การพยากรณ์ยอดขายสำหรับการขายปลีกในอีคอมเมิร์ซโดยใช้การเรียนรู้ของเครื่อง

วรัญชรี ถิรพัฒนาพร[1*], ศุภร คนธภักดี[2]

บทคัดย่อ

การพยากรณ์ยอดขายนั้นช่วยให้ผู้ประกอบการสามารถคาดการณ์ยอดขายหรือมูลค่าของสินค้าที่ต้องผลิตได้ ส่งผลให้สามารถผลิตสินค้าได้เพียงพอตามความต้องการของผู้บริโภค งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาการสร้างแบบจำลองที่ใช้สำหรับการพยากรณ์ยอดขายสินค้า โดยใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึก Long-Short-Term Memory (LSTM) เปรียบเทียบสมรรถนะกับแบบจำลองโครงข่ายปราสาทเทียมแบบ Multi-Layer Perceptron (MLP) แบบจำลองอนุกรมเวลาแบบ Seasonal Autoregressive Integrated Moving Averages (SARIMA) และ Seasonal Autoregressive Integrated Moving Averages with Exogenous Regressors (SARIMAX) ในการพยากรณ์ยอดขายสินค้าล่วงหน้า โดยใช้ชุดข้อมูลการสินค้าจากร้านค้าแห่งหนึ่งที่ขายบนแพลตฟอร์มออนไลน์ ซึ่งมีข้อมูลยอดขายรายวันตั้งแต่เดือน มกราคม 2564 ถึง มิถุนายน 2566 ผู้วิจัยสนใจที่จะเปรียบเทียบผลของการพยากรณ์ โดยพิจารณาจากค่า RMSE MAE และ MAPE ในการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้ง 4 วิธี ผลจากการทดลองพบว่า แบบจำลอง LSTM มีประสิทธิภาพในการพยากรณ์ยอดขายที่แม่นยำกว่า MLP SARIMA และ SARIMAX ในสินค้าแจกันทรงสูง และแจกันทรงเอียงกลาง ซึ่งสินค้าทั้ง 2 ชนิดนี้มีความซับซ้อนของข้อมูลอย่างมาก เนื่องจากในปี 2564 ช่วงสถานการณ์โควิดทำให้มียอดขายสูงขึ้นอย่างรวดเร็ว และในปีต่อมายอดขายกลับลดลงอย่างรวดเร็ว ซึ่งในแบบจำลอง LSTM มีความสามารถในการจัดเรียงข้อมูลและจำความจำในช่วงเวลาย้อนหลังได้ดี ซึ่งช่วยให้แบบจำลองสามารถจับฤดูกาลและแนวโน้มของข้อมูลได้อย่างแม่นยำ และมีความสามารถในการจัดการกับลักษณะข้อมูลที่ซับซ้อนได้อย่างมีประสิทธิภาพ และแบบจำลอง MLP มีประสิทธิภาพในการพยากรณ์ยอดขายที่แม่นยำกว่า LSTM SARIMA และ SARIMAX ในสินค้าแจกันทรงสูงใหญ่ เนื่องจากแจกันทรงสูงใหญ่มีลักษณะของข้อมูลที่มีความซับซ้อนน้อยลง จึงทำให้ MLP มีประสิทธิภาพที่ดีกว่า และจากการทดลองแบบจำลอง SARIMAX เพื่อพิจารณาวันโปรโมชั่นเข้าไปเป็นปัจจัยภายนอกในแบบจำลอง ซึ่งจากผลการทดลองพบว่าแบบจำลอง SARIMAX ให้ประสิทธิภาพที่ดีกว่าแบบจำลอง SARIMA แสดงว่าวันโปรโมชั่นส่งผลให้แบบจำลองมีประสิทธิภาพที่ดีขึ้น

คำสำคัญ : การพยากรณ์ยอดขายสินค้า, การเรียนรู้ของเครื่อง, โครงข่ายประสาทเทียมแบบลึก, อนุกรมเวลา

---

[1] หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

[2] คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ กรุงเทพฯ 10110

* Corresponding author:Tel.: 082-3297083 E-mail address: varuncharee.tpp@g.swu.ac.th

SALES FORECASTING FOR RETAIL SALES IN E-COMMERCE USING MACHINE LEARNINGS

Varuncharee Tiraphattanaporn[1*], Subhorn Khonthapagdee[2]

Abstract

The forecasting of sales enables business owners to anticipate and meet consumer demand, thereby ensuring that products are produced in sufficient quantities, neither exceeding nor falling short of consumer needs. This research aimed to study the creation of models for sales forecasting using Long Short-Term Memory (LSTM) and compare their performance with Multi-Layer Perceptron (MLP), Seasonal Autoregressive Integrated Moving Averages (SARIMA), and Seasonal Autoregressive Integrated Moving Averages with Exogenous Regressors (SARIMAX) models. The dataset used in this study comes from a retail store in Thailand that sells products online. The dataset includes daily transactional sales data from January 2021 to June 2023. The researchers aim to compare the performance of forecasting using the LSTM model MLP model SARIMA model ane SARIMAX model, considering metrics such as RMSE, MAE, and MAPE. The experimental results indicate that the LSTM model outperforms MLP, SARIMA, and SARIMAX in accurately forecasting sales for Tall Vase and Medium Slanted Vase products. Both product types exhibit highly complex data patterns, especially during the COVID-19 situation in 2020, where sales rapidly increased, followed by a sharp decline in the subsequent year. The LSTM model's ability to sequence data and retain memory of past trends allows it to accurately capture seasonal patterns and data trends. Additionally, it efficiently handles the complex data characteristics, leading to superior performance. The MLP model demonstrates better performance in forecasting sales for Large Tall Vase products, which have less complex data characteristics compared to the other product types. The simplified data nature of these products contributes to the MLP model's effectiveness. Furthermore, the SARIMAX model, which incorporates promotional days as external factors, outperforms the SARIMA model. This indicates that promotional days significantly influence the model's forecasting accuracy, improving its performance.

Keywords : Sales Forecast, Machine Learning, Long-Short-Term Memory, time series models

---

[1] Data Science, Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

[2] Faculty of Science, Srinakharinwirot University, Bangkok, 10110, Thailand

* Corresponding author:Tel.: 082-3297083 E-mail address: varuncharee.tpp@g.swu.ac.th

บทนำ

ในปัจจุบันเศรษฐกิจกำลังก้าวเข้าสู้ยุคดิจิทัล อีกทั้งอินเทอร์เน็ตได้เข้ามามีบทบาทอย่างยิ่งในชีวิตประจำวัน จึงส่งผลให้ธุรกิจออนไลน์กลายเป็นสิ่งที่ได้รับความนิยมมากขึ้น ซึ่งจะเห็นได้ชัดในช่วงปี 2563 - 2564 ที่ผ่านมา เป็นช่วงที่ประชาชนต้องเจอกับโรคระบาด Covid-19 ทำให้การใช้ชีวิตประจำวันต้องมีการเปลี่ยนแปลงไป จากการซื้อขายผ่านทางหน้าร้าน กลับกลายเป็นการซื้อขายช่องทางออนไลน์ จึงส่งผลให้อัตราการซื้อขายช่องทางออนไลน์ หรือ E-commerce มีอัตราการเติบโตสูงขึ้นอย่างมีนัยยะ ทำให้หลายธุรกิจเห็นความสำคัญและผลักดันการขายผ่านช่องทางออนไลน์ อีกทั้งประชาชนบางส่วนได้เห็นความสำคัญในการขายออนไลน์เพิ่มขึ้น จึงเป็นเหตุผลที่ทำให้ช่องทางออนไลน์มีร้านใหม่เกิดขึ้นอย่างแพร่หลาย แต่กลับพบว่าก็ไม่ใช่ทุกธุรกิจที่จะประสบความสำเร็จ อย่างบางธุรกิจหากไม่มีการเตรียมพร้อมทางด้านความต้องการของผู้บริโภค ขั้นตอนการผลิต การจัดการคลังสินค้า หรือกลยุทธ์ต่างๆ ที่ใช้ในการดึงดูดผู้บริโภค ก็อาจจะทำให้ธุรกิจนั้นไม่ประสบความสำเร็จ ดังนั้นจากสถานการณ์ที่ไม่แน่นอนที่อาจจะเกิดขึ้นได้ในอนาคต การเตรียมความพร้อมต่างๆ เป็นสิ่งที่สำคัญ หากเจ้าของธุรกิจทราบประมาณการ จำนวนสินค้าที่ผู้บริโภคต้องการหรือยอดขายล่วงหน้าได้ จะทำให้ธุรกิจสามารถรับมือได้กับทุกสถานการณ์ที่เกิดขึ้น และยังทำให้เจ้าของธุรกิจสามารถบริหารงบประมาณที่มีอยู่ให้เกิดประโยชน์สูงสุดกับธุรกิจอีกด้วย

ด้วยเหตุนี้การวางแผนและการจัดการที่ดีจึงเป็นสิ่งที่สำคัญต่อธุรกิจอย่างมาก เนื่องจากการผลิตสินค้าแต่ละชิ้นต่างมีต้นทุนในการผลิต หากมีการสั่งผลิตสินค้าที่มากเกินความต้องการของผู้บริโภค อาจทำให้สินค้าค้างสต๊อกและเงินจมทุนได้ หรืออีกกรณีที่มีการสั่งผลิตสินค้าไม่พอต่อความต้องการของผู้บริโภค อาจทำให้ผู้ประกอบการเสียโอกาสในการขายสินค้าดังกล่าว ด้วยเหตุผลข้างต้นหากมีการพยากรณ์ยอดขายในอนาคตได้อย่างแม่นยำ จะทำให้ผู้ประกอบการสามารถคาดการณ์ความต้องการหรือมูลค่าของสินค้าที่ต้องผลิตได้ ทำให้ธุรกิจสามารถเผชิญกับความผันแปรของความต้องการได้ ส่งผลให้สามารถผลิตสินค้าได้เพียงพอตามความต้องการของผู้บริโภค ไม่มากเกินไปหรือน้อยเกินไป ทำให้ธุรกิจได้กำไรตามที่คาดการณ์ไว้ นอกจากนี้การพยากรณ์ยอดขายยังเปรียบเสมือนเป้าหมายหลักของธุรกิจ ที่ต้องใช้กลยุทธ์ต่างๆเพื่อให้บรรลุเป้าหมาย หรืออาจกล่าวได้ว่าการวางแผนที่ดีจะนำไปสู่การขับเคลื่อนธุรกิจที่ดี

ผู้วิจัยจึงเล็งเห็นปัญหานี้ ดังนั้นจุดมุ่งหมายของการวิจัยเพื่อสร้างแบบจำลองการพยากรณ์ยอดขายสินค้า โดยใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึก Long-Short-Term Memory (LSTM) เปรียบเทียบสมรรถนะกับแบบจำลองโครงข่ายปราสาทเทียมแบบ Multi-Layer Perceptron (MLP) และแบบจำลองอนุกรมเวลาแบบ Seasonal Autoregressive Integrated Moving Averages (SARIMA) และ Seasonal Autoregressive Integrated Moving Averages with Exogenous Regressors (SARIMAX) ในการพยากรณ์ยอดขายสินค้าล่วงหน้า

วิธีดำเนินการ

ขั้นตอนที่ 1 : Data Selection and Understanding

ในงานวิจัยนี้ใช้ข้อมูลธุรกรรมยอดขายจากร้านค้าแห่งหนึ่ง ที่ขายบนแพลตฟอร์มออนไลน์ โดยมีข้อมูลตั้งแต่เดือนมกราคม 2564 – มิถุนายน 2566 โดยมี 1 ตาราง เป็นรายการธุรกรรมการขาย มีคอลัมน์ทั้งหมด 24 คอลัมน์ และมีจำนวนข้อมูล 62,632 แถว ดังตารางที่ 1

ตารางที่ 1 แสดงตัวแปรของข้อมูลรายการธุรกรรมการขาย

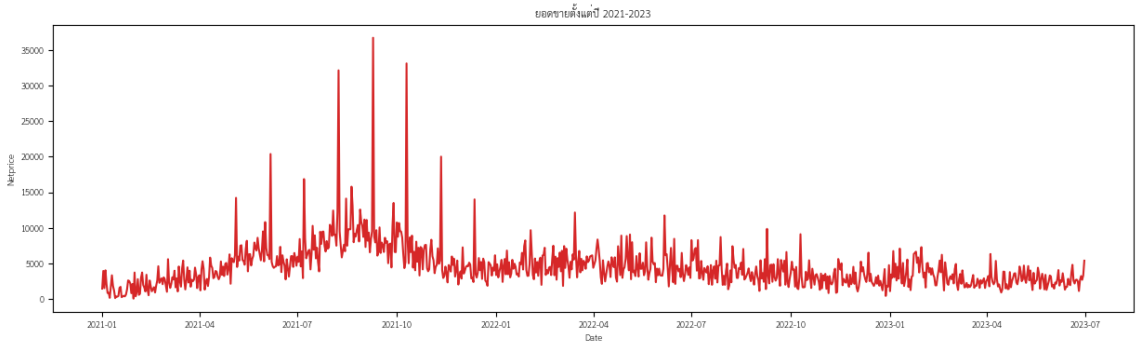
Variable	Description
Order_id	เลขที่ออเดอร์
Status	สถานะคำสั่ง
customer_name	ชื่อลูกค้า
Date	วันที่ที่ทำการชำระเงิน
channel	ช่องทางการชำระเงิน เช่น Online Payment
payment	ช่องทางการชำระเงิน เช่น iBanking Payment (SCB)
Shipping	ชื่อขนส่ง
Type_of_Delivery	ประเภทการจัดส่ง เช่น pick up = เจ้าหน้าที่เข้ามารับสินค้า , Drop off = นำสินค้าไปส่งที่ขนส่ง
Tracking_Number	เลขติดตามพัสดุ
Expected_delivery_date	วันที่คาดหวังว่าจะจัดส่งสินค้า
Delivery_time	วันเวลาที่จัดส่งสินค้า
option	ชื่อสินค้าที่โชว์บนโพสหน้าแอพพลิเคชั่น shopee
category	หมวดหมู่สินค้า
sub_category	หมวดหมู่ย่อยของสินค้า
product	ชื่อสินค้าที่เป็นตัวเลือก
ราคาตั้งต้น	ราคาตั้งต้น
Sale_price	ราคาขาย
quantity	จำนวนสินค้าที่ขาย
Netprice	ราคาสุทธิที่ได้รับ
Discount_by_seller	ส่วนลดที่ออกโดยร้านค้า (เป็น code ที่ร้านค้าให้เก็บ)
Discount_by_Shopee	ส่วนลดที่ออกโดย shopee
CodeDiscount	รหัส code ส่วนลด

Transaction Fee	ค่าธรรมเนียม

ขั้นตอนที่ 2 : Exploratory Data Analysis : EDA

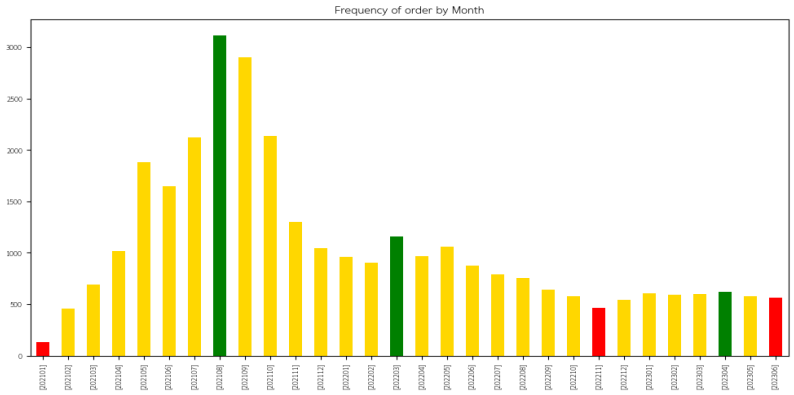
ในงานวิจัยนี้ใช้ภาษาไพทอนในการวิเคราะห์ข้อมูลและการเรียนรู้ของเครื่อง เริ่มต้นด้วยการนำเข้าโมดูลสำคัญสำหรับการสร้างแบบจำลอง ต่อมานำเข้าไฟล์ข้อมูลและข้อมูลที่ใช้สำหรับสร้างแบบจำลอง เบื้องต้นเพื่อหาข้อมูลเชิงลึกจากข้อมูลธุรกรรมการขายโดยใช้ไลบรารี่ Pandas, Numpy, Matplotlib และ seaborn จากการสำรวจข้อมูล ข้อมูลชุดนี้ไม่มีค่า Missing Value

ทำการวิเคราะห์ยอดขาย จากภาพประกอบที่ 1 จะเห็นว่าในปี 2564 มียอดขายที่สูงมาก เนื่องจากปี 2564 มีสถานการณ์โควิด-19 จึงทำให้ผู้คนอยู่บ้านและหันมาซื้อของออนไลน์กันมากขึ้น จึงทำให้ยอดขายสูงโดดขึ้นมา ต่อมาในปี 2565-2566 ยอดขายค่อย ๆ ลดลงมาเรื่อย ๆ เนื่องจากสถานการณ์โควิดเริ่มคลี่คลาย ทำให้ผู้คนเริ่มออกไปใช้ชีวิตตามปกติ ซึ่งยอดขายเพิ่มขึ้นและลดลงอย่างรวดเร็ว ทำให้สั่งผลิตสินค้าให้อยู่ในความต้องการของลูกค้าได้ไม่เหมาะสม บางครั้งสั่งผลิตสินค้ามามากเกินความต้องการของลูกค้า และบางครั้งผลิตสินค้ามาน้อยเกินความต้องการของลูกค้า ทำให้สินค้าขาดสต๊อก และเสียโอกาสในการขายสินค้านั้น



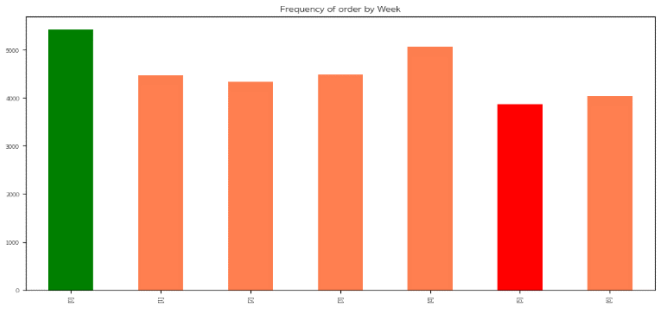
ภาพประกอบ 1 แสดงยอดขายตั้งแต่เดือน มกราคม 2564 – มิถุนายน 2566

ทำการวิเคราะห์จำนวนออเดอร์ในแต่ละเดือนจำแนกตามปี จะเห็นว่าในปี 2564 เดือนสิงหาคมเป็นเดือนที่ขายดีที่สุด และเดือนมกราคมเป็นเดือนที่ขายได้น้อยที่สุดเนื่องจากเดือนมกราคม 2564 เป็นเดือนที่พึ่งเปิดร้านเป็นเดือนแรก ต่อมาในปี 2565 เดือนมีนาคมเป็นเดือนที่ขายดีที่สุด และเดือนพฤศจิกายนเป็นเดือนที่ขายได้น้อยที่สุด และในปี 2566 เดือนเมษายนเป็นเดือนที่ขายดีที่สุด และเดือนมิถุนายนเป็นเดือนที่ขายได้น้อยที่สุด



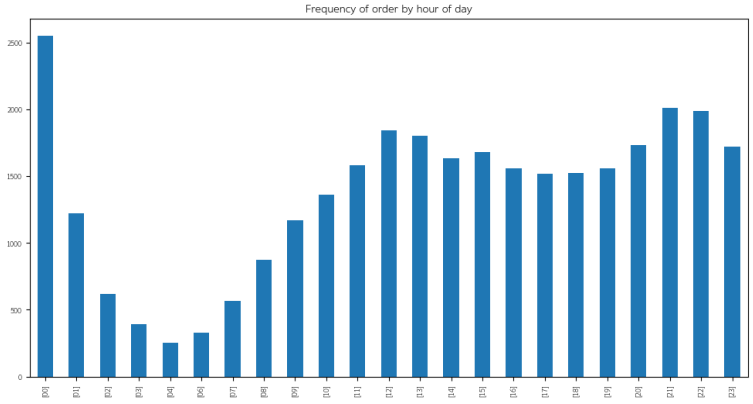
ภาพประกอบ 2 แสดงจำนวนออเดอร์ในแต่ละเดือนและปี

ทำการวิเคราะห์จำนวนออเดอร์แยกเป็นรายวัน จะเห็นว่าวันที่ขายดีที่สุดจะเป็นวันอาทิตย์ เนื่องจากเป็นวันหยุดสุดสัปดาห์ และวันที่ขายได้น้อยที่สุดจะเป็นวันศุกร์ ดังภาพประกอบที่ 3



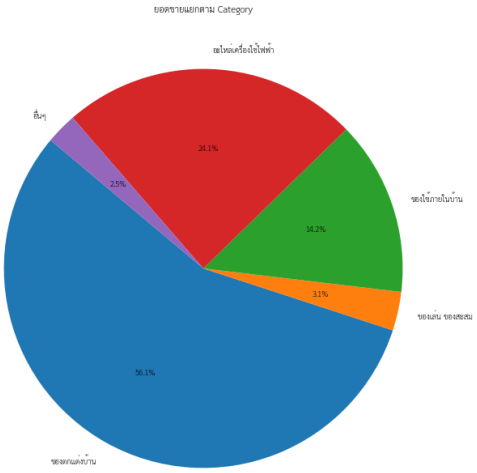
ภาพประกอบ 3 แสดงจำนวนออเดอร์ในแต่วัน

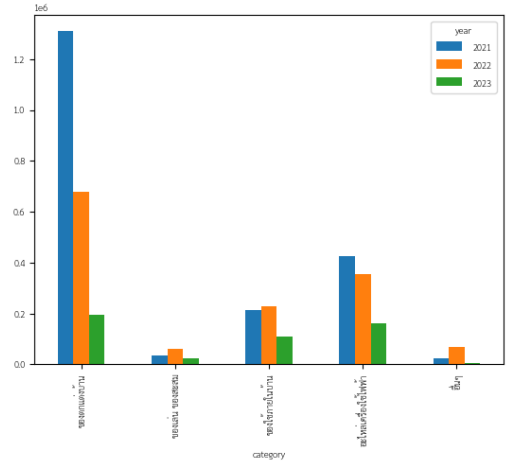
ทำการวิเคราะห์จำนวนออเดอร์ตามช่วงเวลา จะเห็นว่าช่วงเวลาที่ขายดีที่สุดจะเป็นช่วงเวลา 00.00 น. เนื่องจากเป็นช่วงเวลาที่ทางแพลตฟอร์มออนไลน์ จะปล่อยโค้ดส่วนลด หรือ โค้ดส่งฟรีขั้นต่ำ 0 บาท จึงทำให้ช่วงเวลานี้สามารถขายของได้มากที่สุด ดังภาพประกอบที่ 4



ภาพประกอบ 4 แสดงจำนวนออเดอร์ตามช่วงเวลา

ทำการวิเคราะห์ว่าหมวดหมู่ไหนที่ขายดีที่สุดในร้าน จะเห็นว่าหมวดหมู่ของตกแต่งบ้านขายได้ 56.1% ซึ่งเป็นหมวดหมู่ที่ขายดีที่สุดในร้าน รองลงมาเป็นอะไหล่เครื่องใช้ไฟฟ้า 24.1% ของใช้ภายในบ้าน 14.2% ของเล่นของสะสม 3.1% และอื่นๆ 2.5% ตามลำดับ ดังภาพประกอบที่ 5 และจากภาพที่ 6 จะเห็นว่าหมวดหมู่ของตกแต่งบ้าน ขายได้มากที่สุดในทุก ๆ ปี

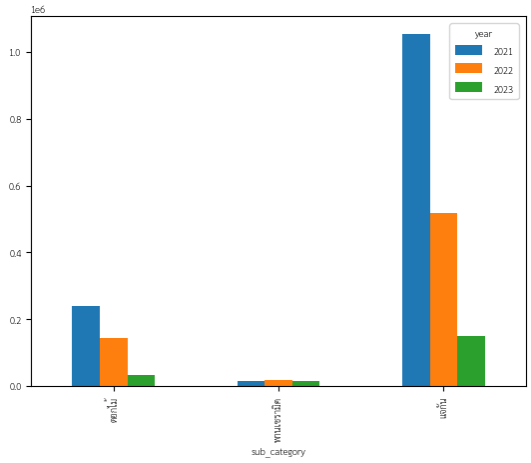


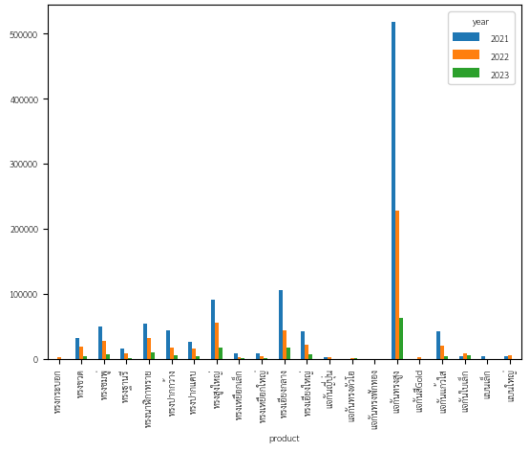


ภาพประกอบ 5 แสดงยอดขายแยกตามหมวดหมู่

ภาพประกอบ 6 แสดงยอดขายแยกตามหมวดหมู่และปี

ทำการวิเคราะห์ว่าหมวดหมู่ของตกแต่งบ้านสินค้าตัวไหนที่ขายดีที่สุด จะเห็นว่าสินค้าแจกันเป็นสินค้าที่ขายดีที่สุดในทุก ๆ ปี ดังภาพประกอบที่ 7 ทำการวิเคราะห์สินค้าประเภทแจกัน เนื่องจากแจกันขายดีที่สุดในหมวดหมู่ของตกแต่งบ้าน จะเห็นว่าแจกันที่ขายดีที่สุดและยอดขายโดดออกมา คือ แจกันทรงสูง แจกันทรงเอียงกลาง และแจกันทรงสูงใหญ่ ตามลำดับ ดังภาพประกอบที่ 8





ภาพประกอบ 7 แสดงยอดขายสินค้าในหมวดหมู่ของตกแต่งบ้าน

ภาพประกอบ 8 แสดงยอดขายสินค้าในสินค้าประเภทแจกัน

ขั้นตอนที่ 3 : Data Preparation and Cleaning

การเตรียมข้อมูลสำหรับการฝึกฝนของแบบจำลอง พบว่าหมวดหมู่ที่ขายดีที่สุดคือหมวดหมู่ของตกแต่งบ้าน และสินค้าที่ขายดีที่สุดในหมวดหมู่นี้คือสินค้าประเภทแจกัน แจกันที่ขายดีที่สุด 3 อันดับแรก คือ แจกันทรงสูง แจกันทรงเอียงกลาง และแจกันทรงสูงใหญ่ ตามลำดับ และจากการสำรวจข้อมูล พบว่าช่วงเวลาที่ขายดีที่สุด คือช่วงเวลา 00.00 น. ซึ่งเป็นช่วงเวลาที่ทาง Shopee ได้ทำการปล่อยโค้ดส่วนลด และโค้ดส่งฟรีออกมาในวันโปรโมชั่นต่างๆ ดังนั้นผู้วิจัยจึงจะทำการแปลงข้อมูลในคอลัมน์ CodeDiscount ให้เป็น 0 และ 1 และทำการเพิ่มฟีเจอร์วันโปรโมชั่นเข้ามาในชุดข้อมูล

เนื่องด้วยข้อมูลในคอลัมน์ CodeDiscount เป็นข้อมูลรหัสส่วนลดที่ลูกค้าใช้ซึ่งอยู่ในรูปแบบที่ไม่พร้อมใช้งาน จึงจำเป็นต้องแปลงข้อมูลในคอลัมน์นี้ โดยหากในคอลัมน์นี้มีการใช้รหัสส่วนลด จะให้เท่ากับ 1 และหากไม่มีการใช้รหัสส่วนลดจะให้เท่ากับ 0 ดังภาพประกอบที่ 9

	Order_id	Status	customer_name	Date	channel	payment	Shipping	Type_of_Delivery	Tracking_Number	Expected_delivery_date	...	CodeDiscount
0	210511BW0C241P	สำเร็จแล้ว	toeyyyyyyyyyyyy	2021-05-11 12:41:00	เก็บเงินปลายทาง	Cash on Delivery	Standard Delivery - ส่งธรรมดาในประเทศ-Shopee X...	pickup	TH019606788965	2021-05-13 12:41:00	...	0
1	210511B6AXWF7Q	สำเร็จแล้ว	ily_c	2021-05-11 05:40:00	ชำระผ่าน ATM	Bank Transfer	Standard Delivery - ส่งธรรมดาในประเทศ-Shopee X...	pickup	TH016025641665	2021-05-13 05:40:00	...	0
2	210511CQFYFH5T	สำเร็จแล้ว	poiipraewx	2021-05-11 20:06:00	Mobile Banking	iBanking Payment V2 (K Plus)	Standard Delivery - ส่งธรรมดาในประเทศ-Shopee X...	pickup	TH011646100865	2021-05-13 20:06:00	...	0
3	210511CWA5AAFC	สำเร็จแล้ว	sattawatkhiaosod	2021-05-11 22:11:00	ชำระผ่าน ATM	Bank Transfer	Standard Delivery - ส่งธรรมดาในประเทศ-Kerry	pickup	SHP5071359736	2021-05-13 22:11:00	...	0
4	210512EES2QMBR	สำเร็จแล้ว	jabbss	2021-05-12 12:43:00	ชำระผ่าน ATM	Bank Transfer	Standard Delivery - ส่งธรรมดาในประเทศ-Shopee X...	pickup	TH011544088665	2021-05-14 12:43:00	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
62626	220703F88MMX1W	สำเร็จแล้ว	kanyapach_z	2022-07-03 12:11:00	ShopeePay	Airpay Wallet V2	Standard Delivery - ส่งธรรมดาในประเทศ-Kerry	pickup	SHP5155381703	2022-07-05 12:11:00	...	1
62627	220703F88MMX1W	สำเร็จแล้ว	kanyapach_z	2022-07-03 12:11:00	ShopeePay	Airpay Wallet V2	Standard Delivery - ส่งธรรมดาในประเทศ-Kerry	pickup	SHP5155381703	2022-07-05 12:11:00	...	1
62628	220722524QMVC6	สำเร็จแล้ว	benjamaswaew	2022-07-22 14:59:00	เก็บเงินปลายทาง	Cash on Delivery	Standard Delivery - ส่งธรรมดาในประเทศ-Kerry	pickup	SHP5160787905	2022-07-25 14:59:00	...	1
62629	2208132EED016P	สำเร็จแล้ว	7to0jemj_b	2022-08-13 17:33:00	เก็บเงินปลายทาง	Cash on Delivery	Standard Delivery - ส่งธรรมดาในประเทศ-Kerry	pickup	SHP5166622534	2022-08-16 17:33:00	...	1
62630	2208132EED016P	สำเร็จแล้ว	7to0jemj_b	2022-08-13 17:33:00	เก็บเงินปลายทาง	Cash on Delivery	Standard Delivery - ส่งธรรมดาในประเทศ-Kerry	pickup	SHP5166622534	2022-08-16 17:33:00	...	1

ภาพประกอบ 9 แสดงคอลัมน์ CodeDiscount ที่ทำการแปลงข้อมูล

ทำการเพิ่มคอลัมน์ promotion_day เป็นข้อมูลวันโปรโมชั่น ซึ่งวันโปรโมชั่นจะมีทั้งหมด 3 วันต่อเดือน ได้แก่ 1. วัน Double Day (เป็นวันที่และเดือนที่เหมือนกัน เช่น วันที่ 2 เดือน 2 , วันที่ 3 เดือน 3 เป็นต้น) 2. วัน Mid Month Sale (เป็นวันที่ 15 ของทุกเดือน) 3. วัน Shopee Pay (เป็นวันที่ 25 ของทุกเดือน) โดยหากวันไหนเป็นวันโปรโมชั่นจะให้เท่ากับ 1 และวันที่ไม่ใช่วันโปรโมชั่นจะให้เท่ากับ 0 และทำการเพิ่มคอลัมน์ SumPromotionDay เป็นการรวมโปรโมชั่นทั้งหมด โดยเอา CodeDiscount (รหัสส่วนลด) และ promotion_day (วันโปรโมชั่น) มารวมกัน

เนื่องจากผู้วิจัยต้องการพยากรณ์ยอดขายเป็นรายสัปดาห์ เพราะทางร้านสั่งผลิตสินค้าเป็นรายสัปดาห์ ดังนั้นผู้วิจัยจึงทำการเพิ่มคอลัมน์ End_of_week ดังภาพประกอบที่ 10

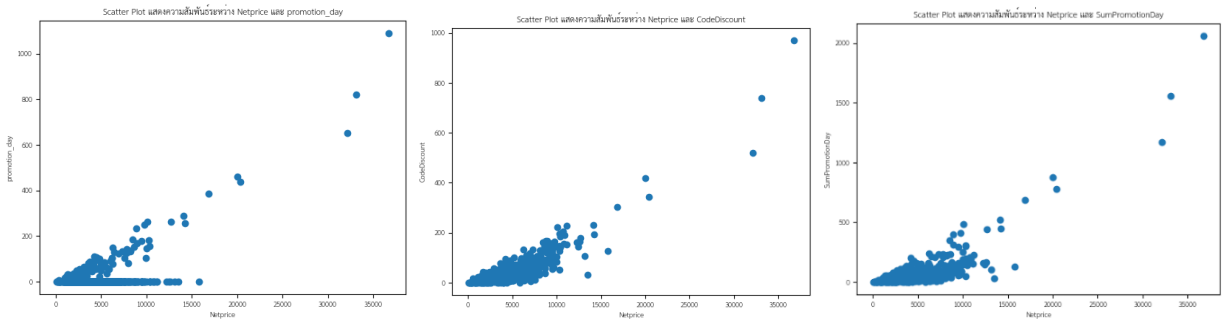
	Order_id	Status	customer_name	Date	End_of_Week
0	210511BW0C241P	สำเร็จแล้ว	toeyyyyyyyyyyy.	2021-05-11	2021-05-16
1	210511B6AXWF7Q	สำเร็จแล้ว	ily_c	2021-05-11	2021-05-16
2	210511CQFYFH5T	สำเร็จแล้ว	poiipraewx	2021-05-11	2021-05-16

ภาพประกอบ 10 แสดงคอลัมน์ End_of_week ที่ทำการเพิ่มเข้ามา

ทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร (Correlations) จากภาพประกอบที่ 11 และ 12 จะเห็นว่ายอดขายสินค้ากับวันโปรโมชั่นมีความสัมพันธ์กัน จะเห็นว่า CodeDiscount (รหัสโค้ดส่วนลด) มีความสัมพันธ์สูงถึง 0.84 ซึ่งมีความสัมพันธ์กันมากกว่า promotion_day เนื่องจากโดยส่วนใหญ่ ลูกค้ามักจะใช้โค้ดส่งฟรีกับสินค้าภายในร้าน เนื่องจากสินค้าภายในร้านจะมีค่าเฉลี่ยอยู่ที่ 55 บาท จึงทำให้ใช้โค้ดส่งฟรีคุ้มกว่า ซึ่งในวันโปรโมชั่น ทาง Shopee มักจะออกโค้ดส่วนลดที่ลดสินค้าได้เยอะ ลูกค้ามักจะนำโค้ดไปใช้กับสินค้าที่ราคาสูงในวันโปรโมชั่น

	Netprice	CodeDiscount	promotion_day	SumPromotionDay
Netprice	1.000000	0.844694	0.665509	0.785371
CodeDiscount	0.844694	1.000000	0.824138	0.949803
promotion_day	0.665509	0.824138	1.000000	0.959963
SumPromotionDay	0.785371	0.949803	0.959963	1.000000

ภาพประกอบ 11 แสดงความสัมพันธ์ระหว่างยอดขายกับวันโปรโมชั่น



ภาพประกอบ 12 แสดงความสัมพันธ์ระหว่างยอดขายกับโปรโมชั่น

ขั้นตอนที่ 4 : Data Modeling – Model Selection

ในงานวิจัยนี้ได้ใช้แบบจำลอง SARIMA SARIMAX LSTM และ MLP ในการพยากรณ์ยอดขาย โดยจะทำการพยากรณ์ยอดขายเป็นรายสัปดาห์ ซึ่งจะพยากรณ์ยอดขายของสินค้า 3 รายการที่มียอดขายสูงสุด ได้แก่ แจกันทรงสูง แจกันทรงเอียงกลาง และแจกันทรงสูงใหญ่ โดยทำการเปรียบเทียบสมรรถนะโดยใช้ค่า Root Mean Squared Error (RMSE), ค่า Mean Absolute

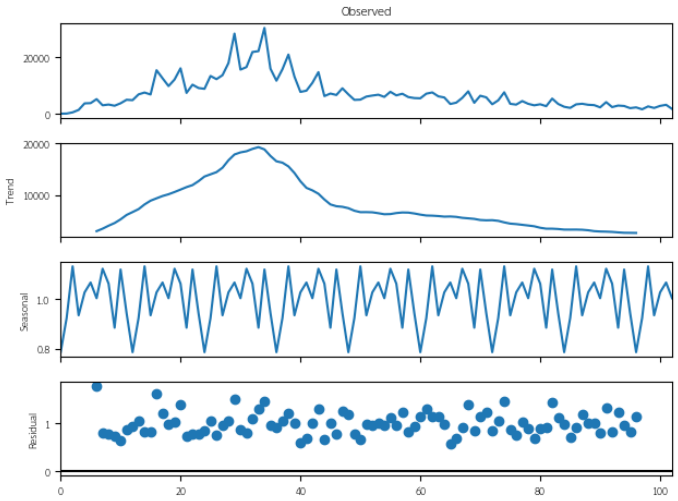
Error (MAE) และ Mean Absolute Percent Error (MAPE) เป็นตัวชี้วัดความแม่นยำของแต่ละแบบจำลอง ซึ่งข้อมูลใน Dataset มีข้อมูลตั้งแต่เดือน มกราคม 2564 - มิถุนายน 2566 โดยผู้วิจัยจะทำการแบ่งข้อมูลเป็น train 80% และ test 20%

ซึ่งจากการวิเคราะห์ความสัมพันธ์ระหว่างยอดขายกับวันโปรโมชั่น จะเห็นว่า 2 ฟีเจอร์นี้มีความสัมพันธ์กัน ดังนั้นในงานวิจัยนี้จะใช้ข้อมูล feature SumPromotion เป็นข้อมูล input เพื่อช่วยเพิ่มประสิทธิภาพในการพยากรณ์ยอดขายในแบบจำลอง SARIMAX

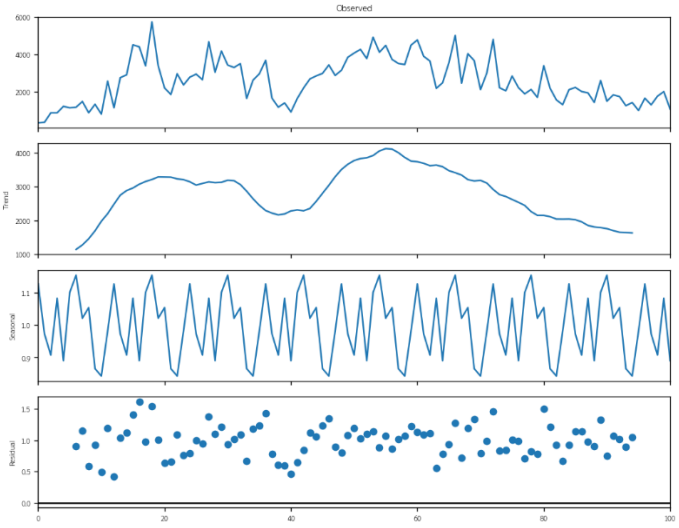
4.1 แบบจำลอง SARIMA และ SARIMAX

หลักการทำงานของ SARIMA และ SARIMAX คือ จะนำข้อมูลที่ต้องการพยากรณ์มาใช้ในการ train โดยแบบบจำลอง SARIMA และ SARIMAX เหมาะกับข้อมูลที่มีฤดูกาล และข้อจำกัดของแบบจำลองนี้ข้อมูลจะต้องมีความคงที่ (Stationary) โดยที่แบบจำลอง SARIMAX จะเพิ่ม Exogenous variable (exog) ซึ่งเป็นตัวแปรวันโปรโมชั่นเข้าไปในแบบจำลอง

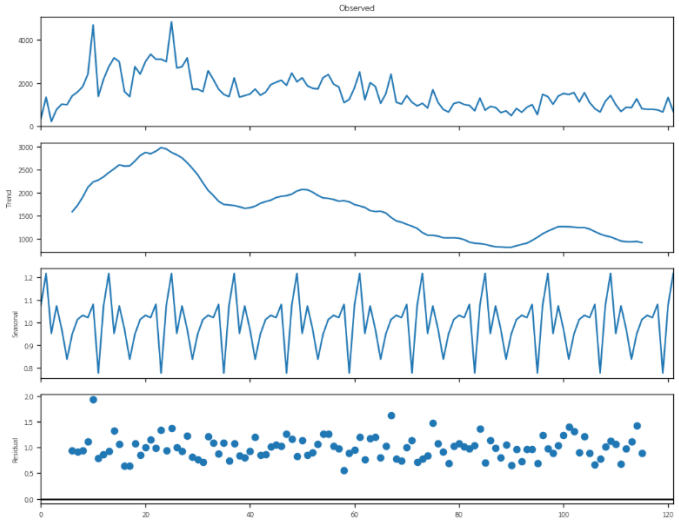
นำข้อมูลยอดขายที่ทำการแบ่งเป็น train set ของสินค้าประเภทแจกันทรงสูง แจกันทรงเอียงกลาง และแจกันทรงสูงใหญ่ มาตรวจสอบดูค่า Trend Seasonal และ Residual จากภาพประกอบที่ 13 - 15 จะเห็นได้ว่าข้อมูลยอดขายนั้นมี Trend ขึ้นและลง และมีความเป็น Seasonal



ภาพประกอบ 13 แสดงค่า Trend Seasonal และ Residual ของ train data แจกันทรงสูง



ภาพประกอบ 14 แสดงค่า Trend Seasonal และ Residual ของ train data แจกันทรงเอียงกลาง



ภาพประกอบ 15 แสดงค่า Trend Seasonal และ Residual ของ train data แจกันทรงสูงใหญ่

จากข้อมูลสินค้าประเภทแจกันทั้ง 3 ประเภท ยังมียอดขายที่ไม่คงที่ (Non-Stationary) ดังนั้นจึงต้องทำการ Differencing เพื่อทำให้ข้อมูลเป็น stationary โดยในงานวิจัยนี้จะใช้สถิติทดสอบ Augmented Dickey-Fuller test (ADF-Test) เทียบกับค่าวิกฤต (ค่าวิกฤต = 0.05) หากค่า p-value < 0.05 แสดงว่าข้อมูลมีความคงที่ (Stationary)

จากนั้นใช้ Grid Search ในการ Tuning Parameter เพื่อหาค่า p,d,q ที่ดีที่สุด ซึ่งดูได้จากค่า Akaike Information Criterion (AIC) ที่ต่ำที่สุด โดยกำหนดให้ p = range (1,3) , d = 1 (เนื่องจากทำการ diff 1 ครั้งก็สามารถทำข้อมูลให้เป็น stationary) , q = rang (1,3) และ s = 12 เนื่องจากดูจากกราฟ Seasonal แล้วชุดข้อมูลนี้มี Seasonal เป็นรายไตรมาส ดังนั้น 1 ไตรมาส คือ 12 week ซึ่งได้ค่าพารามิเตอร์ที่ดีสุด ดังตารางที่ 2

สินค้า	(p,d,q) x (P,D,Q,s)
แจกันทรงสูง	(1,1,1) x (1,1,1,12)
แจกันทรงเอียงกลาง	(2,1,1) x (1,1,1,12)
แจกันทรงสูงใหญ่	(1,1,1) x (1,1,1,12)

ตารางที่ 2 แสดงค่าพารามิเตอร์ที่ใช้ของสินค้าแต่ละประเภทใน SARIMA และ SARIMAX

นำค่าพารามิเตอร์ที่ได้ไป fit กับแบบจำลอง SARIMA และ แบบจำลอง SARIMAX โดยทำการเพิ่ม Exogenous variable (exog) ซึ่งเป็นตัวแปรวันโปรโมชั่นเข้าไปในแบบจำลอง SARIMAX โดยกำหนดตัวแปรเป็น SumPromotionDay ทำการพยากรณ์ยอดขายใน week ถัดไป แล้วนำข้อมูลที่ได้จากการพยากรณ์มาเปรียบเทียบกับ test (ข้อมูลจริง) และนำค่าที่ได้มาหาค่า RMSE, MAPE และ MAE เพื่อวัดประสิทธิภาพของแบบจำลอง

4.2 แบบจำลอง LSTM

ในการพยากรณ์ยอดขายสินค้า ของแบบจำลอง LSTM จะทำการแบ่งข้อมูลออกเป็น 2 ชุด คือ train 80% และ test 20% แล้วจึงแบ่งชุดข้อมูลของ train และ test ออกเป็นข้อมูลที่จะใช้ในการ input และ output ของแบบจำลอง โดยแปลงค่าของชุดข้อมูลให้อยู่ในรูปแบบของอาร์เรย์ 3 มิติ ซึ่งถูกแบ่งออกเป็น [sample, timesteps, feature] เพื่อใช้ในแบบจำลอง LSTM แล้วทำการ tuning parameter เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด

และทำการกำหนด Timestep = 7 เข้ามาช่วยในการขยับข้อมูลย้อนหลังหรือขยับข้อมูลไปข้างหน้า โดยที่แบบจำลอง LSTM จะใช้ข้อมูลแบบ sequence ซึ่งต้องการการจัดเรียงให้เป็นลำดับของข้อมูลตามเวลา เพื่อให้โมเดลเรียนรู้และเข้าใจลำดับของข้อมูลได้ ซึ่งข้อมูล input ที่ใช้ในแบบจำลองนี้ จะใช้ข้อมูลยอดขายของวันก่อนหน้าเพื่อใช้ในการพยากรณ์ ส่วนข้อมูล output คือข้อมูลที่ต้องการพยารกรณ์นั้นได้กำหนดค่าเป็น 19 เพราะต้องการจะพยากรณ์ยอดขายทั้งหมด 19 week

ซึ่งได้ค่าพารามิเตอร์ที่ดีที่สุด คือ epochs = 300 ค่า hidden neuron (n_node) = 16 โดยกำหนดค่า activation function = relu และ optimize = adam เพื่อเพิ่มประสิทธิภาพในการช่วยทำนายผลของแบบจำลอง แล้วจึงทำการ fit ข้อมูลกับแบบจำลองเพื่อให้แบบจำลองนั้นทำการพยากรณ์ยอดขายไปอีก 19 week ข้างหน้า

4.3 แบบจำลอง MLP :

ในการพยากรณ์ยอดขายสินค้า ของแบบจำลอง MLP จะทำการแบ่งข้อมูลออกเป็น 2 ชุด คือ train 80% และ test 20% แล้วจึงแบ่งชุดข้อมูลของ train และ test ออกเป็นข้อมูลที่จะใช้ในการ input และ output ของแบบจำลอง โดยแปลงค่าของชุดข้อมูลให้อยู่ในรูปแบบของอาร์เรย์ 3 มิติ ซึ่งถูกแบ่งออกเป็น [sample, timesteps, feature] เพื่อใช้ในแบบจำลอง MLP แล้วทำการ tuning parameter เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด โดยแบบจำลอง MLP จะประกอบด้วยเลเยอร์หลายชั้น (multi-layer) โดยแต่ละเลเยอร์จะประกอบด้วยโหนด (nodes) ซึ่งแต่ละโหนดจะเชื่อมต่อกับโหนดในเลเยอร์ถัดไป

ทำการกำหนด Timestep = 7 คือจำนวนข้อมูลย้อนหลังที่จะนำมาใช้ในการทำนายข้อมูลปัจจุบันในแต่ละชุดข้อมูล ซึ่งข้อมูล input ที่ใช้ในแบบจำลองนี้ จะใช้ข้อมูลยอดขายของวันก่อนหน้าเพื่อใช้ในการพยากรณ์ ส่วนข้อมูล output คือข้อมูลที่ต้องการพยารกรณ์นั้นได้กำหนดค่าเป็น 19 เพราะต้องการจะพยากรณ์ยอดขายทั้งหมด 19 week

ซึ่งได้ค่าพารามิเตอร์ที่ดีที่สุด คือ มี 2 Layer โดยมีจำนวนโหนด (n_node) ในแต่ละเลเยอร์คือ 16 และ 1 ตามลำดับ และ epochs = 300 โดยกำหนดค่า activation function = relu และ optimize = adam เพื่อเพิ่มประสิทธิภาพในการช่วยทำนายผลของแบบจำลอง แล้วจึงทำการ fit ข้อมูลกับแบบจำลองเพื่อให้แบบจำลองนั้นทำการพยากรณ์ยอดขายไปอีก 19 week ข้างหน้า

ขั้นตอนที่ 5 : Model Evaluation

ในการประเมินวัดประสิทธิภาพของแบบจำลอง ในงานวิจัยนี้จะใช้ค่า Root Mean Squared Error (RMSE) , ค่า Mean Absolute Error (MAE) และ Mean Absolute Percent Error (MAPE) ในการวัดค่าความแม่นยำ

5.1 รากที่สองของค่าเฉลี่ยความผิดพลาดกำลังสอง (Root Mean Squared Error : RMSE) คือ การวัดความแตกต่างระหว่างค่าจริงและค่าพยากรณ์ ถ้าหากค่า RMSE นั้นมีค่าน้อย แสดงว่าค่าพยากรณ์นั้นประมาณค่าได้ใกล้เคียงกับค่าจริง ดังสมการ

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

โดยที่

Y_i คือ ค่า actual value ของ samples test ที่ i

$\hat{Y}_i$ คือ ค่า predict ของ samples test ที่ i

n คือ จำนวน samples ทั้งหมด

5.2 ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) คือ การหาค่าเฉลี่ยของความแตกต่างสมบูรณ์ระหว่างค่าพยากรณ์และค่าจริง ถ้าหากค่า MAE นั้นมีค่าน้อย แสดงว่าค่าทำนายนั้นมีค่าใกล้เคียงกับค่าจริง ดังสมการ

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

โดยที่

Y_i คือ ค่าจริงหรือ ค่า *actual value* ของ *samples test* ที่ *i*

$\hat{Y}_i$ คือ ค่าที่ประมาณได้หรือ ค่า *predict* ของ *samples test* ที่ *i*

n คือ จำนวน *samples* ทั้งหมด

5.3 ค่าเฉลี่ยความผิดพลาดร้อยละสัมบูรณ์ (Mean Absolute Percent Error : MAPE) คือ การหาเปอร์เซ็นค่าเฉลี่ยของความแตกต่างสมบูรณ์ระหว่างค่าพยากรณ์และค่าจริง ถ้าหากค่า MPAE นั้นมีค่าน้อย แสดงว่าค่าพยากรณ์นั้นประมาณค่าได้ใกล้เคียงกับค่าจริงดัง สมการ

$$\text{MAPE} = \frac{\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_t}\right| \times 100}{N}$$

โดยที่

Y_i คือ ค่า actual value ของ samples test ที่ i

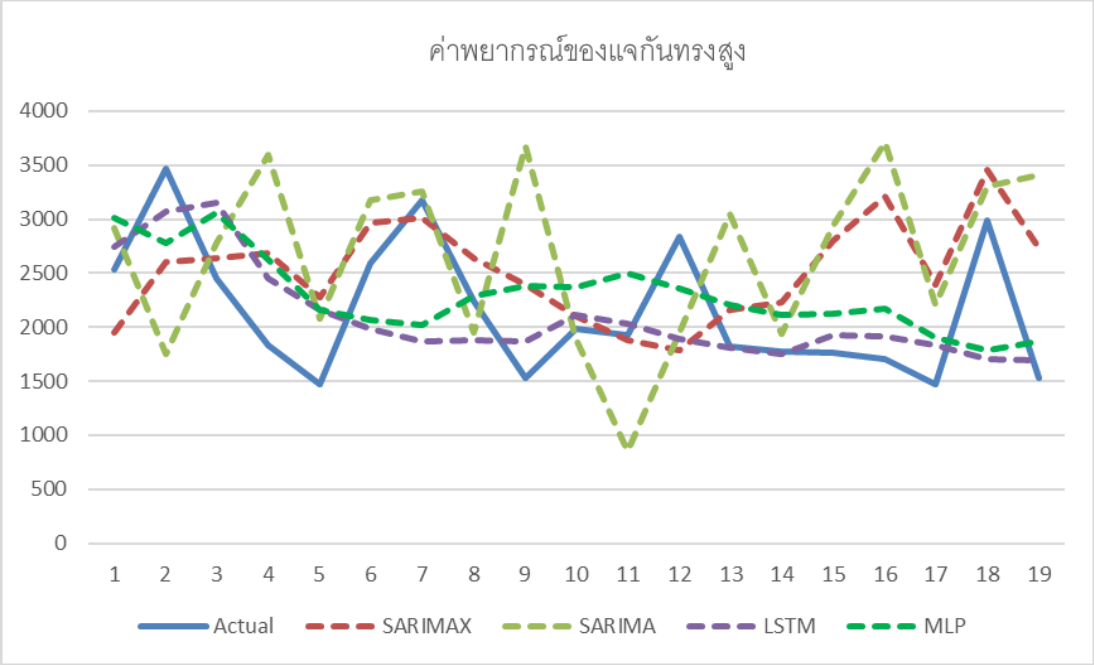
$\hat{Y}_i$ คือ ค่า predict ของ samples test ที่ i

n คือ จำนวน samples ทั้งหมด

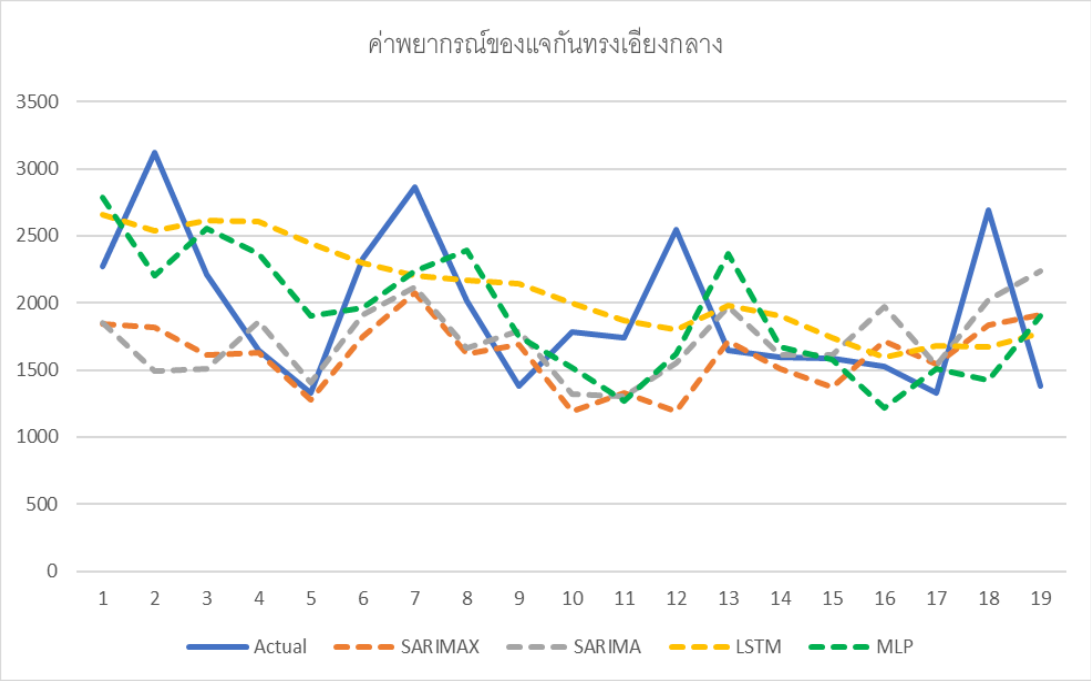
ถ้าค่า MAPE น้อยกว่า 10% จัดว่าการพยากรณ์ค่อนข้างแม่นยำ ถ้าค่า MAPE อยู่ระหว่าง 10% ถึง 20% จัดว่าการพยากรณ์ใช้ได้ดี ถ้าค่า MAPE อยู่ระหว่าง 20% ถึง 50% จัดว่าการพยากรณ์พอใช้ ถ้าค่า MAPE มากกว่า 50% จัดว่าการพยากรณ์ไม่แม่นยำ และถ้าค่า MAPE น้อยกว่า 10% จัดว่า การพยากรณ์ค่อนข้างแม่นยำ (Taha Falatouria & Farzaneh Darbaniana & Patrick Brandtnera & Chibuzor Udokwua, 2022)

ผลการวิจัยและอภิปรายผลการวิจัย

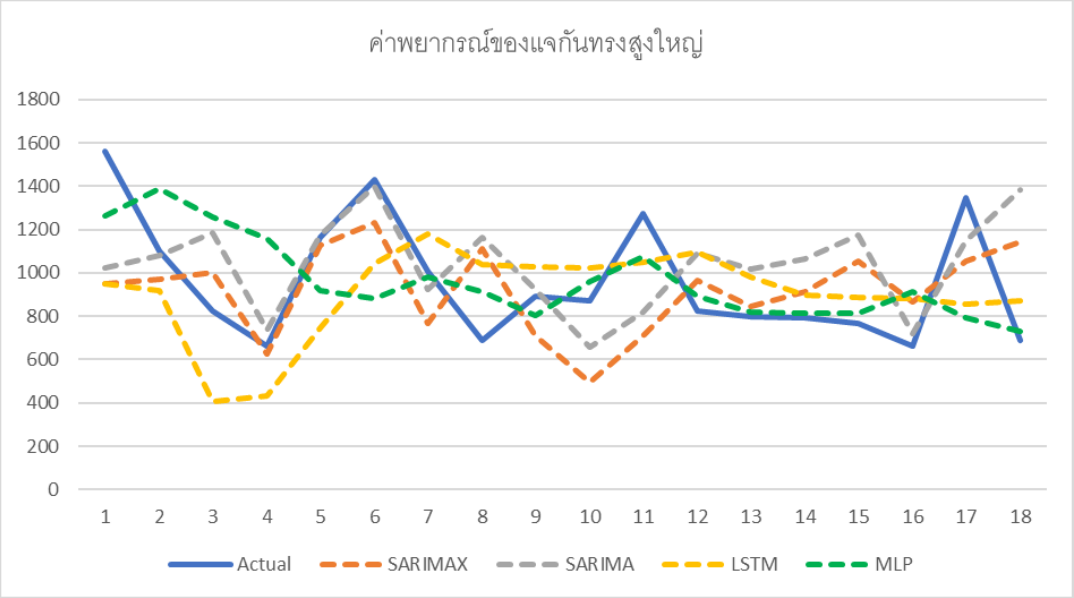
ในการวิจัยนี้ ได้ทำการศึกษาและวิเคราะห์เปรียบเทียบสมรรถนะระหว่างแบบจำลอง SARIMA SARIMAX LSTM และ MLP โดยผู้วิจัยได้เลือกศึกษาข้อมูลการขายสินค้าจากร้านค้าแห่งหนึ่ง ที่ขายบนแพลตฟอร์มออนไลน์ ซึ่งมีข้อมูลยอดขายรายวันตั้งแต่เดือน มกราคม 2564 ถึง มิถุนายน 2566 โดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อเป็นแนวทางในการพัฒนาการสร้างแบบจำลองในการพยากรณ์ยอดขายสินค้าในอีก 19 สัปดาห์ข้างหน้า โดยผลลัพธ์ที่ได้จากการทดลองด้วยแบบจำลอง SARIMA SARIMAX LSTM และ MLP ของสินค้าทั้ง 3 รายการ นำค่าพยากรณ์ที่ได้มาเปรียบเทียบกับยอดขายจริง ด้วยการพล็อตกราฟ ดังภาพประกอบที่ 16 - 18



ภาพประกอบ 16 แสดงการเปรียบเทียบการพยากรณ์ของแต่ละแบบจำลองของแจกันทรงสูง



ภาพประกอบ 17 แสดงการเปรียบเทียบการพยากรณ์ของแต่ละแบบจำลองของแจกันทรงเอียงกลาง



ภาพประกอบ 18 แสดงการเปรียบเทียบการพยากรณ์ของแต่ละแบบจำลองของแจกันทรงสูงใหญ่

ผู้วิจัยได้วัดประสิทธิภาพของแบบจำลองแต่ละอัลกอริทึมเพื่อนำมาเปรียบเทียบและสรุปผล ซึ่งในการวิจัยนี้ได้สร้างแบบจำลอง SARIMA SARIMAX LSTM และ MLP เพื่อใช้ในการพยากรณ์ยอดขายล่วงหน้า ดังนั้นจึงเลือกใช้ Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) และ Mean Absolute Percent Error (MAPE) มาเป็นตัววัดประสิทธิภาพของแบบจำลอง หรือการหาค่า error ซึ่งได้จากการหาค่าความแตกต่างระหว่างค่าที่ได้จากการพยากรณ์กับค่าจริงแล้วนำไปหาค่าเฉลี่ย โดย RMSE แตกต่างกับ MAE ตรงที่ RMSE จะนำผลรวมค่าเฉลี่ย error ที่ได้มายกกำลังสองแล้วถอดรากที่สอง ส่วน MAE จะนำผลรวมค่าเฉลี่ยของ error ที่ได้มาถอดค่าสัมบูรณ์ และ MAPE จะวัดความแม่นยำของการพยากรณ์โดยการหาค่าเฉลี่ยของความแตกต่างระหว่างค่าพยากรณ์และค่าจริงเป็นเปอร์เซ็นต์ โดยที่ค่า RMSE จะมีผลต่อค่า error ที่ค่อนข้างสูงหรือเป็น outlier ซึ่งได้จากการพยากรณ์มากกว่า MAE เนื่องจาก RMSE มีขั้นตอนการคำนวณที่นำค่าผิดพลาดแต่ละค่าไปยกกำลังสอง และหลังจากนั้นนำมาหาค่าเฉลี่ย ส่งผลให้ค่าผิดพลาดที่มีความผิดพลาดมาก ๆ มีน้ำหนักมากขึ้นในการคำนวณ RMSE เมื่อเปรียบเทียบกับค่าผิดพลาดที่น้อยลง ซึ่งทำให้ RMSE มักจะได้ค่าที่มากกว่า MAE เนื่องจาก MAE หาค่าเฉลี่ยของ error ที่ได้มาถอดค่าสัมบูรณ์โดยไม่ต้องยกกำลังสอง ทำให้ค่า Error ทุกตัวมีผลต่อค่า MAE เท่ากัน โดยวัดประสิทธิภาพของแบบจำลองได้ ดังตารางที่ 3

Product	Method	parameter	RMSE	MAE	MAPE
แจกันทรงสูง	SARIMA	(p,d,q) x (P,D,Q,s) = (1,1,1) x (1,1,1,12)	1222.19	968.65	44.96
	SARIMAX	(p,d,q) x (P,D,Q,s) = (1,1,1) x (1,1,1,12)	810.37	690.57	31.45
	LSTM	n_step = 7, node = 16, epochs = 300, Activation = relu, optimize = adam	592.50	452.65	19.84
	MLP	n_step = 7, 2 Layer node = 16, epochs = 300, Activation = relu, optimize = adam	629.83	566.07	26.91

แจกันทรงเอียงกลาง	SARIMA	(p,d,q) x (P,D,Q,s) = (2,1,1) x (1,1,1,12)	879.41	701.46	29.26
	SARIMAX	(p,d,q) x (P,D,Q,s) = (2,1,1) x (1,1,1,12)	847.23	673.86	27.48
	LSTM	n_step = 7, node = 16, epochs = 300, Activation = relu, optimize = adam	563.48	461.39	25.16
	MLP	n_step = 7, 2 Layer node = 16, epochs = 300, Activation = relu, optimize = adam	587.66	503.07	25.46
แจกันทรงสูงใหญ่	SARIMA	(p,d,q) x (P,D,Q,s) = (1,1,1) x (1,1,1,12)	368.45	276.26	27.03
	SARIMAX	(p,d,q) x (P,D,Q,s) = (1,1,1) x (1,1,1,12)	349.95	281.21	26.14
	LSTM	n_step = 7, node = 16, epochs = 300, Activation = relu, optimize = adam	303.47	269.95	28.00
	MLP	n_step = 7, 2 Layer node = 16, epochs = 300, Activation = relu, optimize = adam	283.79	218.88	22.64

ตาราง 3 แสดงค่า RMSE MAE และ MAPE ของแต่ละแบบจำลอง

จากตารางที่ 3 แสดงการเปรียบเทียบประสิทธิภาพของแต่ละแบบจำลอง จะเห็นว่าในแจกันทรงสูง แบบจำลอง LSTM ให้ประสิทธิภาพที่ดีที่สุด เนื่องจากมีค่า RMSE MAE และ MAPE ต่ำที่สุด คือ ค่า RMSE = 592.50, MAE = 452.65 และ MAPE = 19.84 หมายความว่า รากที่สองของความคลาดเคลื่อนโดยเฉลี่ยของการพยากรณ์เป็นจำนวน 592.50 บาท ค่าพยากรณ์มีความแตกต่างจากค่าจริงโดยเฉลี่ยอยู่ที่ประมาณ 452.65 บาท และ ค่าความคลาดเคลื่อนเฉลี่ยของการพยากรณ์เป็น 19.84% ของค่าจริง

แจกันทรงเอียงกลาง แบบจำลองแบบจำลอง LSTM ให้ประสิทธิภาพที่ดีที่สุด เนื่องจากมีค่า RMSE MAE และ MAPE ต่ำที่สุด คือ ค่า RMSE = 563.48, MAE = 461.39 และ MAPE = 25.16 หมายความว่า รากที่สองของความคลาดเคลื่อนโดยเฉลี่ยของการพยากรณ์เป็นจำนวน 563.48 บาท ค่าพยากรณ์มีความแตกต่างจากค่าจริงโดยเฉลี่ยอยู่ที่ประมาณ 461.39 บาท และ ค่าความคลาดเคลื่อนเฉลี่ยของการพยากรณ์เป็น 25.16% ของค่าจริง แต่จะเห็นได้ว่าแบบจำลอง MLP ก็มีประสิทธิภาพใกล้เคียงกับแบบจำลอง LSTM อาจเป็นเพราะว่าข้อมูลของแจกันทรงเอียงกลางไม่ซับซ้อนมากพอที่จะให้ LSTM มีประสิทธิภาพที่มากเหมือนเคสอื่น ๆ ในกรณีที่ข้อมูลไม่ซับซ้อนมาก โมเดลทั้งสองอาจมีประสิทธิภาพใกล้เคียงกัน

แจกันทรงสูงใหญ่ แบบจำลอง MLP ให้ประสิทธิภาพที่ดีที่สุด เนื่องจากมีค่า RMSE MAE และ MAPE ต่ำที่สุด คือ ค่า RMSE = 283.79, MAE = 218.88 และ MAPE = 22.64 หมายความว่า รากที่สองของความคลาดเคลื่อนโดยเฉลี่ยของการพยากรณ์เป็นจำนวน 283.79 บาท ค่าพยากรณ์มีความแตกต่างจากค่าจริงโดยเฉลี่ยอยู่ที่ประมาณ 218.88 บาท และ ค่าความคลาดเคลื่อนเฉลี่ยของการพยากรณ์เป็น 22.64% ของค่าจริง

จากแจกันทั้ง 3 รายการ แบบจำลองที่ให้ประสิทธิภาพที่ดีที่สุดของแจกันทรงสูง และทรงเอียงกลาง คือ แบบจำลอง LSTM และแบบจำลองที่ให้ประสิทธิภาพที่ดีที่สุดของแจกันทรงสูงใหญ่ คือ แบบจำลอง MLP ซึ่งในแบบจำลอง LSTM มีความสามารถในการจัดเรียงข้อมูลและจำความจำในช่วงเวลาย้อนหลังได้ดี ซึ่งช่วยให้แบบจำลองสามารถจับฤดูกาลและแนวโน้ม

ของข้อมูลได้อย่างแม่นยำ และแบบจำลอง LSTM มีความสามารถในการจัดการกับลักษณะซับซ้อนและลำดับข้อมูลได้อย่างมีประสิทธิภาพ จากข้อมูลชุดนี้ในช่วงปีโควิดทำให้ข้อมูลมียอดขายสูงโดดขึ้นมา จึงอาจทำให้ข้อมูลมีลักษณะที่ซับซ้อน แบบจำลอง LSTM จึงมีประสิทธิภาพมากกว่า SARIMA หรือ SAIMAX และแบบจำลอง MLP ที่ให้ประสิทธิภาพที่ดีที่สุดในของแจกันทรงสูงใหญ่ อาจเกิดจากลักษณะของข้อมูลที่มีความซับซ้อนน้อยลงหรือมีลักษณะที่เหมาะสมกับการใช้ MLP มากกว่า LSTM

ในงานวิจัยนี้ได้ทดลองใช้แบบจำลอง SARIMAX เพื่อพิจารณาวันโปรโมชั่นเข้าไปเป็นปัจจัยภายนอกในแบบจำลอง ซึ่งจากผลการทดลองพบว่าแบบจำลอง SARIMAX ให้ประสิทธิภาพที่ดีกว่าแบบจำลอง SARIMA แสดงว่าวันโปรโมชั่นส่งผลให้แบบจำลองมีประสิทธิภาพที่ดีขึ้น

สรุปผลการวิจัย

งานวิจัยนี้เป็นการศึกษาการพยากรณ์ยอดขายสินค้า โดยใช้แบบจำลอง SARIMA SARIMAX LSTM และ MLP โดยทำการวัดประสิทธิภาพของแบบจำลอง โดยพิจารณาจากค่า RMSE MAE และ MAPE จากผลการวิจัย สรุปได้ว่าแจกันทรงสูงและแจกันทรงเอียงกลาง แบบจำลองที่ให้ประสิทธิภาพที่ดีที่สุด คือ แบบจำลอง LSTM เนื่องจากในปี 2564 เกิดสถานการณ์โควิดทำให้ยอดขายของ 2 รายการนี้มียอดขายที่สูงขึ้นอย่างรวดเร็ว และในปีต่อมายอดขายก็ลดลงอย่างรวดเร็ว ทำให้ข้อมูลมีความซับซ้อนมากขึ้น แบบจำลอง LSTM มีความสามารถในการจัดการกับลักษณะซับซ้อนและลำดับข้อมูลได้อย่างมีประสิทธิภาพ จึงส่งผลให้แบบจำลอง LSTM มีประสิทธิภาพที่ดีที่สุด และแจกันทรงสูงใหญ่ แบบจำลองที่มีประสิทธิภาพที่ดีที่สุด คือ แบบจำลอง MLP เนื่องจากในข้อมูลของแจกันทรงสูงใหญ่อาจมีลักษณะของข้อมูลที่มีความซับซ้อนน้อยลง จึงทำให้ MLP มีประสิทธิภาพที่ดีที่สุด

จากผลการทดลอง ทำให้สามารถพยากรณ์ยอดขายหรือจำนวนของสินค้าที่ต้องผลิตได้ ทำให้ธุรกิจสามารถวางแผนการผลิตให้เหมาะสมกับความต้องการนั้น ๆ โดยลดความขัดแย้งในการจัดการคลังสินค้าและปรับปรุงระบบจัดการสต็อก ลดความสูญเสียจากสินค้าที่เหลือค้างอยู่ในคลังสินค้าหรือสินค้าที่ขายไม่ได้ และมีสินค้าเพียงพอตามความต้องการของผู้บริโภค ไม่มากเกินไปหรือน้อยเกินไป

กิตติกรรมประกาศ

การจัดทำวิจัยได้รับการสนับสนุนจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ในการนำเสนอผลงานวิจัย ผู้วิจัยจึงขอขอบคุณมา ณ ที่นี้

เอกสารอ้างอิง

[1] Kheawpeam, N., & Sinthupinyo, S. (2023, 23-25 July 2023). Demand Forecasting Using Machine Learning to Manage Product Inventory for Multi-channel Retailing Store. 2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS),

[2] บรรจงพัฒนา, ภ., & Banjongpattana, P. (2021). การศึกษาและเปรียบเทียบตัวแบบการพยากรณ์ ARIMA, ARIMAX และ GARCH ต่อดัชนี

[3] Xinxiang, Z., Bo, Z., & Huijuan, F. (2017, 15-17 Dec. 2017). A comparison study of outpatient visits forecasting effect between ARIMA with seasonal index and SARIMA. 2017 International Conference on Progress in Informatics and Computing (PIC)

[4] Falatouri, T., Darbanian, F., Brandtner, P., & Udokwu, C. (2022). Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM. Procedia Computer Science, 200, 993-1003. https://doi.org/https://doi.org/10.1016/j.procs.2022.01.298