ดีที่สุดของแต่ละตัวแบบ เพื่อเลือกแบบจำลองที่มีความแม่นยำที่สุดนั้น จะเห็นว่าแบบจำลองที่เหมาะสมที่สุดในการพยากรณ์ปริมาณสายโทรเข้า ได้แบบจำลองที่ดีที่สุดคือ ตัวแบบจำลองผสมระหว่างตัวแบบ SARIMA กับตัวแบบโครงข่ายประสาทเทียม ที่มีค่า MAPE เท่ากับ 24.02%

สรุปผลการวิจัย

ผลการเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์ทั้ง 4 ตัวแบบ สามารถสรุปผลการดำเนินวิจัยได้ว่าตัวแบบผสม โดยนำค่าส่วนเหลือที่ได้จากการพยากรณ์ด้วยตัวแบบ ARIMA(1,1,1) x SARIMA$(0,1,1)_7$ และ ตัวแปรปัจจัยภายนอกซึ่งประกอบไปด้วย X1 คือวันหยุดทางราชการ, X2 คือวันหยุดชดเชยทางราชการ, X3 คือวันหยุดพิเศษ, X4 คือวันหยุดชดเชยพิเศษ, X5 คือวันหยุดสงกรานต์-ปีใหม่, X6 คือวันหยุดสงกรานต์-ปีใหม่ตรงกับวันเสาร์อาทิตย์, X7 คือวันหยุดชดเชยสงกรานต์-ปีใหม่ มาสร้างตัวแบบโครงข่ายประสาทเทียมด้วยวิธีเพอร์เซ็ปตรอนหลายชั้น พบว่าตัวแบบโครงข่ายประสาทเทียมชั้นข้อมูลนำเข้ามีจำนวนโหนด 7 โหนด ชั้นซ่อนมีจำนวนโหนด 1 โหนด และชั้นผลลัพธ์มีจำนวนโหนด 1 โหนด หรือเรียกว่าตัวแบบโครงข่ายประสาทเทียม 7-1-1 เป็นตัวแบบที่เหมาะสมที่สุด และเมื่อนำไปทดสอบความแม่นยำของตัวแบบให้ค่า MAPE เท่ากับ 24.02% โดยตัวแบบดังกล่าวสามารถนำไปพยากรณ์ล่วงหน้า เพื่อเป็นแนวทางหรือข้อมูลเพิ่มเติมในการตัดสินใจในการวางแผนการทำงานของศูนย์บริการข้อมูลลูกค้า

เอกสารอ้างอิง

Anansapsuk, A. (2017). *A Comparative Study of Hybrid Time Series Models for Forecasting Seasonal Time Series* [Unpublished master's thesis]. Chulalongkorn University. (in Thai)

Ljung, G., & Box, G. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, *65*(2), 297-303.

Chatfield, C. (2016). *The Analysis of Time Series: An Introduction* (6th ed.). New York, Chapman and Hall/CRC Press.

Ibrahim, R., Ye, H., L'Ecuyer, P., & Shen, H. (2016). Modeling and forecasting call center arrivals: A literature survey and a case study. *International Journal of Forecasting*, *32*(3), 865-874.

Njeru, K., Na, K., & Ivivi, J. (2018). Forecasting future customer call volumes: A case study. *International Journal on Future Revolution in Computer Science & Communicaton Engineering*, *4*(6), 12-16.

Taylor, J. W. (2008). A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Science*, *54*(2), 253–265.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175.

Zhang, G. P. (2012). Neural Networks for Time-Series Forecasting. In: Rozenberg G., Bäck T., Kok J.N. (eds.) *Handbook of Natural Computing*. Berlin, Heidelberg, Springer. https://doi.org/10.1007/978-3-540-92910-9_14