การพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุรกิจค้าปลีกออนไลน์ โดยใช้การเรียนรู้ของเครื่อง

FORECASTING CUSTOMER DEMAND FOR NEW PRODUCTS IN AN ONLINE RETAIL BUSINESS USING MACHINE LEARNING

สุพรรษา วัฒนบุตร

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2566

การพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุรกิจค้าปลีกออนไลน์ โดยใช้การเรียนรู้ของเครื่อง

สุพรรษา วัฒนบุตร

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2566
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

FORECASTING CUSTOMER DEMAND FOR NEW PRODUCTS IN AN ONLINE RETAIL BUSINESS USING MACHINE LEARNING

SUPANSA WATTANABUTR

A Master's Project Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Data Science)

Faculty of Science, Srinakharinwirot University

2023

Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

การพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุรกิจค้าปลีกออนไลน์

โดยใช้การเรียนรู้ของเครื่อง

ของ

สุพรรษา วัฒนบุตร

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร.นุวีย์ วิวัฒนวัฒนา)

..... ประธาน

(รองศาสตราจารย์ ดร.ดวงดาว วิชาดากุล)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ศิริสรรพ เหล่าหะเกียรติ)

ชื่อเรื่อง	การพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุรกิจค้าปลีกออนไลน์ โดยใช้การเรียนรู้ของเครื่อง
ผู้วิจัย	สุพรรษา วัฒนบุตร
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. นุวีย์ วิวัฒนวัฒนา

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาการสร้างแบบจำลองเบื้องต้นที่ใช้สำหรับการพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุรกิจค้าปลีกออนไลน์ ด้วยเทคนิคการเรียนรู้ของเครื่อง เพื่อพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรกที่สินค้าใหม่เริ่มวางขาย จากข้อมูลการขายของสินค้าปัจจุบันที่มีความคล้ายคลึงกันกับสินค้าใหม่ โดยประกอบไปด้วย 2 วิธีหลักดังนี้ 1. ทำการเลือกสินค้าที่มีความคล้ายคลึงกันระหว่างสินค้าปัจจุบันกับสินค้าใหม่ 2. สร้างแบบจำลองการพยากรณ์ข้อมูลด้วยแบบจำลองโครงข่ายประสาทเทียมแบบหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) แบบจำลองอนุกรมเวลาแบบ Autoregressive Integrated Moving Average (ARIMA) และการพยากรณ์อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA) โดยใช้ชุดข้อมูลร้านค้าปลีกแห่งหนึ่งในประเทศไทย ที่มีการขายสินค้าผ่านช่องทางออนไลน์ควบคู่กับการขายผ่านหน้าร้าน มีข้อมูลรายการธุรกรรมการขายรายวันตั้งแต่เดือนมกราคม ค.ศ. 2021 ถึงเดือนกันยายน ค.ศ. 2023 ผู้วิจัยเปรียบเทียบผลของการพยากรณ์ด้วยแบบจำลอง LSTM, ARIMA และ MA โดยพิจารณาจากค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) ในการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้ง 3 วิธี ผลจากการทดลองพบว่าผลลัพธ์ที่ได้จากการพยากรณ์ด้วยวิธี MA มีประสิทธิภาพในการพยากรณ์ที่แม่นยำมากกว่าในกรณียอดขายของสินค้าใหม่นั้นมีแนวโน้มคงที่ และ ARIMA มีประสิทธิภาพในการพยากรณ์ที่แม่นยำมากกว่าในกรณียอดขายของสินค้าใหม่นั้นมีแนวโน้มลดลง ดังนั้นการพยากรณ์ด้วยวิธี MA เป็นวิธีที่น่าสนใจหากนำเอาวิธีดังกล่าวไปประยุกต์ใช้งานจริงกับการพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรก เพราะเป็นวิธีการที่ง่าย ใช้เวลาน้อยในการพยากรณ์ และเหมาะกับข้อมูลที่ไม่ซับซ้อนและมีความคงที่ แม้จะเป็นวิธีการดั้งเดิมแต่ให้ค่าความแม่นยำสูงเมื่อเทียบกับอีก 2 วิธีที่ใช้เวลาและมีวิธีการที่ซับซ้อนมากกว่า ซึ่งผลการพยากรณ์ที่ได้นั้นไม่ได้แตกต่างจากวิธี MA มากนักโดยเฉพาะค่าพยากรณ์ยอดขายเดือนแรก เมื่อพิจารณาเฉพาะค่าพยากรณ์ยอดขายเดือนแรกนั้น LSTM, ARIMA และ MA ไม่ได้แตกต่างกันมากและมีความใกล้เคียงกับยอดขายเดือนแรกของสินค้าใหม่

คำสำคัญ : การพยากรณ์ความต้องการสินค้าใหม่, การเรียนรู้ของเครื่อง, สินค้าใหม่

Title	FORECASTING CUSTOMER DEMAND FOR NEW PRODUCTS IN AN ONLINE RETAIL BUSINESS USING MACHINE LEARNING
Author	SUPANSA WATTANABUTR
Degree	MASTER OF SCIENCE
Academic Year	2023
Thesis Advisor	Assistant Professor Nuwee Wiwatwattana , Ph.D.

This research investigates the development of a preliminary model for forecasting customer demand for new products in an online retail business using machine learning techniques. The objective is to predict the demand for new products in the first three months of their releases, based on sales data of similar current products. The study employed two main approaches: (1) selection of products with similar characteristics to new products; (2) construction of predictive models using Long Short-Term Memory (LSTM) models, Autoregressive Integrated Moving Average (ARIMA) models, and Moving Averages (MA) models. It utilized data from a Thai retail store, which conducts sales through both online and physical stores (from January 2021-September 2023), this study compared the outcomes of predictions made by LSTM, ARIMA, and MA models, considering Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) metrics. The experimental findings revealed that the MA model outperformed the others in accurately forecasting the sales trends of new products that exhibit stable trends. Conversely, the ARIMA model excels in predicting products with declining sales trends. Consequently, employing the MA method for forecasting the demand for new products in the first three months appears promising due to its simplicity, minimal computational time, and compatibility with stable and less complex data. Despite being a traditional method, The MA models yield high accuracy compared to the other two methods, which are more time-consuming and intricate. Particularly, when focusing solely on the prediction of first-month sales, LSTM, ARIMA, and MA models show insignificant differences and closely approximate the actual sales of new products.

Keyword : Forecasting Customer Demand for New Products, Machine Learning, New Products

กิตติกรรมประกาศ

สารนิพนธ์นี้สำเร็จลุล่วงได้ด้วยความอนุเคราะห์จาก ผศ.ดร.นุวีย์ วิวัฒนวัฒนา อาจารย์ที่ปรึกษาที่ให้คำปรึกษา และคำแนะนำในการทำสารนิพนธ์

ขอกราบขอบพระคุณร้านค้าเจ้าของข้อมูลที่ได้ให้ความอนุเคราะห์ข้อมูลรายการธุรกรรมการขายมาใช้ในการทำสารนิพนธ์นี้

ขอกราบขอบพระคุณคณะกรรมการสอบสารนิพนธ์ทุกท่านที่ได้ให้คำแนะนำและข้อเสนอแนะ สำหรับการปรับปรุงสารนิพนธ์ และการใช้เทคนิคการเรียนรู้ของเครื่องโดยใช้แบบจำลองโครงข่ายปราสาทเทียมแบบหน่วยความจำระยะสั้นแบบยาว และแบบจำลองอนุกรมเวลามาเป็นเครื่องมือในการวิเคราะห์และแก้ไขปัญหาทางข้อมูล

ขอกราบขอบพระคุณบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ และคณาจารย์ทุกท่านในภาควิชาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่กรุณาให้ความรู้ คำแนะนำ และสนับสนุนการทำวิจัย

สุพรรษา วัฒนบุตร

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญรูปภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ความสำคัญและความเป็นมาของงานวิจัย 1

1.2 วัตถุประสงค์ 4

1.3 ขอบเขตของการวิจัย 4

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย 5

บทที่ 2 วรรณกรรมและงานวิจัยที่เกี่ยวข้อง 6

2.1 นิยามของสินค้าใหม่หรือผลิตภัณฑ์ใหม่ 6

2.2 ทฤษฎีเกี่ยวกับอัลกอริทึม 7

2.3 งานวิจัยที่เกี่ยวข้อง 14

บทที่ 3 การดำเนินการวิจัย 22

3.1 กระบวนการทำงานของแบบจำลอง 23

3.2 การเก็บรวบรวมข้อมูล (Data Collection) 28

3.3 การสำรวจข้อมูล Exploratory Data Analysis (EDA) 30

3.4 การเตรียมข้อมูล (Data Preparation) 39

3.5 การสร้างแบบจำลองพยากรณ์ 46

บทที่ 4 ผลการดำเนินการวิจัย 52
4.1 ผลลัพธ์ของการสร้างแบบจำลอง Long Short-Term Memory (LSTM) 52
4.2 ผลลัพธ์ของการสร้างแบบจำลอง Autoregressive Integrated Moving Average (ARIMA) 54
4.3 ผลลัพธ์ของการพยากรณ์แบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA) 55
4.4 ผลลัพธ์จากการเปรียบเทียบการพยากรณ์ความต้องการสินค้าระหว่างแบบจำลอง Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA) และ Moving Averages (MA) 59
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ 64
5.1 สรุปผลการวิจัย 64
5.2 อภิปรายผลการวิจัย 67
5.3 ข้อเสนอแนะ 69
บรรณานุกรม 70
ประวัติผู้เขียน 73

สารบัญตาราง

หน้า

ตาราง 1 แสดงผลการทดลองวิธี Machine Learning Models แต่ละวิธีกับแบบจำลองที่เจ้าของงานวิจัยพัฒนาขึ้น 14

ตาราง 2 สรุปงานวิจัยที่เกี่ยวข้อง 20

ตาราง 3 แสดงตัวแปรของข้อมูลรายการธุรกรรมการขาย 29

ตาราง 4 แสดงข้อมูลสินค้าปัจจุบัน และสินค้าใหม่ที่มีความคล้ายคลึงกัน 46

ตาราง 5 แสดงค่าพารามิเตอร์ที่ใช้ของแต่ละสินค้า 51

ตาราง 6 แสดงผลการพยากรณ์ความต้องการสินค้าของสินค้าใหม่ โดยใช้แบบจำลอง LSTM 53

ตาราง 7 แสดงผลการพยากรณ์ความต้องการสินค้า 3 เดือนแรกของสินค้าใหม่ โดยใช้แบบจำลอง ARIMA 55

ตาราง 8 แสดงผลการพยากรณ์ความต้องการสินค้า 3 เดือนแรกของสินค้าใหม่ โดยวิธี MA 56

ตาราง 9 แสดงผลการพยากรณ์ความต้องการสินค้าใหม่ 1 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA เทียบกับค่าจริงและจำนวนสั่งซื้อที่ทำการปัดค่าพยากรณ์เป็นจำนวนเต็ม 63

ตาราง 10 แสดงผลการพยากรณ์ความต้องการสินค้าใหม่ของแต่ละสินค้าระหว่างแบบจำลอง LSTM, ARIMA และ MA 65

ตาราง 11 แสดงผลการพยากรณ์ความต้องการสินค้าใหม่ 1 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA เทียบกับจำนวนสั่งซื้อที่ทำการปัดค่าพยากรณ์เป็นจำนวนเต็ม 66

สารบัญรูปภาพ

หน้า

ภาพประกอบ 1 แสดงรูปภาพการทำงานของ LSTM 8

ภาพประกอบ 2 แสดง Flowchart กระบวนการสร้างแบบจำลอง LSTM 23

ภาพประกอบ 3 แสดง Flowchart กระบวนการสร้างแบบจำลอง ARIMA 25

ภาพประกอบ 4 แสดง Flowchart กระบวนการสร้างแบบจำลองด้วยวิธี MA..... 27

ภาพประกอบ 5 แสดงตัวอย่างข้อมูล 30

ภาพประกอบ 6 แสดงข้อมูลด้วยสถิติเชิงพรรณา 31

ภาพประกอบ 7 แสดงรายละเอียดข้อมูลแต่ละคอลัมน์..... 32

ภาพประกอบ 8 แสดงแนวโน้มจำนวนยอดขายแยกรายปี..... 33

ภาพประกอบ 9 แสดงจำนวนยอดขายแยกรายเดือน..... 34

ภาพประกอบ 10 แสดงจำนวนยอดขายแยกรายวัน 35

ภาพประกอบ 11 แสดงจำนวนยอดขายแยกตามช่วงเวลา..... 36

ภาพประกอบ 12 แสดงจำนวนยอดขายแยกตามหมวดหมู่..... 37

ภาพประกอบ 13 แสดงจำนวนยอดขาย 5 อันดับแรกแยกตามหมวดหมู่ 37

ภาพประกอบ 14 แสดงจำนวนยอดขายแยกตามหมวดหมู่และแยกตามปี 5 อันดับแรก 38

ภาพประกอบ 15 แสดงแถวที่ทำการ Clean ข้อมูล 39

ภาพประกอบ 16 แสดงคอลัมน์ที่ทำการ Clean ข้อมูล 40

ภาพประกอบ 17 แสดงคอลัมน์ที่มีค่าเริ่มต้นเหมือนกัน 40

ภาพประกอบ 18 แสดงคอลัมน์ที่ทำการเพิ่ม 41

ภาพประกอบ 19 แสดงจำนวนธุรกรรมการขายของสินค้าแยกตามปีที่สินค้าเริ่มวางขาย 41

ภาพประกอบ 20 แสดงแนวโน้มจำนวนยอดขายของสินค้าปัจจุบัน..... 42

ภาพประกอบ 21 แสดงหมวดหมู่สินค้าที่ขายดีที่สุด 5 อันดันแรก ของสินค้าปัจจุบัน 43

ภาพประกอบ 22 แสดงแนวโน้มจำนวนยอดขายของสินค้าใหม่ 43

ภาพประกอบ 23 แสดงหมวดหมู่สินค้าที่ขายดีที่สุด 5 อันดันแรก ของสินค้าใหม่ 44

ภาพประกอบ 24 แสดงค่า Trend Seasonal และ Residual ของ Train Data ครีมย้อมผม C1 . 48

ภาพประกอบ 25 แสดงค่า Trend Seasonal และ Residual ของ Train Data ครีมย้อมผม C2 . 49

ภาพประกอบ 26 แสดงค่า Trend Seasonal และ Residual ของ Train Data สบู่วิตามิน C3 ... 49

ภาพประกอบ 27 แสดงค่า Trend Seasonal และ Residual ของ Train Data ดินสอเขียนคิ้ว C4 50

ภาพประกอบ 28 กราฟแสดงค่า RMSE และ MAE เปรียบเทียบระหว่าง LSTM, ARIMA และ MA ของสินค้าครีมย้อมผม N1..... 57

ภาพประกอบ 29 กราฟแสดงค่า RMSE และ MAE เปรียบเทียบระหว่าง LSTM, ARIMA และ MA ของสินค้าครีมย้อมผม N2..... 57

ภาพประกอบ 30 กราฟแสดงค่า RMSE และ MAE เปรียบเทียบระหว่าง LSTM, ARIMA และ MA ของสินค้าสบู่วิตามิน N3..... 58

ภาพประกอบ 31 กราฟแสดงค่า RMSE และ MAE เปรียบเทียบระหว่าง LSTM, ARIMA และ MA ของสินค้าดินสอเขียนคิ้ว N4 58

ภาพประกอบ 32 กราฟแสดงผลการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าครีมย้อมผม N1 ใน 3 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA..... 59

ภาพประกอบ 33 กราฟแสดงผลการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าครีมย้อมผม N2 ใน 3 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA..... 60

ภาพประกอบ 34 กราฟแสดงผลการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าสบู่วิตามิน N3 ใน 3 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA..... 60

ภาพประกอบ 35 กราฟแสดงผลการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าดินสอเขียนคิ้ว N4 ใน 3 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA..... 61

บทที่ 1
บทนำ

1.1 ความสำคัญและความเป็นมาของงานวิจัย

เนื่องจากปัจจุบันธุรกิจค้าปลีกสมัยใหม่มีการเพิ่มบริการการขายผ่านช่องทางออนไลน์ (Online) ควบคู่กับการขายผ่านหน้าร้าน (Offline) เพื่อให้เข้าถึงกลุ่มลูกค้าได้กว้างขึ้น ซึ่งมีผู้ประกอบการมากมายทั้งรายใหญ่รายย่อย และยังต้องแข่งขันกับธุรกิจการค้าออนไลน์ที่เติบโตต่อเนื่องจากการเข้าถึงสมาร์ทโฟนและอินเทอร์เน็ตได้สะดวก ทำให้การแข่งขันของธุรกิจมีแนวโน้มสูงขึ้น (Tunpaiboon, 2562) ดังนั้นผู้ประกอบการจึงต้องมีการปรับตัวและพัฒนาการขายสินค้าให้ตรงกับความต้องการของผู้บริโภคอยู่ตลอดเวลา และการพัฒนาการขายอย่างหนึ่ง คือการเพิ่มความหลากหลายของสินค้า เพื่อให้ตอบสนองความต้องการของลูกค้าได้มากยิ่งขึ้น ให้ลูกค้าสามารถเลือกซื้อสินค้าได้สะดวกครบถ้วนในที่เดียวได้มากที่สุด ซึ่งส่งผลต่อการรักษาฐานลูกค้าและการขยายฐานลูกค้า อีกทั้งเพื่อเพิ่มโอกาสในการขายของผู้ประกอบการด้วย ดังนั้นผู้ประกอบการจึงต้องสรรหาผลิตภัณฑ์ใหม่ ๆ เข้ามาขายโดยเฉพาะผ่านช่องทางออนไลน์ เพราะสามารถขายสินค้าได้หลากหลายประเภทกว่าผ่านหน้าร้าน เนื่องจากสินค้าหลายรายการไม่สามารถที่จะนำมาวางขายผ่านหน้าร้านได้ ด้วยข้อจำกัดหลายอย่าง เช่น ขนาดสินค้ากับพื้นที่ของร้าน เป็นต้น ซึ่งการดำเนินธุรกิจให้มีความหลากหลายของสินค้าและบริการมากขึ้น อาจทำให้ผู้ประกอบการธุรกิจนั้นมีปัญหาในการจัดการพื้นที่จัดเก็บ และการบริหารจัดการสินค้าที่ยุ่งยากมากขึ้น โดยเฉพาะกับสินค้าใหม่ที่เตรียมจะนำเข้ามาวางขาย เนื่องจากยังไม่มีข้อมูลปริมาณการขายในอดีตมาช่วยในการพยากรณ์ความต้องการของลูกค้าได้

ด้วยเหตุนี้หากผู้ประกอบการไม่ได้มีการประมาณการจำนวนยอดขายสินค้าใหม่ที่นำเข้ามาเพื่อวางขายว่าจะสามารถขายได้ปริมาณเท่าใดตามจำนวนหน่วยขายเป็นจำนวนชิ้น และใช้ระยะเวลาเท่าใดในการขายสินค้านั้น ๆ ได้หมด ซึ่งวิธีที่ใช้ในปัจจุบัน คือใช้วิธีการสั่งซื้อสินค้าใหม่ตามจำนวนที่ทำให้ได้ราคาทุนสินค้าที่ถูกยกตัวอย่างเช่น ซื้อสินค้ารายการหนึ่งจำนวน 1 ลัง จะซื้อได้ในราคาลังละ 500 บาท หากซื้อในจำนวน 10 ลัง จะซื้อได้ในราคาลังละ 480 บาท เป็นต้น ทำให้ผู้ประกอบการมักจะเลือกซื้อสินค้าในปริมาณมาก เพื่อให้ได้ราคาทุนถูกแล้วสามารถนำไปแบ่งขายปลีกตามจำนวนหน่วยขายเป็นจำนวนชิ้นแล้วได้กำไรที่เพิ่มขึ้น แต่การพิจารณาจำนวนในการซื้อเพื่อให้ได้ราคาทุนที่ถูกเพียงอย่างเดียวนั้น ส่งผลให้เกิดการสั่งซื้อสินค้าใหม่เข้ามามากเกินความต้องการของลูกค้า ซึ่งเป็นปัญหาที่พบอยู่ในปัจจุบันของการสั่งซื้อสินค้าใหม่ สินค้าบางรายการไม่สามารถขายได้หรือขายได้น้อยกว่าที่คาดการณ์ไว้มาก ส่งผลทำให้เกิดปัญหาตามมา เช่น ปัญหา

พื้นที่จัดเก็บ ค่าใช้จ่ายในการเก็บรักษา การหมุนเวียนเงินทุน รวมทั้งขาดทุนหากเก็บสินค้านั้น ๆ ไว้ปริมาณมากจนเกินไปไม่สามารถขายได้ทันก่อนสินค้าหมดอายุหรือชำรุด เป็นต้น หากต้องทำการวิจัยตลาด เพื่อสำรวจความต้องการสินค้าใหม่ของลูกค้าก่อนจะนำสินค้าใหม่เข้ามาวางขาย ก็จะต้องใช้ระยะเวลา และค่าใช้จ่ายที่ค่อนข้างสูง ซึ่งปกติการดำเนินการทางธุรกิจค้าปลีกจะมีการนำสินค้าใหม่เข้ามาขายอยู่เรื่อย ๆ เพื่อเพิ่มยอดขาย ซึ่งการทำวิจัยตลาดกับสินค้าใหม่ทุกตัวที่จะนำมาวางขายนั้นไม่สามารถที่จะทำได้ โดยเฉพาะกับผู้ประกอบการค้าปลีกรายย่อยทั่วไป ดังนั้นหากมีวิธีการที่สามารถช่วยคาดการณ์จำนวนยอดขายสินค้าใหม่ตามจำนวนหน่วยขายเป็นจำนวนชิ้นได้ ก็จะช่วยให้ปัญหาเหล่านี้ลดลงจากเดิม เมื่อพิจารณาหลักการในการบริหารจัดการสินค้าโดยทั่วไป เพื่อดูหลักการในการสั่งสินค้า มีแนวทางการสั่งซื้อสินค้าดังนี้

1. มีการกำหนดรอบระยะเวลาในการสั่งซื้อที่แน่นอนสม่ำเสมอ เช่น สั่งซื้อสัปดาห์ละครั้ง
2. ทำการตรวจนับสินค้าคงเหลือก่อนที่จะทำการสั่งซื้อสินค้าตามรอบการสั่ง
3. สั่งซื้อสินค้าจากข้อมูลยอดขายในอดีต และจากข้อมูลการพยากรณ์ยอดขาย
4. หาข้อมูลใหม่ ๆ เพื่อช่วยในการตัดสินใจ เช่น ความนิยมของลูกค้าในปัจจุบัน
5. สั่งซื้อตามจำนวนที่สามารถจะขายในช่วงเวลาที่กำหนด
6. ไม่กักตุนสินค้ามากเกินไป เพราะจะทำให้คุณภาพสินค้าตกได้ หากเก็บรักษาไม่ดีพอ
7. บริหารจัดการสต๊อกไม่ให้สินค้าขาด เพื่อลดการเสียโอกาสในการขาย
8. สั่งซื้อจากแหล่งที่น่าเชื่อถือและสามารถส่งให้ถึงหน้าร้านหรือคลังสินค้า เพื่อประหยัดเวลาและค่าใช้จ่ายในการเดินทาง
9. ไม่ซื้อสินค้าจากแหล่งที่ไม่แน่นอน เพราะอาจได้สินค้าที่ไม่ได้มาตรฐาน

นอกจากนี้แล้วผู้ประกอบการยังต้องทำการตรวจเช็คสินค้าอย่างสม่ำเสมอ และพิจารณาการบริหารจัดการคำสั่งซื้อให้เหมาะสมดังนี้

1. ทำการสั่งซื้อสินค้าที่ขายดี และบริหารจัดการไม่ให้สต๊อกสินค้าขายดีนั้นขาด
2. ในกรณีที่สินค้าที่ขายไม่ดี ไม่ควรสั่งซื้อมาเยอะ เพราะเสี่ยงสินค้าหมดอายุ
3. นำเสนอสินค้าใหม่ ๆ เพื่อเป็นทางเลือกและดึงดูดใจทั้งลูกค้าเก่าและลูกค้าใหม่
4. จัดกิจกรรมส่งเสริมการขายเพื่อระบายสินค้าเก่า เช่น ลดราคา สะสมแต้มแลกของ ("แนวทางการสั่งซื้อสินค้า," 2022)

จะเห็นได้ว่าการบริหารจัดการสินค้าในธุรกิจค้าปลีกโดยทั่วไป การนำเสนอสินค้าใหม่ ๆ นั้นก็เป็นส่วนหนึ่งที่สำคัญ แต่ปัญหาในการบริหารจัดการ และการคาดการณ์ยอดขายสินค้าใหม่ตาม

จำนวนหน่วยขายเป็นจำนวนชิ้นนั้นค่อนข้างยาก เนื่องจากยังไม่มีข้อมูลปริมาณการขายในอดีตมาช่วยในการพยากรณ์ความต้องการสินค้าเหมือนกับสินค้าที่มีขายในปัจจุบัน

ผู้วิจัยจึงเล็งเห็นถึงปัญหานี้ ดังนั้นจุดมุ่งหมายของการวิจัยเพื่อสร้างแบบจำลองการพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุรกิจค้าปลีกออนไลน์ โดยใช้การเรียนรู้ของเครื่อง (Machine Learning) เพื่อนำมาใช้ในการพยากรณ์ปริมาณความต้องการสินค้าใหม่จากข้อมูลของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายกัน โดยพยากรณ์ความต้องการสินค้าใหม่ตามจำนวนหน่วยขายเป็นจำนวนชิ้น เพื่อเป็นข้อมูลช่วยประกอบการตัดสินใจของผู้ประกอบการในการประเมินจำนวนยอดขายเป็นจำนวนชิ้น ที่คาดว่าจะขายได้ก่อนที่จะตัดสินใจนำสินค้าใหม่เข้ามาวางขาย เพื่อลดค่าใช้จ่ายในการทำการวิจัยตลาดในการสำรวจความต้องการสินค้าของลูกค้า และลดปัญหาการสั่งซื้อสินค้าเข้ามามากเกินความต้องการของลูกค้าที่เป็นต้นเหตุต่อปัญหาค่าใช้จ่ายในการลงทุนและดูแลรักษาสินค้า โดยทำการพยากรณ์แยกเป็นรายสินค้า Stock Keeping Unit (SKU) ซึ่ง SKU คือ หน่วยสินค้าที่เล็กที่สุดในระบบคลังสินค้า ที่ถูกจำแนกออกตามความแตกต่างใน 1 รหัสสินค้า ด้วย สี ขนาด น้ำหนัก ความกว้าง ความยาว รสชาติ ยี่ห้อ รุ่น เป็นต้น SKU จะช่วยให้สามารถแยกความแตกต่างของสินค้าแต่ละชิ้นได้ง่ายและเป็นการสร้างความเข้าใจตัวสินค้าตรงกันทั้งระบบ ตั้งแต่ ผู้ผลิต ผู้จัดซื้อ จัดส่ง จนถึงผู้ขาย ลดโอกาสในการผิดพลาดในตลอดการจัดการหลังร้านไปจนถึงการขายหน้าร้าน ("สินค้า SKU," 2020)

1.2 วัตถุประสงค์

1. เพื่อศึกษาการสร้างแบบจำลองการพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุรกิจค้าปลีกออนไลน์ โดยพยากรณ์ความต้องการสินค้าใหม่ตามจำนวนหน่วยขายเป็นจำนวนชิ้น ด้วยแบบจำลองโครงข่ายปราสาทเทียมแบบหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) เปรียบเทียบกับวิธีการพยากรณ์ด้วยแบบจำลองอนุกรมเวลาแบบ Autoregressive Integrated Moving Average (ARIMA) และวิธีการพยากรณ์อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA) เพื่อพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรกที่สินค้าใหม่เริ่มวางขาย หน่วยเป็นจำนวนชิ้น/SKU โดยพยากรณ์ความต้องการสินค้าใหม่จากข้อมูลของสินค้าปัจจุบันที่มีขายอยู่และมีความคล้ายคลึงกันกับสินค้าใหม่ โดยใช้มาตรวัดเพื่อการวิเคราะห์แบบถดถอย (Regression Analysis)

2. เพื่อทำความเข้าใจกับชุดข้อมูลของสินค้าในประเภทต่าง ๆ

3. เพื่อทราบถึงหลักการทำงานของอัลกอริทึม และสามารถนำไปใช้ประโยชน์ในการเพิ่มประสิทธิภาพในการพยากรณ์ความต้องการสินค้าใหม่

4. เพื่อศึกษาว่าการสร้างแบบจำลองวิธีใด มีประสิทธิภาพมากที่สุดในการพยากรณ์ความต้องการสินค้าใหม่กับข้อมูลชุดนี้ และสามารถนำแบบจำลองนี้ไปประยุกต์ใช้งานได้จริงกับข้อมูลชุดอื่นที่มีลักษณะคล้ายกันได้

1.3 ขอบเขตของการวิจัย

1. ข้อมูลที่ใช้ในงานวิจัย

งานวิจัยนี้ใช้ข้อมูลจากร้านค้าปลีกแห่งหนึ่งในประเทศไทย ที่มีการขายสินค้าผ่านช่องทางออนไลน์ควบคู่กับการขายผ่านหน้าร้าน มีข้อมูลรายการธุรกรรมการขายรายวัน ตั้งแต่เดือนมกราคม ค.ศ. 2021 ถึงเดือนกันยายน ค.ศ. 2023 โดยมีข้อมูล 1 ตาราง มีคอลัมน์ทั้งหมด 28 คอลัมน์ และมีจำนวนข้อมูล 55,524 แถว มีข้อมูลของสินค้า ซึ่งมีรายละเอียดตัวแปร คือ รหัสสินค้า ชื่อสินค้า หมวดหมู่สินค้า ราคาสินค้า และมีตัวแปรอื่น ๆ เช่น วัน เดือน ปี จำนวนยอดขาย รูปแบบการชำระสินค้า เป็นต้น

2. เตรียมข้อมูลและทำความสะอาดข้อมูลให้อยู่ในรูปแบบที่พร้อมสำหรับนำมาวิเคราะห์มีจำนวนสินค้าทั้งหมด 1,776 SKU ทำการสำรวจข้อมูลโดยรวม จากนั้นทำการแบ่งข้อมูลออกเป็น 2 ชุด ดังนี้

ข้อมูลชุดที่ 1 กำหนดให้สินค้าที่มียอดขายออกครั้งแรกในปี ค.ศ. 2021-2022 เป็นสินค้าปัจจุบัน

ข้อมูลชุดที่ 2 กำหนดให้สินค้าที่มียอดขายออกครั้งแรกในปี ค.ศ. 2023 เป็นสินค้าใหม่

จากนั้นทำการเลือกสินค้าที่มีความคล้ายคลึงกันระหว่างสินค้าปัจจุบันกับสินค้าใหม่ ที่อยู่ในหมวดหมู่เดียวกัน จำนวน 4 คู่ รวมเป็น 8 SKU แบ่งเป็นสินค้าปัจจุบัน 4 SKU และแบ่งเป็นสินค้าใหม่ 4 SKU มาพยากรณ์ความต้องการสินค้าใหม่จากข้อมูลของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายกัน

3. สร้างแบบจำลองการพยากรณ์ข้อมูลด้วยแบบจำลองโครงข่ายปราสาทเทียมแบบหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) แบบจำลองอนุกรมเวลาแบบ Autoregressive Integrated Moving Average (ARIMA) และวิธีการพยากรณ์อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA)

4. ทำการวัดประสิทธิภาพแบบจำลอง โดยพิจารณาจาก

4.1 ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE)

4.2 ค่าเฉลี่ยของค่าสัมบูรณ์ความคลาดเคลื่อน (Mean Absolute Error: MAE)

5. เปรียบเทียบประสิทธิภาพของแบบจำลองแต่ละวิธี แล้วนำเสนอแบบจำลองที่มีประสิทธิภาพดีที่สุด

1.4 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. สามารถนำแบบจำลองที่ได้ไปประยุกต์ใช้งานจริง เพื่อใช้งานในการพยากรณ์ความต้องการสินค้าใหม่ของลูกค้า เพื่อให้ทราบว่าสินค้าใหม่แต่ละ SKU จะมีความต้องการปริมาณกี่ชิ้นใน 3 เดือนแรกที่สินค้าใหม่เริ่มวางขาย

2. เพื่อลดขั้นตอนการทำงาน ลดระยะเวลาและค่าใช้จ่ายในการวิจัยตลาด เพื่อสำรวจความต้องการของสินค้าใหม่

บทที่ 2
วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาทฤษฎี และงานวิจัยที่เกี่ยวข้องกับการพัฒนาแบบจำลองการพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุรกิจค้าปลีกออนไลน์ โดยพยากรณ์ความต้องการสินค้าใหม่ตามจำนวนหน่วยขายเป็นจำนวนชิ้น ด้วยแบบจำลองโครงข่ายปราสาทเทียมแบบหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM), แบบจำลองอนุกรมเวลาแบบ Autoregressive Integrated Moving Average (ARIMA) และการพยากรณ์อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA) ได้นำเสนอตามหัวข้อดังต่อไปนี้

1. นิยามของสินค้าใหม่หรือผลิตภัณฑ์ใหม่
2. ทฤษฎีเกี่ยวกับอัลกอริทึม
3. งานวิจัยที่เกี่ยวข้อง

2.1 นิยามของสินค้าใหม่หรือผลิตภัณฑ์ใหม่

สินค้าใหม่หรือผลิตภัณฑ์ใหม่ (New Product) หมายถึง สินค้าหรือผลิตภัณฑ์ใด ๆ ของกิจการ อาจเป็นสินค้าหรือผลิตภัณฑ์ที่มาจากแนวคิดใหม่ หรือสินค้าที่มีการปรับปรุงเปลี่ยนแปลงบางอย่างในตัวผลิตภัณฑ์สินค้าที่มีอยู่แล้ว เพื่อให้ผู้บริโภคหรือลูกค้ามีความพึงพอใจในตัวผลิตภัณฑ์สินค้า โดยมีความพึงพอใจมากขึ้นกว่าการบริโภคผลิตภัณฑ์สินค้าตัวเดิม หรืออาจเป็นสินค้าเดิมมีการที่นำเสนอในตลาดใหม่ ๆ ที่ยังไม่เคยจำหน่ายสินค้านี้มาก่อน(ถนอม บริคุต, 2014) จำแนกผลิตภัณฑ์สินค้า ได้ 3 ลักษณะ คือ

1. สินค้าใหม่เชิงนวัตกรรม (Innovative Product) คือ สินค้าที่ไม่เคยมีผู้ใดนำเสนอในตลาดมาก่อน สร้างจุดต่างและดึงดูดลูกค้าด้วยความแปลกใหม่ ทันสมัย

2. สินค้าดัดแปลง (Modify Product) คือ สินค้าที่มีการพัฒนาต่อยอดและปรับปรุงมาจากสินค้าเดิม มุ่งเน้นการปรับปรุงประสิทธิภาพ เพิ่มเติมฟังก์ชันการใช้งาน หรือเปลี่ยนดีไซน์ให้ทันสมัย

3. สินค้าลอกเลียนแบบ (Imitative Product) คือ สินค้าที่ลอกเลียนแบบสินค้าที่มีอยู่แล้วในตลาด มุ่งเน้นการนำเสนอราคาที่ถูกกว่าและเจาะกลุ่มลูกค้าเฉพาะเจาะจง หรือสินค้าใหม่ของกิจการแต่ไม่ได้ใหม่ในท้องตลาด ซึ่งสินค้านั้นได้รับการยอมรับและเป็นที่นิยมของผู้บริโภค มีโอกาสในการทำกำไร จึงนำสินค้าที่ลอกเลียนแบบนั้นเข้าสู่ท้องตลาด

2.2 ทฤษฎีเกี่ยวกับอัลกอริทึม

ทฤษฎีโครงข่ายปราสาทเทียมแบบหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM)

โครงข่ายปราสาทเทียมแบบหน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) คือ ระบบโครงข่ายประสาท (Neural Network) รูปแบบหนึ่ง เป็นเทคนิคหนึ่งที่ถูกพัฒนาจาก Recurrent neural network (RNN) ให้มีความเสถียรและมีประสิทธิภาพมากขึ้น มีการเสนอวิธี Long Short-Term Memory หรือ LSTM ขึ้นมาโดย Sepp Hochreiter และ Juergen Schmidhuber ซึ่ง LSTM ทำงานได้ดีในการเรียนรู้แบบ Long-Term หลักการทำงานของ LSTM การที่ภายในอัลกอริทึมมีการเก็บข้อมูล และส่งไปประมวลผลในเวลาถัดไป โดยจะมี Gate ต่าง ๆ ได้แก่ Forget Gate, Input Gate, Input Modulation Gate, Output Gate ทำหน้าที่ควบคุมการไหลของข้อมูล คัดเลือกข้อมูล และคำนวณว่าควรเก็บข้อมูลไว้ใน Cell State และHidden State เท่าใด หากค่าความสำคัญของข้อมูลไม่เพียงพอ จะไม่สามารถผ่าน Gate นั้นไปได้ (Olah, 2015)

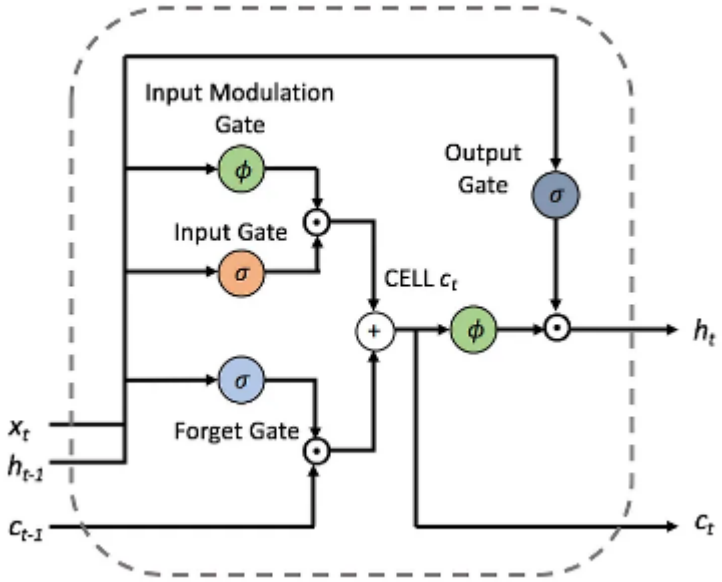
การทำงานของ LSTM

1. Forget คือ การล้าง Cell State เดิมออกไป เพื่อเตรียมพื้นที่รับข้อมูลใหม่ โดย Forget Gate Layer นั้นเป็น Gate ที่มีหน้าที่ในการกำหนดว่าข้อมูลที่จะเข้ามาใน Cell นั้นควรถูกเก็บไว้หรือควรทิ้งไปโดย การสร้าง Forget Gate นี้ จะดู Input Data ที่เข้ามาประกอบกับ Hidden State ก่อนหน้านั้น โดยจะใช้ Sigmoid Activation Function เป็นตัวตัดสิน ผลลัพธ์ที่ได้จาก Forget Gate Layer จะอยู่ระหว่างค่า 0 และ 1 ซึ่งหากได้ค่าเป็น 0 คือให้ลบค่า Cell State เดิมออกไป แต่หากได้ค่าเป็น 1 คือให้เก็บค่า Cell State นี้ต่อไป

2. Write โดย Input Gate Layer เป็น Gate ที่มีหน้าที่รับข้อมูล Input เข้ามาใหม่แล้วจึงทำการบันทึกหรือเขียน (Write) ข้อมูลลงไปในแต่ละโหนด ซึ่งการทำงานแบ่งออกเป็น 2 ส่วน โดยที่ส่วนแรกถ้าต้องการ Update cell state เมื่อทำการรับข้อมูล Input เข้ามาแล้ว จะใช้ Sigmoid Function เป็นตัวตัดสินว่าอนุญาตให้อัพเดทหรือไม่ Function ที่เป็นตัวควบคุมจะเรียกใช้ Input Gate เพื่อที่จะเลือกว่าให้มีการ Update cell state หรือไม่ และส่วนที่สองถ้า Input Gate เลือกที่จะทำการอัพเดท Cell State ซึ่ง Tanh Function ก็จะทำการสร้าง Candidate Values ขึ้นมาใน State การคำนวณนี้ใช้ค่า Input Data ที่เข้ามา กับ Hidden State ก่อนหน้านั้น การอัพเดท Cell State ได้ข้อมูลจาก Forget Gate, Input Gate และInput Modulation Gate เพื่อนำประกอบการพิจารณาการอัพเดท Cell State

3. Read คือ การอนุญาตให้อ่านผลลัพธ์ของการทำงานของอัลกอริทึม โดย Output Gate Layer เป็น Gate ที่มีหน้าที่เตรียมทำการส่งออกข้อมูล (Output Data) โดยข้อมูลที่จะทำการ Output นั้นดูจาก Cell State ที่ผ่านกระบวนการคำนวณต่าง ๆ แล้ว โดย Sigmoid Function จะเป็นตัวเลือกกว่าข้อมูลส่วนไหน Cell State ที่จะถูก Output จากนั้นก็จะนำค่า Cell State เข้า Tanh Function (Output Tanh Function อยู่ในช่วง -1, 1) แล้วนำค่าที่ได้จาก Tanh Function มาทำการคำนวณกับค่า Output ที่ได้จาก Sigmoid Function จากนั้นจะได้ค่า Output ที่ต้องการ การทำงานของ LSTM แสดงดังภาพประกอบที่ 1

(Tangruamsub, 2017)



ภาพประกอบ 1 แสดงรูปภาพการทำงานของ LSTM

ที่มา : (Tangruamsub, 2017)

ทฤษฎีแบบจำลอง Autoregressive integrated average model (ARIMA)

แบบจำลอง ARIMA เป็นรูปแบบของแบบจำลองชุดข้อมูลเวลา (Time Series Data) ที่ใช้ในการวิเคราะห์และการทำนายค่าในอนาคต การพยากรณ์ด้วยวิธี ARIMA เป็นการอาศัยพฤติกรรมของข้อมูลในอดีต เพื่อกำหนดรูปแบบในปัจจุบัน และอธิบายแนวโน้มหรือปรากฏการณ์ต่าง ๆ ของตัวข้อมูลเองในอนาคต แบบจำลอง ARIMA เป็นแบบจำลองที่พัฒนาโดย George E.P.Box และ Gwilym M. Jenkins ในปี ค.ศ.1970 โดยพื้นฐานแล้วแบบจำลอง ARIMA เป็นวิธีที่ให้ค่าพยากรณ์ในระยะสั้นที่ดี หรือเหมาะกับการพยากรณ์ไปข้างหน้าในช่วงเวลาสั้น ๆ และต้องมีช่วงของข้อมูลที่ยาวพอสมควร

ARIMA มีจุดเด่นในการพยากรณ์ด้วยความง่ายในการนำไปประยุกต์ใช้ และสามารถประมาณค่าพารามิเตอร์ได้ด้วยวิธี Ordinary Least Square (OLS) หรือ Maximum Likelihood Estimator (MLE) นอกจากนี้ยังสามารถประยุกต์ใช้กับแนวคิดต่าง ๆ เช่น ARIMAX, Seasonal ARIMA, ARIMA Intervention รวมถึงการร่วมกับ ARIMA-GARCH Model ในการพิจารณาวิกฤติต่าง ๆ อย่างมีประสิทธิภาพในการทำนาย และการจัดการโครงสร้างข้อมูลที่ซับซ้อนได้อย่างมีประสิทธิภาพ (ผู้ช่วยศาสตราจารย์ ดร. เฉลิมพล จตุพรม, 2562)

แบบจำลอง ARIMA (p,d,q) ประกอบด้วย 3 ส่วนหลัก ๆ ได้แก่

1. แบบจำลอง Auto Regressive (AR (p))

แบบจำลอง Auto Regressive เป็นรูปแบบที่แสดงค่าสังเกต y_t ถูกกำหนดจากค่าของ $y_{t-1},\ldots, y_{t-p}$ หรือค่าสังเกตที่เกิดขึ้นก่อนหน้า p ซึ่งจะนำข้อมูลก่อนหน้ามาใช้ในการทำนายหาค่าในอนาคตหรือก็คือค่าปัจจุบัน โดยกระบวนการ AR นั้นใช้พารามิเตอร์ p ซึ่งเป็นค่า Lag (ค่าความล่าช้า) ซึ่งเขียนอยู่ในรูปสมการได้ดังสมการ

$$y_t = c + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \varepsilon t$$

โดยที่

c คือ ค่าคงที่

$\emptyset_p$ คือ Order ของ Auto Regressive

y_{t-p} คือ ค่า Time Series ที่เวลา t-p

ε_t คือ ค่า Error ของโมเดล (ค่า white noise)

2. แบบจำลอง Moving Average (MA (q))

แบบจำลอง Moving Average (MA) เป็นรูปแบบที่แสดงว่า y_t ถูกกำหนดจากค่าความคลาดเคลื่อน ε_{t-1} ,..., ε_{t-q} หรือค่าความคลาดเคลื่อนที่อยู่ก่อนหน้า กล่าวคือ เทคนิค Moving Average (MA) มีหลักการทำงานคล้ายกับ (Auto Regressive) AR แต่แตกต่างกันตรงที่ Moving Average จะสนใจความสัมพันธ์ระหว่าง Error ของช่วงเวลาปัจจุบัน โดยกระบวนการ MA นั้นใช้พารามิเตอร์ q ในการหา Error ซึ่งเขียนให้อยู่ในรูปสมการได้ดังสมการ

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

โดยที่

c คือ ค่าคงที่

ε_t คือ ค่า Error ของ Model (ค่า White Noise)

θ_q คือ Order ของ Moving Average

ε_{t-q} คือ ผลรวมระหว่างค่าปัจจุบันกับค่าก่อนหน้าของ Error Term

3. กระบวนการ Integrated (I (d))

แบบจำลอง ARIMA ถูกออกแบบมาเพื่อวิเคราะห์ข้อมูลอนุกรมเวลาที่มีคุณสมบัติคงที่ (Stationary) หมายความว่าค่าเฉลี่ยและความแปรปรวนของข้อมูลไม่เปลี่ยนแปลงตามเวลา ในกรณีข้อมูลอนุกรมเวลาที่ใช้ในการวิเคราะห์มีคุณสมบัติไม่คงที่ (Non-Stationary) จะต้องทำการแปลงข้อมูลดังกล่าวนั้นให้เป็นข้อมูลที่มีคุณสมบัติคงที่ด้วยกระบวนการ Integrated (I (d)) โดยใช้วิธีการหาผลต่าง (Differencing) ของข้อมูลอนุกรมเวลา เปรียบเทียบค่าข้อมูลในปัจจุบันกับค่าข้อมูลในอดีต d คาบเวลาก่อนหน้า

การหาผลต่างอันดับที่ d สามารถเขียนในรูปของ I (d) ดังนี้

I(d) คือ $\chi_t = \Delta_{d-1} (\chi_t - \chi_{t-1})$

หรือ $(1-B)^d \chi_t$

(มหาวิทยาลัยเชียงใหม่, 2553)

แบบจำลอง Autoregressive Integrated Moving Average Model (ARIMA)

นำแบบจำลอง Auto Regressive แบบจำลอง Moving Average และ กระบวนการ Integrated มาพิจารณาร่วมกัน รูปแบบทั่วไปของแบบจำลอง คือ แบบจำลอง ARIMA (p,d,q)

$$\Delta_d y_d = \delta + \Delta_d y_{t-1} + \phi\Delta_d y_{t-2} + \ldots + \Delta_d y_{t-p} + \varepsilon t - \theta\varepsilon_{t-1} - \ldots - \theta_t\varepsilon_{t-q}$$

โดยที่

y_t คือ ค่าสังเกตในอนุกรมเวลา ณ เวลา t

d คือ จำนวนครั้งของการหาผลต่างเพื่อให้ข้อมูลอนุกรมเวลาคงที่ (Stationary)

p คือ อันดับของ Autoregressive

q คือ อันดับของ Moving Average

δ คือ ค่าคงที่ (Constant Term)

t คือ เวลา

Δ_d คือ ผลต่างอันดับที่ d

$\phi_1,\ldots,\phi_p$ คือ พารามิเตอร์ของ Auto Regressive

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

ε_t คือ กระบวนการ White Noise ค่าความคลาดเคลื่อน ณ เวลา t เป็นตัวแปรสุ่มที่อิสระต่อกัน โดยมีการแจกแจงปกติที่มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนคงที่ (มหาวิทยาลัยเชียงใหม่, 2553)

ทฤษฎีการทดสอบ Unit root

การทดสอบ Unit root สามารถทดสอบโดยใช้วิธีการทดสอบ DF (Dickey-Fuller test) และการทดสอบ ADF (Augment Dickey-Fuller test) การมี Unit Root หมายถึง ลักษณะเฉพาะของข้อมูลอนุกรมเวลาที่ค่าเฉลี่ยหรือความแปรปรวนของข้อมูลมีการเปลี่ยนแปลงไปตามเวลา

การทดสอบ Unit Root ว่าข้อมูลอนุกรมเวลานั้นมีความคงที่หรือไม่ จะทำการทดสอบทั้งหมด 3 รูปแบบ ดังนี้

1. กรณีที่ไม่มีค่าคงที่ และแนวโน้มเวลา $\Delta X_t = \theta X_{t-1} + \varepsilon t$

2. กรณีที่มีเฉพาะค่าคงที่ $\Delta X_t = \propto + X_{t-1} + \varepsilon t$

3. กรณีที่มีทั้งค่าคงที่ และแนวโน้มเวลา $\Delta X_t = \propto + \beta t \theta X t_{-1} + \varepsilon t$

โดยที่

θ คือ ค่าพารามิเตอร์ของการทดสอบ Unit Root

β_t คือ ค่าคงที่ของแนวโน้ม

$\propto$ คือ ค่าคงที่ระดับ

(มหาวิทยาลัยเชียงใหม่, 2553)

ทดสอบโดยใช้การเปรียบเทียบค่าสถิติ ADF Test เทียบกับค่าวิกฤต Mackinnon ซึ่งมีสมมติฐานดังนี้

สมมติฐานหลัก $H_0 : \delta = 0$ ข้อมูลมี Unit Root

สมมติฐานรอง $H_1 : \delta < 0$ ข้อมูลไม่มี Unit Root

ผลลัพธ์ของการทดสอบ

– ยอมรับสมมติฐานหลัก H_0 ข้อมูลมี Unit Root แสดงว่าข้อมูลนั้น Non-Stationary

– ปฏิเสธสมมติฐานหลัก H_1 ข้อมูลไม่มี Unit Root แสดงว่าข้อมูลนั้น Stationary

(มหาวิทยาลัยเชียงใหม่, 2553)

ทฤษฎีการพยากรณ์อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA)

การพยากรณ์โดยวิธีค่าเฉลี่ยเคลื่อนที่ เป็นการพยากรณ์จากข้อมูลยอดขายจริงในอดีตจำนวน N ข้อมูล มาคำนวณหาค่าเฉลี่ย เพื่อพยากรณ์ยอดขายหรือความต้องการสินค้าในช่วงเวลาที่ต้องการ ซึ่งเขียนให้อยู่ในรูปสมการได้ดังสมการ

$$\text{MA} = \frac{\sum_{t=1}^{n} A_t}{N}$$

MA คือ ค่าพยากรณ์

t คือ ช่วงเวลา

A_t คือ ค่าข้อมูลจริง ณ ช่วงเวลา t

N คือ จำนวนข้อมูล

ทฤษฎีเกี่ยวกับการประเมินความแม่นยำของแบบจำลอง

งานวิจัยนี้ใช้ค่า Root Mean Squared Error (RMSE) และ Mean Absolute Error (MAE) ในการวัดค่าความแม่นยำของแบบจำลอง

1. รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Squared Error: RMSE) คือ การวัดความแตกต่างระหว่างค่าจริงและค่าพยากรณ์ โดยคำนวณค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนทั้งหมดแล้วถอดรากที่ 2 (วีรภัทร สาธิตคณิตกุล, 2023) ถ้าหากค่า RMSE นั้นมีค่าที่ต่ำ แสดงว่าค่าพยากรณ์นั้นใกล้เคียงกับค่าจริง ดังสมการ

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

โดยที่

Y_i คือ ค่า Actual หรือค่าจริงของข้อมูลตัวอย่างที่ i

$\hat{\text{Y}}_i$ คือ ค่า Predict หรือค่าพยากรณ์ของข้อมูลตัวอย่างที่ i

n คือ จำนวนข้อมูลตัวอย่างทั้งหมด

2. ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) คือ การหาค่าเฉลี่ยของความแตกต่างแบบสมบูรณ์ระหว่างค่าพยากรณ์กับค่าจริง (วีรภัทร สาธิตคณิตกุล, 2023) ถ้าหากค่า MAE นั้นมีค่าที่ต่ำ แสดงว่าค่าพยากรณ์นั้นมีค่าใกล้เคียงกับค่าจริง ดังสมการ

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

โดยที่

Y_i คือ ค่า Actual หรือค่าจริงของข้อมูลตัวอย่างที่ i

$\hat{\text{Y}}_i$ คือ ค่า Predict หรือค่าพยากรณ์ของข้อมูลตัวอย่างที่ i

n คือ จำนวนข้อมูลตัวอย่างทั้งหมด

2.3 งานวิจัยที่เกี่ยวข้อง

1. บทความวิจัยเรื่อง A Sales Forecast Method for Products with No Historical Data (Dai & Huang, 2021)

งานวิจัยนี้ได้กล่าวถึง วิธีการพยากรณ์ยอดขายของผลิตภัณฑ์ใหม่ที่ไม่มีข้อมูลยอดขายย้อนหลัง กับชุดข้อมูลการขายรถยนต์ ซึ่งประกอบด้วยยอดขายรายเดือนในมณฑลกวางตุ้งตั้งแต่ปี ค.ศ. 2018 ถึงปี ค.ศ. 2019 ประกอบด้วย Feature ของผลิตภัณฑ์ 6 แบรนด์ รุ่นรถ 168 รุ่น ผลปรากฏว่าแบบจำลองที่ผู้วิจัยนำเสนอ ด้วยวิธีการวัดความคล้ายคลึงระหว่างผลิตภัณฑ์เดิมกับผลิตภัณฑ์ใหม่ และใช้ยอดขายของผลิตภัณฑ์เดิมที่คล้ายคลึงนั้นในการพยากรณ์ยอดขายของผลิตภัณฑ์ใหม่ สามารถทำงานได้ดีในข้อมูลขนาดเล็ก และมีประสิทธิภาพในการคาดการณ์ยอดขาย 3 เดือนแรก หลังจากการเปิดตัวได้มีประสิทธิภาพมากกว่า เมื่อเปรียบเทียบกับวิธี Machine Learning Models ที่ใช้ Features และยอดขายในรอบ 3 เดือนหลังการเปิดของผลิตภัณฑ์เดิม สำหรับการฝึกอบรม (Train) และใช้ Features ของผลิตภัณฑ์ใหม่ เพื่อทำนายยอดขายในรอบ 3 เดือน โดยประเมินความแม่นยำในการทำนายด้วยค่า Root Mean Square Percentage Error (RMSPE) แสดงดังตารางที่ 1

ตาราง 1 แสดงผลการทดลองวิธี Machine Learning Models แต่ละวิธีกับแบบจำลองที่เจ้าของงานวิจัยพัฒนาขึ้น

Methods	RMSPE
Linear Regression	4.75
SVR	4.32
KNN	5.68
Random Forest	6.44
CatBoost	8.27
XGBoost	4.99
Neural Network	5.84
Method of This Paper	0.82

2. บทความวิจัยเรื่อง A Machine Learning-based Framework for Forecasting Sales of New Products with Short Life Cycles using Deep Neural Networks (Elalem et al., 2023)

งานวิจัยนี้ได้พัฒนากรอบวิธีการพยากรณ์ยอดขายของผลิตภัณฑ์ใหม่ที่เพิ่งเปิดตัว และมีวงจรชีวิตสั้น รวมทั้งมีข้อมูลยอดขายย้อนหลังที่จำกัด โดยใช้ชุดข้อมูลขนาดใหญ่สองชุด คือ ชุดข้อมูลแรกมาจาก Dell และชุดข้อมูลที่สองมาจาก Retailer X (ชื่อบริษัทไม่เปิดเผย) วิธีการ คือหาผลิตภัณฑ์ที่มีอยู่ที่คล้ายกันและมีอายุสั้นเหมือนกับผลิตภัณฑ์ใหม่ แล้วหาค่าเฉลี่ยของยอดขายย้อนหลังของผลิตภัณฑ์ที่มีอยู่ และใช้ยอดขายเฉลี่ยนั้นเป็นการพยากรณ์หลัก โดยมีขั้นตอนหลัก 3 ขั้นตอน คือ (1) เตรียมข้อมูลการขายสำหรับผลิตภัณฑ์ที่มีอยู่โดยการปรับความเรียบของยอดขายตามวงจรชีวิต เพื่อหายอดขายที่เป็นตัวแทนของวงจรชีวิต แล้วจัดกลุ่มผลิตภัณฑ์ที่คล้ายกันโดยใช้การแบ่งกลุ่ม (2) กำหนดผลิตภัณฑ์ใหม่ให้เข้ากับหนึ่งในกลุ่มของผลิตภัณฑ์ที่มีอยู่โดยใช้สัปดาห์แรกของยอดขายผลิตภัณฑ์ใหม่ (3) ดำเนินการเพิ่มข้อมูลยอดขายของผลิตภัณฑ์ที่มีอยู่ที่ปรับความเรียบแล้ว และยอดขายผลิตภัณฑ์ใหม่ที่ปรับความเรียบในช่วง 2-3 สัปดาห์แรกที่ผลิตภัณฑ์เพิ่งเปิดตัว จากนั้นใช้ข้อมูลเพื่อทำนายยอดขายในระหว่างช่วงวงจรชีวิตของผลิตภัณฑ์ใหม่ โดยประยุกต์ใช้ทั้งวิธีทางสถิติแบบดั้งเดิม (ARIMAX) และใช้การเรียนรู้ของเครื่อง 3 วิธีการ ได้แก่ Deep Neural Networks (DNNs) - Long Short-Term Memory, Gated Recurrent Units และ Convolutional Neural Networks

ผลการศึกษา โดยเปรียบเทียบประสิทธิภาพการพยากรณ์ทั้ง 4 วิธีการ พบว่าการแบ่งกลุ่มผลิตภัณฑ์ตามช่วงเวลาจะทำให้ข้อผิดพลาดในการทำนายน้อยลงอย่างมีนัยสำคัญ ผลการทดลองที่สำคัญคือ การใช้แบบจำลอง ARIMAX ที่เป็นเทคนิคแบบดั้งเดิมสามารถทำนายได้ดีกว่า DNNs โดยมีค่าเฉลี่ยของค่าสัมบูรณ์ความคลาดเคลื่อน (MAE) ต่ำลงถึง 21%–24% อย่างไร้เงื่อนไข อย่างไรก็ตาม เมื่อลองเพิ่มสัญญาณรบกวนแบบเกาส์เซียนในการวิเคราะห์ความทนทาน พบว่าประสิทธิภาพของ ARIMAX ลดลงอย่างมาก ในขณะที่ DNNs มีประสิทธิภาพที่คงที่ ผลการวิจัยนี้เสนอข้อมูลความรู้สำหรับผู้ประกอบการ ว่าเมื่อใดควรเลือกใช้เทคนิค Deep Learning Methods หรือวิธี ARIMAX

3. บทความวิจัยเรื่อง Forecasting Demand Profiles of New Products (van Steenbergen & Mes, 2020)

งานวิจัยนี้ได้พัฒนาวิธีการใหม่ที่เรียกว่า DemandForest ซึ่งมาจากการรวม K-means, Random Forest และ Quantile Regression Forest เพื่อสร้างการทำนายความต้องการของผลิตภัณฑ์ใหม่ในช่วงเริ่มต้นเปิดตัว ด้วยอัลกอริทึมเรียนรู้ของเครื่อง โดยใช้ชุดข้อมูลจาก 5 บริษัท (ชื่อบริษัทไม่เปิดเผย) โดยใช้ตัวอักษร A ถึง E แทนชื่อบริษัท ซึ่งมีข้อมูลลักษณะของผลิตภัณฑ์ ข้อมูลประวัติการขายในผลิตภัณฑ์ที่มีอยู่และผลิตภัณฑ์ใหม่ ซึ่ง DemandForest สามารถจัดกลุ่มทำนายรูปแบบความต้องการ และทำนายปริมาณที่ไม่แน่นอนของความต้องการทั้งหมดในช่วงเวลาการเปิดตัว เจ้าของงานวิจัยตรวจสอบ และแสดงตัวอย่างวิธีการในการทำนาย และการจัดการสินค้าคงคลังโดยใช้ข้อมูลจริงของบริษัทหลาย ๆ แห่ง โดยใช้ DemandForest ซึ่งมาจากการรวมแบบจำลอง K-means, Random Forest และ Quantile Regression Forest พบว่า DemandForest ให้การทำนายที่แม่นยำ ส่งผลทำให้ประหยัดค่าสินค้าคงคลังได้ประมาณ 15% ขึ้นอยู่กับเวลานำส่งและระดับบริการ โดยประเมินความแม่นยำในการทำนายด้วยค่า Root Mean Square Error (RMSE)

4. บทความวิจัยเรื่อง Analysis and Demand Forecasting Based on E-Commerce Data (Tang, 2023)

งานวิจัยนี้ศึกษาวิธีพยากรณ์ความต้องการของสมาร์ทโฟนใน E-Commerce โดยแบบจำลอง ARIMA และ XGBoost ใช้ข้อมูลยอดขายสมาร์ทโฟนของแพลตฟอร์มอีคอมเมิร์ซหนึ่งเพื่อการทดลอง หลังจากการวิเคราะห์เปรียบเทียบผลการทดลอง และความคลาดเคลื่อนของแบบจำลอง พบว่า แม้ว่า ARIMA สามารถทำการพยากรณ์ข้อมูลชุดเวลาได้ดี แต่ XGBoost ซึ่งเป็นอัลกอริทึมในการเรียนรู้เครื่องสามารถจัดการกับปัญหาที่เป็นไม่เชิงเส้น และปัญหาที่ไม่แน่นอนได้ดีกว่า โดย ARIMA มีประสิทธิภาพในการทำนาย โดยมีค่า $R^2 = 0.79$ และ XGBoost มีประสิทธิภาพในการทำนายที่ดีกว่า โดยมีค่า $R^2 = 0.86$

5. บทความวิจัยเรื่อง Demand Forecasting for E-Commerce Platforms (Jain et al., 2020)

งานวิจัยนี้ศึกษาวิธีการทำนายความต้องการของผลิตภัณฑ์ในบริษัท E-Commerce เพื่อช่วยบริษัท E-Commerce ป้องกันไม่ให้เกิดการขาดสต๊อกของสินค้าหรือสต๊อกสินค้ามากเกินความต้องการ งานวิจัยนี้เป็นการศึกษาเปรียบเทียบระหว่างแบบจำลอง Seasonal Autoregressive Integrated Moving Average (SARIMA) และ Long Short-Term Memory Network (LSTM) เจ้าของงานวิจัยได้นำทั้งสองแบบจำลองมาใช้กับชุดข้อมูลของร้านค้าออนไลน์และเปรียบเทียบผลลัพธ์ พบว่า SARIMA ดีกว่าในการทำนายยอดขายระยะยาวในขณะที่ LSTM ให้ผลลัพธ์ที่มีความแม่นยำมากกว่าในระยะเวลาสั้น ๆ โดยรวม SARIMA ให้ผลลัพธ์ที่มีความเสถียรมากกว่า สำหรับข้อมูลทดสอบเดียวกันของหมวดหมู่เฟอร์นิเจอร์ พบว่าค่า RMSE สำหรับโมเดล ARIMA คือ 1.24 ในขณะที่สำหรับโมเดล LSTM คือ 1.55

6. บทความวิจัยเรื่อง Sales Forecasting: A Comparison of Traditional and Modern Times-Series Forecasting Models on Sales Data with Seasonality (Ali & Nakti, 2023)

งานวิจัยนี้ได้กล่าวถึง วิธีการพยากรณ์ยอดขายของผลิตภัณฑ์ โดยใช้ข้อมูลยอดขายขององค์กรการค้าในอินเดีย และการผลิตอาหารตั้งแต่ปี ค.ศ. 2017 ถึงกลางปี ค.ศ. 2022 งานวิจัยนี้เป็นการศึกษาเปรียบเทียบระหว่างแบบจำลอง Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), Long Short-Term Memory (LSTM) และ Fb-Prophet วัดประสิทธิภาพด้วย Root Mean Square Error (RMSE) ผลของการศึกษา พบว่าแบบจำลอง SARIMA เหมาะสำหรับองค์กรที่มีการเคลื่อนไหวอย่างรวดเร็ว มีลำดับฤดูกาล และมีประสิทธิภาพดีกว่าแบบจำลอง LSTM

7. บทความวิจัยเรื่อง Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM (Falatouri et al., 2022)

งานวิจัยนี้เปรียบเทียบคุณภาพของแบบจำลอง SARIMA และ LSTM สำหรับการพยากรณ์ความต้องการในการจัดการห่วงโซ่อุปทานค้าปลีก โดยใช้ข้อมูลยอดขายจริงมากกว่า 37 เดือนรวมถึงข้อมูลเกี่ยวกับการขายผลิตภัณฑ์ต่าง ๆ จากร้านค้าปลีกในออสเตรีย สำหรับการวิเคราะห์และเปรียบเทียบ โดยพบว่า LSTM มีประสิทธิภาพดีกว่าสำหรับผลิตภัณฑ์ที่มีความต้องการที่คงที่ และ SARIMA ทำงานได้ดีกว่าสำหรับผลิตภัณฑ์ที่มีพฤติกรรมตามฤดูกาล ซึ่งรวมถึงปัจจัยภายนอก คือ โปรโมชั่น ผลิตภัณฑ์ที่มีโปรโมชั่น พบว่ามีประสิทธิภาพดีขึ้นอย่างมีนัยสำคัญ ประเมินประสิทธิภาพแบบจำลองด้วยค่า Mean Absolute Percentage Error (MAPE) และ Root Mean Square Percentage Error (RMSE)

8. บทความวิจัยเรื่อง Retail Demand Forecasting using CNN-LSTM Model (Nithin et al., 2022)

งานวิจัยนี้ได้กล่าวถึงการคาดการณ์ความต้องการค้าปลีกตามข้อมูลในอดีต คือ ข้อมูลยอดขายที่ผ่านมา ด้วยแบบจำลองการเรียนรู้เชิงลึก CNN-LSTM (Convolutional Neural Network-Long Short-Term Memory Network) โดยโมเดลได้รับการฝึกอบรมและทดสอบในชุดข้อมูลที่ประกอบด้วยร้านค้า 10 แห่งและ 50 รายการ โดยมีการสังเกตมากกว่า 900,000 รายการจัดเรียงตามวันที่ ร้านค้า และรายการ โดยเปรียบเทียบการใช้ Swish Activation Function และ Activation Function ReLU (Rectified Linear Unit) ซึ่งพบว่า Swish Activation Function มีประสิทธิภาพดีกว่า Activation Function ReLU โดยโมเดล CNN-LSTM มีค่า Root Mean Square Percentage Error (RMSE) ต่ำกว่าอย่างมีนัยสำคัญ

9. บทความวิจัยเรื่อง Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM (Dave et al., 2021)

งานวิจัยนี้ได้กล่าวถึงการพยากรณ์การส่งออกในอนาคตของอินโดนีเซีย ซึ่งรัฐบาลสามารถใช้ในการตัดสินใจ และส่งเสริมเศรษฐกิจในอนาคต โดยใช้ข้อมูลส่งออกรายเดือนของประเทศอินโดนีเซีย ตั้งแต่ ค.ศ. 1998 – 2019 จากเว็บไซต์ Federal Reserve Economic

งานวิจัยนี้เป็นการศึกษาเปรียบเทียบระหว่างแบบจำลอง ARIMA, LSTM และ Hybrid LSTM– ARIMA ทำการประเมินวัดความแม่นยำด้วยค่า Mean Absolute Percentage Error (MAPE) พบว่า แบบจำลอง Hybrid LSTM– ARIMA มีความแม่นยำที่ดีกว่า โดยมีค่า MAPE เท่ากับ 7.38% เมื่อเทียบกับแบบจำลอง ARIMA เท่ากับ 9.38% และ LSTM เท่ากับ 8.56%

10. บทความวิจัยเรื่อง Demand Forecasting in Production Planning for Dairy Products using Machine Learning and Statistical Method
(Vithitsoontorn & Chongstitvatana, 2022)

งานวิจัยนี้ได้กล่าวถึงการพยากรณ์ความต้องการในการวางแผนการผลิตผลิตภัณฑ์นม โดยข้อมูลธุรกรรมการขายรายวันที่รวบรวมจากระบบ ERP ขององค์การส่งเสริมการเลี้ยงโคนมประเทศไทย (อ.ส.ค.) ภายในเดือนตุลาคม พ.ศ. 2559 – เดือนกันยายน พ.ศ.2564 (5 ปี) บันทึกโดยโรงงาน 5 แห่ง มีผลิตภัณฑ์ขายมากกว่า 100 รายการ แต่เลือกผลิตภัณฑ์เพียง 25 รายการที่ต้องการการคาดการณ์ เนื่องจากครอบคลุม 90% ของรายได้ทั้งหมด และผลผลิตประจำเดือนที่ผลิตโดยโรงงานทั้งหมด เลือกสินค้า 8 รายการ จากทั้งหมด 40 ซีรีส์ที่แตกต่างกัน โดยมีปริมาณยอดขายที่มีความผันผวนสูงตั้งแต่ 10 หน่วยต่อเดือน จนถึงมากกว่า 800,000 หน่วยต่อเดือนและข้อมูลวันหยุดรวบรวมจากธนาคารแห่งประเทศไทย

งานวิจัยนี้เป็นการศึกษาเปรียบเทียบระหว่างแบบจำลอง ARIMA และ LSTM ทำการประเมินวัดความแม่นยำด้วยค่า RMSE, MAE และ MASE พบว่า ARIMA มีประสิทธิภาพดีกว่าในข้อมูลแนวโน้มขาขึ้น (Uptrend) ส่วนข้อมูลแนวโน้มขาลง (Downtrend) หรือชุดข้อมูลที่ซับซ้อนมากขึ้น LSTM มีประสิทธิภาพดีกว่า และการฝึกแบบจำลองด้วยการสังเกตรายเดือนให้ผลลัพธ์ที่ดีกว่า เมื่อเทียบกับการสังเกตรายสัปดาห์

ผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้อง สามารถสรุปผลแต่ละงานวิจัยได้ดังตารางที่ 2

ตาราง 2 สรุปงานวิจัยที่เกี่ยวข้อง

ลำดับ	ชื่องานวิจัย	แบบจำลอง	แบบจำลองที่มีประสิทธิภาพดีที่สุด
1	A Sales Forecast Method for Products with No Historical Data (Dai & Huang, 2021)	Linear Regression SVR, KNN Random Forest CatBoost XGBoost Neural Network Method of This Paper	Method of This Paper
2	A Machine Learning-based Framework for Forecasting Sales of New Products with Short Life Cycles using Deep Neural Networks (Elalem, Maier, & Seifert, 2023)	ARIMAX DNNs – LSTM Gated Recurrent Units Convolutional Neural Networks	ARIMAX กรณีไม่มีสัญญาณรบกวนแบบเกาส์เซียน DNNs กรณีมีสัญญาณรบกวนแบบเกาส์เซียน
3	Forecasting Demand Profiles of New Products (van Steenbergen & Mes, 2020)	DemandForest K-means Random Forest Quantile Regression Forest	DemandForest
4	Analysis and Demand Forecasting Based on E-Commerce Data (Tang, 2023)	ARIMA XGBoost	XGBoost

ตาราง 2 (ต่อ)

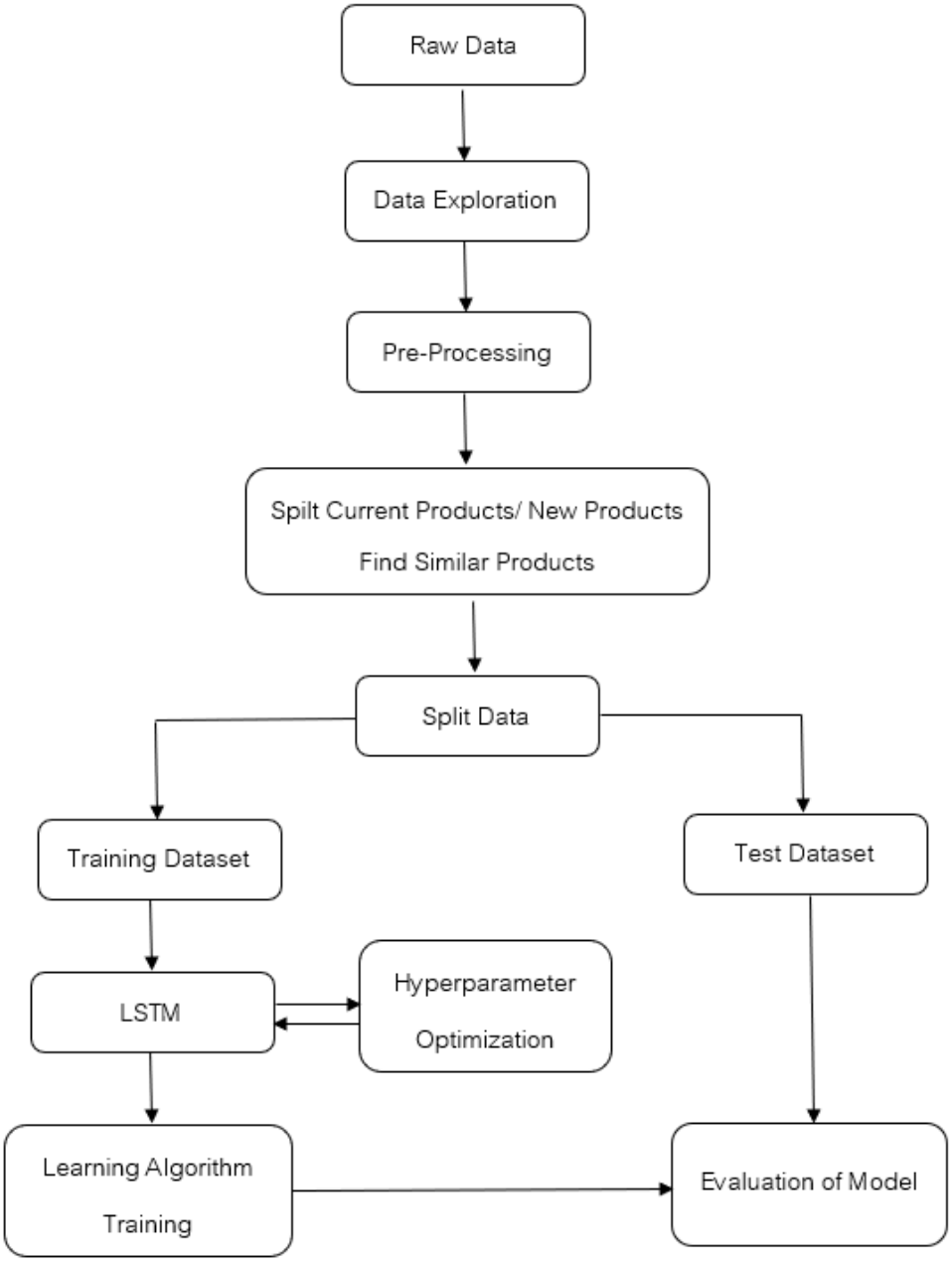
ลำดับ	ชื่องานวิจัย	แบบจำลอง	แบบจำลองที่มีประสิทธิภาพดีที่สุด
5	Demand Forecasting for E-Commerce Platforms (Jain et al., 2020)	SARIMA LSTM	SARIMA
6	Sales Forecasting: A Comparison of Traditional and Modern Times-Series Forecasting Models on Sales Data with Seasonality (Ali & Nakti, 2023)	ARIMA SARIMA LSTM Fb-Prophet	SARIMA
7	Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM (Falatouri, Darbanian, Brandtner, & Udokwu, 2022)	SARIMA LSTM	SARIMA กรณีมีพฤติกรรมตามฤดูกาล LSTM กรณีมีความต้องการคงที่
8	Retail Demand Forecasting using CNN- LSTM Model (Nithin, Rajasekar, Jayanthy, Karthik, & Rithick, 2022)	CNN-LSTM ใช้ Activation Function แบบ 1.Swish Activation Function 2.Activation Function ReLU	CNN-LSTM ที่ใช้ Activation Function แบบ Swish Activation Function
9	Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM (Dave, Leonardo, Jeanice, & Hanafiah, 2021)	ARIMA LSTM Hybrid LSTM- ARIMA	Hybrid LSTM- ARIMA
10	Demand Forecasting in Production Planning for Dairy Products using Machine Learning and Statistical Method (Vithitsoontorn & Chongstitvatana, 2022)	ARIMA LSTM	ARIMA ดีกว่าในข้อมูลแนวโน้มขึ้น (Uptrend) LSTM ดีกว่าในข้อมูลแนวโน้มขาลง (Downtrend) หรือชุดข้อมูลที่ซับซ้อนมาก

บทที่ 3
การดำเนินการวิจัย

ในงานวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. กระบวนการทำงานของแบบจำลอง
2. การเก็บรวบรวมข้อมูล (Data Collection)
3. การสำรวจข้อมูล (Exploratory Data Analysis: EDA)
4. การเตรียมข้อมูล (Data Preparation)
5. การสร้างแบบจำลองพยากรณ์

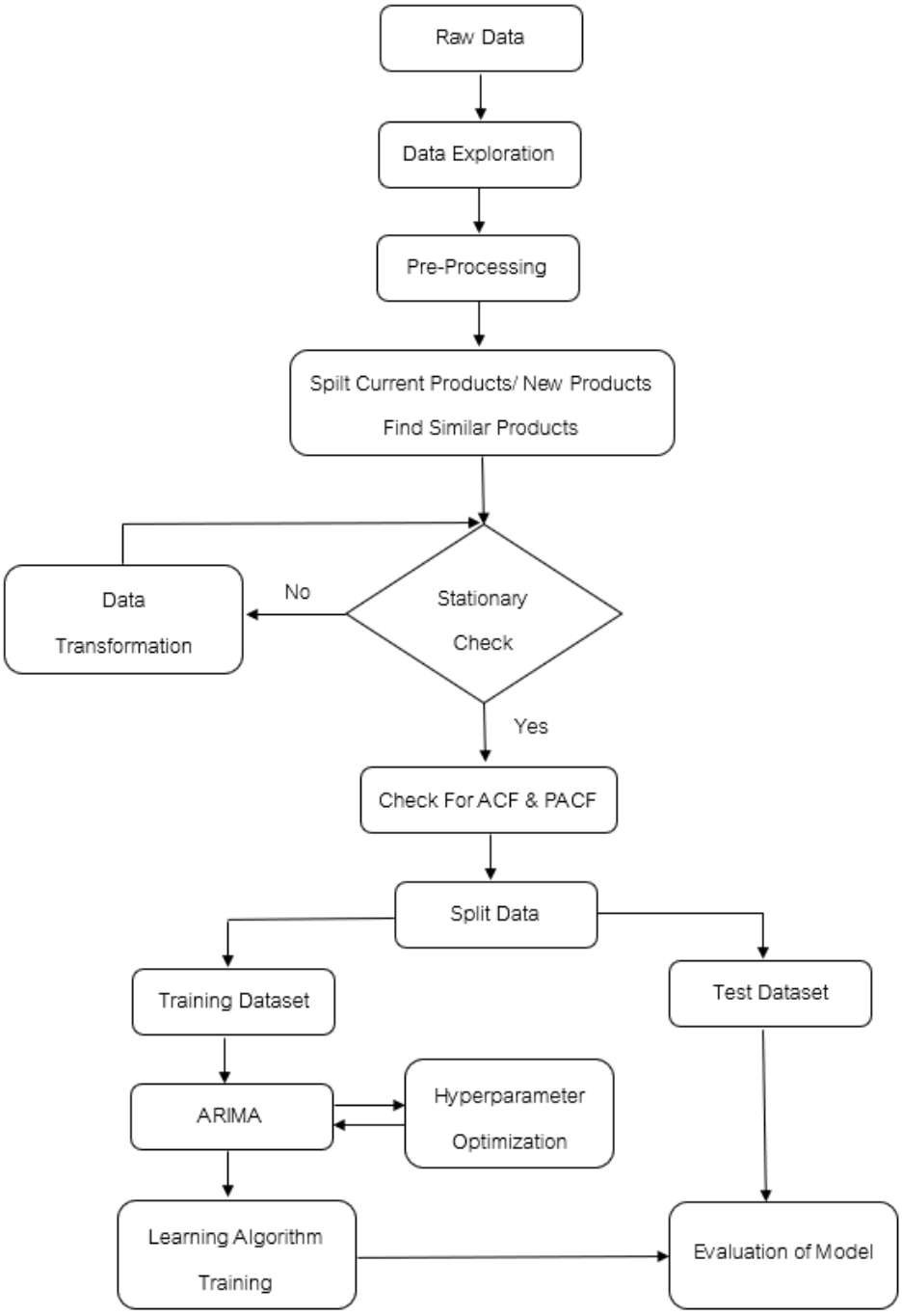
3.1 กระบวนการทำงานของแบบจำลอง



ภาพประกอบ 2 แสดง Flowchart กระบวนการสร้างแบบจำลอง LSTM

จากภาพประกอบที่ 2 Flowchart อธิบายถึงกระบวนการสร้างแบบจำลอง โดยเริ่มจากขั้นตอนการนำเข้าข้อมูล การสำรวจข้อมูล (Exploratory Data Analysis : EDA) ทำความสะอาดข้อมูลและแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ต่อได้ จากนั้นทำการแบ่งข้อมูลออกเป็นสินค้าปัจจุบันและสินค้าใหม่ หาสินค้าปัจจุบันที่มีความคล้ายคลึงกันกับสินค้าใหม่ และแบ่งข้อมูล (Split Data) เป็น 2 ส่วน คือ ข้อมูลฝึกสอน (Training Set) และข้อมูลทดสอบ (Test Set) โดยกำหนดให้ข้อมูลของสินค้าปัจจุบันเป็นข้อมูลฝึกสอน และข้อมูลของสินค้าใหม่เป็นข้อมูลทดสอบ

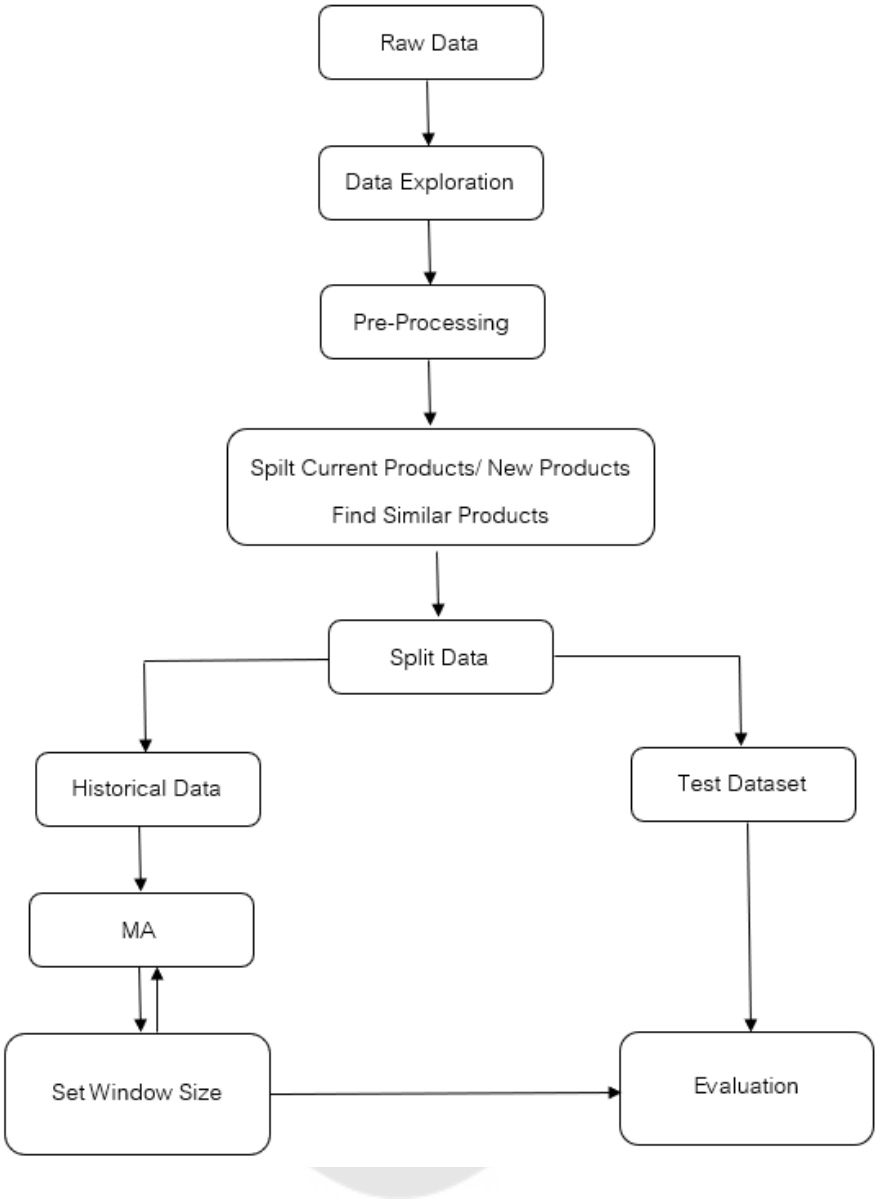
ขั้นตอนการสร้างแบบจำลองในการทำนาย โดยใช้เทคนิค Long Short-Term Memory (LSTM) ซึ่งมีประสิทธิภาพในการพยากรณ์แบบช่วงเวลา และทำการปรับจูนพารามิเตอร์ (Hyperparameter Optimization) เพื่อเพิ่มประสิทธิภาพของแบบจำลองจากการประเมินผลประสิทธิภาพของแบบจำลอง (Evaluation) และนำเสนอผลการวัดประสิทธิภาพแบบจำลองในการพยากรณ์ความต้องการสินค้าใหม่ผ่านข้อมูลทดสอบ (Test Set) โดยใช้ค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE)



ภาพประกอบ 3 แสดง Flowchart กระบวนการสร้างแบบจำลอง ARIMA

จากภาพประกอบที่ 3 Flowchart ได้อธิบายถึงกระบวนการสร้างแบบจำลอง โดยเริ่มจากขั้นตอนการนำเข้าข้อมูล การสำรวจข้อมูล (Exploratory Data Analysis : EDA) ทำความสะอาดข้อมูลและแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ต่อได้ จากนั้นทำการแบ่งข้อมูลออกเป็นสินค้าปัจจุบันและสินค้าใหม่ หาสินค้าปัจจุบันที่มีความคล้ายคลึงกันกับสินค้าใหม่ และแบ่งข้อมูล (Split Data) เป็น 2 ส่วน คือ ข้อมูลฝึกสอน (Training Set) และข้อมูลทดสอบ (Test Set) โดยกำหนดให้ข้อมูลของสินค้าปัจจุบันเป็นข้อมูลฝึกสอน และข้อมูลของสินค้าใหม่เป็นข้อมูลทดสอบ

ขั้นตอนการสร้างแบบจำลองในการทำนาย โดยใช้เทคนิค Autoregressive Integrated Moving Average (ARIMA) ซึ่งมีประสิทธิภาพในการพยากรณ์แบบช่วงเวลา และทำการปรับจูนพารามิเตอร์ (Hyperparameter Optimization) เพื่อเพิ่มประสิทธิภาพของแบบจำลองจากการประเมินผลประสิทธิภาพของแบบจำลอง (Evaluation) และนำเสนอผลการวัดประสิทธิภาพแบบจำลองในการพยากรณ์ความต้องการสินค้าใหม่ผ่านข้อมูลทดสอบ (Test Set) โดยใช้ค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE)



ภาพประกอบ 4 แสดง Flowchart กระบวนการสร้างแบบจำลองด้วยวิธี MA

จากภาพประกอบที่ 4 Flowchart ได้อธิบายถึงกระบวนการสร้างแบบจำลอง โดยเริ่มจากขั้นตอนการนำเข้าข้อมูล การสำรวจข้อมูล (Exploratory Data Analysis : EDA) ทำความสะอาดข้อมูลและแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ต่อได้ จากนั้นทำการแบ่งข้อมูลออกเป็นสินค้าปัจจุบันและสินค้าใหม่ หาสินค้าปัจจุบันที่มีความคล้ายคลึงกันกับสินค้าใหม่ และแบ่งข้อมูล (Split Data) เป็น 2 ส่วน คือ ข้อมูลย้อนหลัง (Historical Data) และข้อมูลทดสอบ (Test Set) โดยกำหนดให้ข้อมูลของสินค้าปัจจุบันเป็นข้อมูลที่ใช้ในการคำนวณค่าเฉลี่ยในการนำมาพยากรณ์ และข้อมูลของสินค้าใหม่เป็นข้อมูลทดสอบ

ขั้นตอนในการพยากรณ์ โดยใช้เทคนิค Moving Averages (MA) ซึ่งมีประสิทธิภาพในการพยากรณ์แบบช่วงเวลา ทำการ Set Window Size เพื่อเลือกจำนวนข้อมูลที่ใช้ในการคำนวณค่าเฉลี่ยในการนำมาพยากรณ์ จากนั้นทำการประเมินผลประสิทธิภาพการพยากรณ์ (Evaluation) และนำเสนอผลการวัดประสิทธิภาพในการพยากรณ์ความต้องการสินค้าใหม่ผ่านข้อมูลทดสอบ (Test Set) โดยใช้ค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE)

ขั้นตอนสุดท้ายนำแบบจำลองแต่ละอัลกอริทึม คือ Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA) และการพยากรณ์อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA) มาทำการเปรียบเทียบประสิทธิภาพของแบบจำลองโดยใช้ค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE)

3.2 **การเก็บรวบรวมข้อมูล** (Data Collection)

ในงานวิจัยนี้ใช้ข้อมูลจากร้านค้าปลีกแห่งหนึ่งในประเทศไทย ที่มีการขายสินค้าผ่านช่องทางออนไลน์ควบคู่กับการขายผ่านหน้าร้าน มีข้อมูลธุรกรรมการขายที่เป็นยอดขายรายวันตั้งแต่เดือนมกราคม ค.ศ. 2021 ถึงเดือนกันยายน ค.ศ. 2023 โดยมีข้อมูล 1 ตาราง เป็นรายการธุรกรรมการขาย มีคอลัมน์ทั้งหมด 28 คอลัมน์ และมีจำนวนข้อมูล 55,524 แถว ดังตารางที่ 3

ตาราง 3 แสดงตัวแปรของข้อมูลรายการธุรกรรมการขาย

Variable	Description
CF_BRANCH_NAME	ชื่อร้านสาขา
CF_PERSONID	รหัสร้านสาขา
CF_FNAME	รูปแบบการชำระเงิน
CF_SYSPERSONID	รหัสพนักงาน
CF_ITEMID	รหัสสินค้า
CF_ITEMNAME	ชื่อสินค้า
CF_BASEQUANTITY	จำนวนยอดขายตามหน่วยขายสินค้า
CF_UNITNAME	หน่วยขายสินค้า เป็น Pcs. หรือชิ้น
CF_PRICE	ราคาสินค้า
CF_AMOUNT	ราคารวมของสินค้า
CF_AVGCOST	ต้นทุนเฉลี่ย
CF_VOUCHER_INCOME	จำนวนบัตรกำนัล
CF_VOUCHER_VOUCHERTYPE	ประเภทบัตรกำนัล
TRANDATE	วันที่และเวลาที่ทำธุรกรรม
CF_TRANNO	เลขที่การทำธุรกรรม
CF_SYSITEMID	รหัสระบบ
CF_SYSITEMGROUPL1ID	กลุ่มระบบที่ 1
CF_SYSITEMGROUPL2ID	กลุ่มระบบที่ 2
CF_SYSITEMGROUPL3ID	กลุ่มระบบที่ 3
CF_SYSITEMGROUPL4ID	กลุ่มระบบที่ 4
CF_SYSITEMGROUPL5ID	กลุ่มระบบที่ 5

ตาราง 3 (ต่อ)

Variable	Description
CF_ITEMGROUPL1_GROUPNAME	หมวดหมู่สินค้าลำดับที่ 1
CF_ITEMGROUPL2_GROUPNAME	หมวดหมู่สินค้าลำดับที่ 2
CF_ITEMGROUPL3_GROUPNAME	หมวดหมู่สินค้าลำดับที่ 3
CF_ITEMGROUPL4_GROUPNAME	หมวดหมู่สินค้าลำดับที่ 4
CF_ITEMGROUPL5_GROUPNAME	หมวดหมู่สินค้าลำดับที่ 5
CF_COMPANY	ใบแจ้งยอดเงิน
CF_TRANDATE	วันที่ทำธุรกรรม

3.3 **การสำรวจข้อมูล** Exploratory Data Analysis (EDA)

การสำรวจข้อมูล และวิเคราะห์ข้อมูลธุรกรรมการขายจากข้อมูลจากร้านค้าปลีกแห่งหนึ่งในประเทศไทย ที่มีการขายสินค้าผ่านช่องทางออนไลน์ควบคู่กับการขายผ่านหน้าร้าน โดยมีข้อมูลยอดขายรายวันตั้งแต่เดือนมกราคม ค.ศ. 2021 ถึงเดือนกันยายน ค.ศ. 2023 แสดงตัวอย่างข้อมูลดังภาพประกอบที่ 5

_BRANCH_NAME	CF_PERSONID	CF_FNAME	CF_SYSPERSONID	CF_ITEMID	CF_ITEMNAME	CF_BASEQUANTITY	CF_UNITNAME	CF_PRICE	CF_AMOUNT	...	CF_SYSITEMGROUPL3ID	CF_SYSITEMGROUPL4ID	CF_SYSITEMGROUPL5ID	CF_ITEMGROUPL1_GROUPNAME
Head Branch	999	เงินสด	0	2001000203079	ครก	1	Pcs.	120	120	...	NaN	NaN	NaN	ของใช้ในบ้านและครัว
Head Branch	999	เงินสด	0	2001000203079	ครก	1	Pcs.	120	120	...	NaN	NaN	NaN	ของใช้ในบ้านและครัว
Head Branch	999	เงินสด	0	2001000203079	ครก	1	Pcs.	120	120	...	NaN	NaN	NaN	ของใช้ในบ้านและครัว
Head Branch	999	เงินสด	0	2001000203079	ครก	1	Pcs.	120	120	...	NaN	NaN	NaN	ของใช้ในบ้านและครัว
Head Branch	999	เงินสด	0	2001000203079	ครก	1	Pcs.	120	120	...	NaN	NaN	NaN	ของใช้ในบ้านและครัว

ภาพประกอบ 5 แสดงตัวอย่างข้อมูล

เริ่มกระบวนการ EDA เพื่อหาข้อมูลเชิงลึกจากข้อมูล จากการแสดงข้อมูลด้วยสถิติเชิงพรรณา ดังภาพประกอบที่ 6 พบว่า

1. ชุดข้อมูล ประกอบด้วย 55,524 แถว 28 คอลัมน์

2. มีสินค้าจำนวน 1,776 SKU

3. มีหมวดหมู่สินค้าจำนวน 54 หมวดหมู่

4. หมวดหมู่ผลิตภัณฑ์เปลี่ยนสีผมเป็นหมวดหมู่สินค้าที่ขายสุด

5. ไม่มี VOUCHER ส่วนลด หรือ โปรโมชั่น

6. ค่า Min ของ CF_BASEQUANTITY, CF_PRICE, CF_AMOUNT เป็น 0 จึงทำการเช็คข้อมูลแถวที่เป็น 0 ในคอลัมน์ดังกล่าว พบว่า มีจำนวนยอดขายเป็นจำนวนชิ้น แต่ไม่มีราคาขาย ทำให้ผลรวมยอดขายเป็น 0 จำนวน 2 แถว และมีจำนวนยอดขายเป็นจำนวนชิ้นเท่ากับ 0 แต่มีราคาขาย ทำให้ผลรวมยอดขายเป็น 0 จำนวน 2 แถว รวมทั้งสองกรณีเป็น 4 แถว จึงทำการลบแถวนี้ออก

	CF_BRANCH_NAME	CF_FNAME	CF_ITEMID	CF_ITEMNAME	CF_UNITNAME	CF_VOUCHER_VOUCHERTYPE	CF_TRANNO	CF_ITEMGROUPL1_GROUPNAME	CF_COMPANY
count	55254	55254	55254	55254	55254	55254	55254	55254	55254
unique	1	1	1874	1776	1	1	24587	54	1
top	Head Branch	เงินสด	2001869660143	รวมสินค้า 20บาท	Pcs.	NN	ORCM6301000208	ผลิตภัณฑ์เปลี่ยนสีผม	ใบแจ้งยอดเงิน
freq	55254	55254	3050	3050	55254	55254	23	8336	55254

	CF_PERSONID	CF_SYSPERSONID	CF_BASEQUANTITY	CF_PRICE	CF_AMOUNT	CF_AVGCOST	CF_VOUCHER_INCOME	CF_SYSITEMID	CF_SYSITEMGROUPL1ID	CF_SYSITEMGROUPL2ID
count	55254.0	55254.0	55254.000000	55254.000000	55254.000000	55254.000000	55254.0	55254.000000	55254.000000	0.0
mean	999.0	0.0	1.072719	49.275437	51.196167	38.049129	1.0	1479.622797	45.839577	NaN
std	0.0	0.0	0.529885	50.647261	57.559021	38.449847	0.0	970.503938	30.143651	NaN
min	999.0	0.0	0.000000	0.000000	0.000000	0.000000	1.0	15.000000	3.000000	NaN
25%	999.0	0.0	1.000000	20.000000	20.000000	16.800000	1.0	765.000000	21.000000	NaN
50%	999.0	0.0	1.000000	35.000000	39.000000	28.000000	1.0	1367.000000	45.000000	NaN
75%	999.0	0.0	1.000000	49.000000	55.000000	42.000000	1.0	2040.000000	73.000000	NaN
max	999.0	0.0	22.000000	890.000000	3582.000000	800.000000	1.0	4451.000000	100.000000	NaN

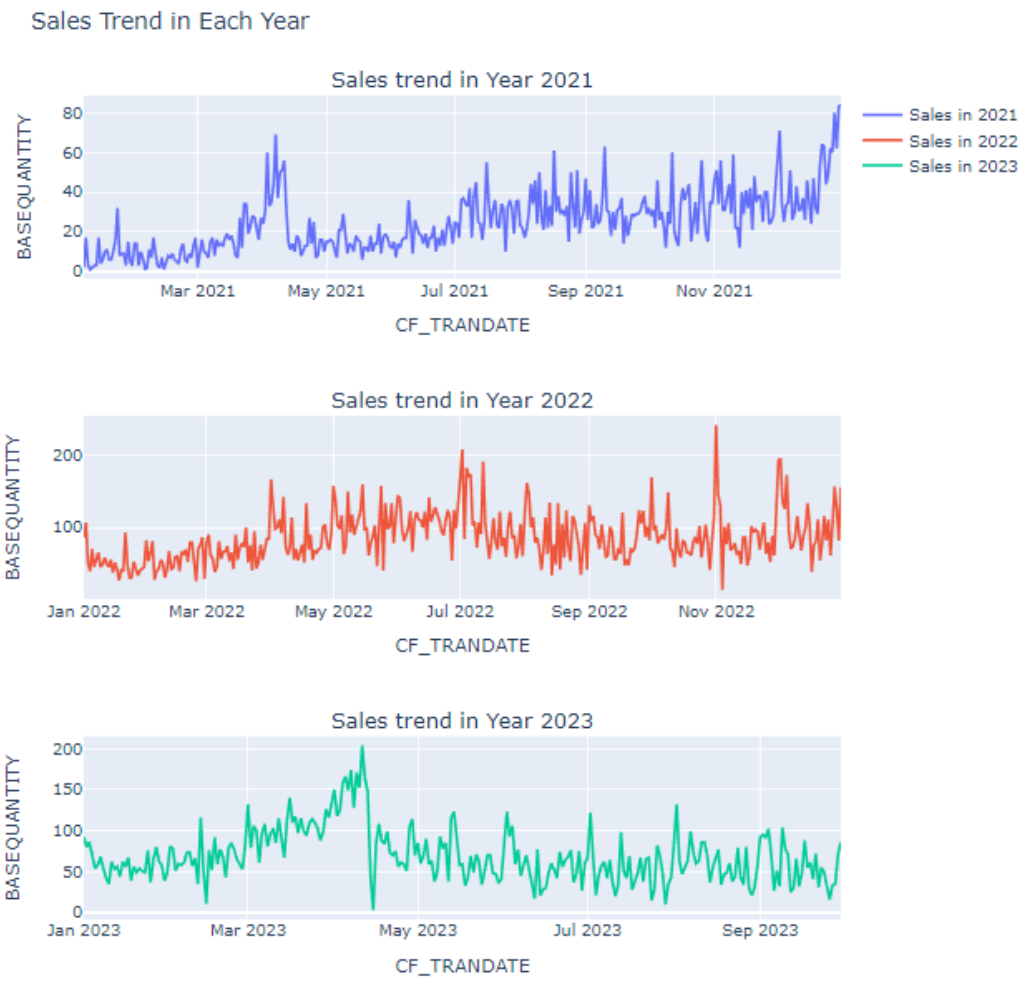
ภาพประกอบ 6 แสดงข้อมูลด้วยสถิติเชิงพรรณา

ทำการวิเคราะห์ข้อมูล พบว่า มีคอลัมน์ที่มีค่าขาดหายไปทั้งหมดจำนวน 8 คอลัมน์ ดังนั้น จึงทำการลบคอลัมน์ดังกล่าวออก โดยลบตั้งแต่คอลัมน์ CF_SYSITEMGROUPL2ID ถึง CF_SYSITEMGROUPL5ID และลบคอลัมน์ตั้งแต่ CF_ITEMGROUPL2_GROUPNAME ถึง CF_ITEMGROUPL5_GROUPNAME นอกจากนี้ยังทำการลบคอลัมน์ที่ไม่จำเป็นต่อการวิเคราะห์ข้อมูล เพื่อหาข้อมูลเชิงลึกอีกเป็นจำนวน 2 คอลัมน์ ได้แก่ คอลัมน์ CF_VOUCHER_INCOME และ CF_VOUCHER_VOUCHERTYPE เนื่องจากข้อมูลทั้ง 2 คอลัมน์ มีค่าเริ่มต้นเหมือนกันทั้งหมด ดังภาพประกอบที่ 7

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 55254 entries, 0 to 55253
Data columns (total 28 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   CF_BRANCH_NAME            55254 non-null  object
 1   CF_PERSONID               55254 non-null  int64
 2   CF_FNAME                  55254 non-null  object
 3   CF_SYSPERSONID            55254 non-null  int64
 4   CF_ITEMID                 55254 non-null  object
 5   CF_ITEMNAME               55254 non-null  object
 6   CF_BASEQUANTITY           55254 non-null  int64
 7   CF_UNITNAME               55254 non-null  object
 8   CF_PRICE                  55254 non-null  int64
 9   CF_AMOUNT                 55254 non-null  int64
 10  CF_AVGCOST                55254 non-null  float64
 11  CF_VOUCHER_INCOME         55254 non-null  int64
 12  CF_VOUCHER_VOUCHERTYPE    55254 non-null  object
 13  TRANDATE                  55254 non-null  datetime64[ns]
 14  CF_TRANNO                 55254 non-null  object
 15  CF_SYSITEMID              55254 non-null  int64
 16  CF_SYSITEMGROUPL1ID       55254 non-null  int64
 17  CF_SYSITEMGROUPL2ID       0 non-null      float64
 18  CF_SYSITEMGROUPL3ID       0 non-null      float64
 19  CF_SYSITEMGROUPL4ID       0 non-null      float64
 20  CF_SYSITEMGROUPL5ID       0 non-null      float64
 21  CF_ITEMGROUPL1_GROUPNAME  55254 non-null  object
 22  CF_ITEMGROUPL2_GROUPNAME  0 non-null      float64
 23  CF_ITEMGROUPL3_GROUPNAME  0 non-null      float64
 24  CF_ITEMGROUPL4_GROUPNAME  0 non-null      float64
 25  CF_ITEMGROUPL5_GROUPNAME  0 non-null      float64
 26  CF_COMPANY                55254 non-null  object
 27  CF_TRANDATE               55254 non-null  datetime64[ns]
dtypes: datetime64[ns](2), float64(9), int64(8), object(9)
memory usage: 11.8+ MB
```

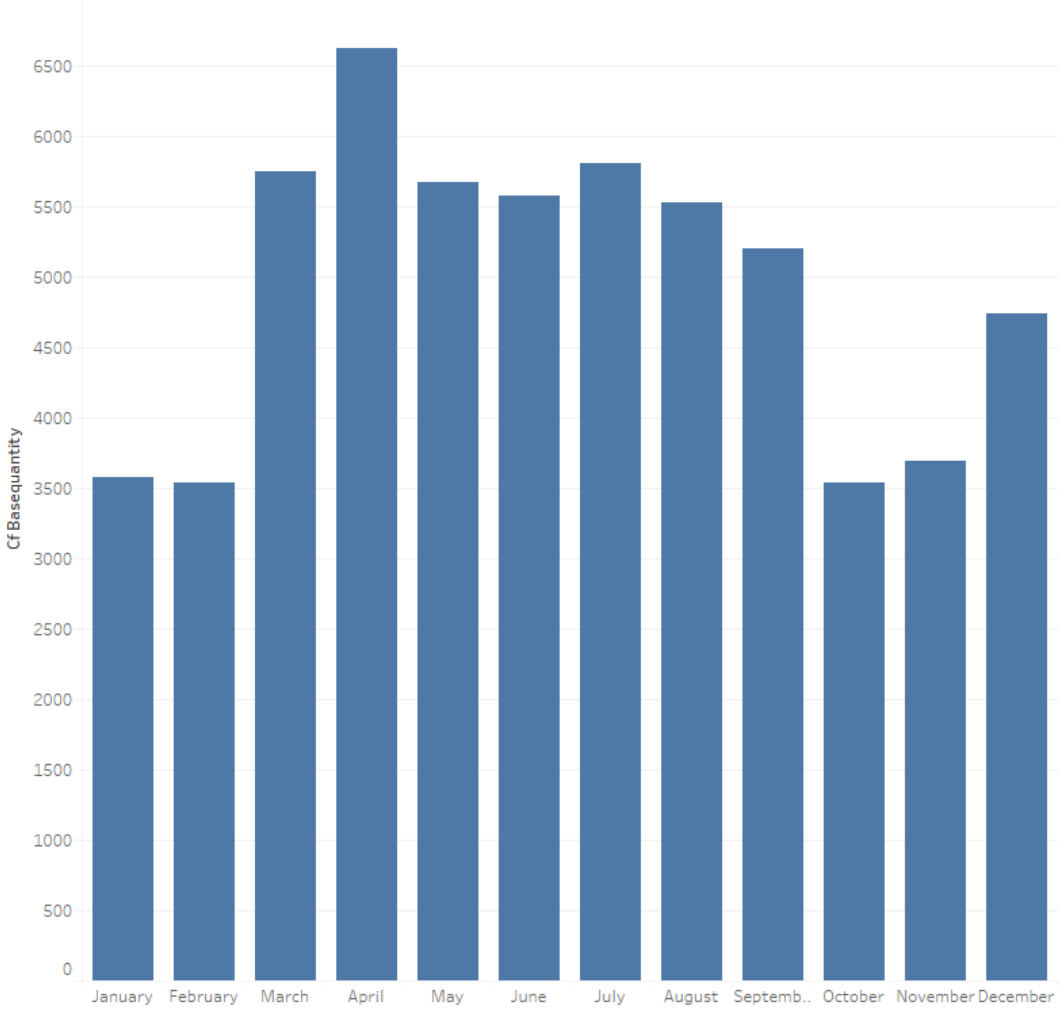
ภาพประกอบ 7 แสดงรายละเอียดข้อมูลแต่ละคอลัมน์

ทำการวิเคราะห์แนวโน้มจำนวนยอดขายสินค้ารวมทุก SKU ตามจำนวนหน่วยขายเป็นจำนวนชิ้นแยกรายปี ตั้งแต่ปี ค.ศ. 2021–2023 พบว่า จำนวนยอดขายมีแนวโน้มสูงในเดือนเมษายน คาดว่าเนื่องจากมีวันหยุดราชการในเทศกาลวันสงกรานต์หลายวัน ทำให้ลูกค้ามีการซื้อสินค้าจำนวนมากในช่วงนี้ ดังภาพประกอบที่ 8



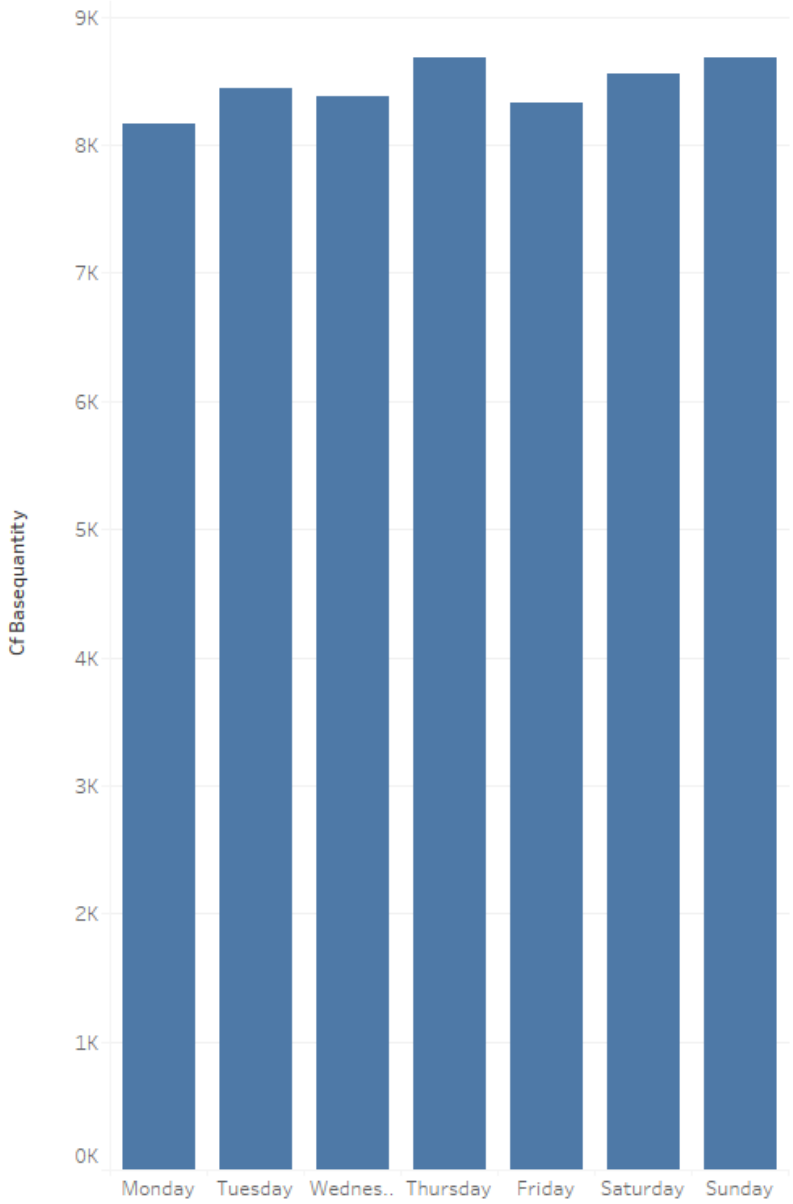
ภาพประกอบ 8 แสดงแนวโน้มจำนวนยอดขายแยกรายปี

ทำการวิเคราะห์จำนวนยอดขายภาพรวมแยกรายเดือน ตามจำนวนหน่วยขายเป็นจำนวนชิ้น พบว่า มีจำนวนยอดขายสูงสุดในเดือนเมษายน และมีจำนวนยอดขายต่ำสุดในเดือนกุมภาพันธ์ ดังภาพประกอบที่ 9



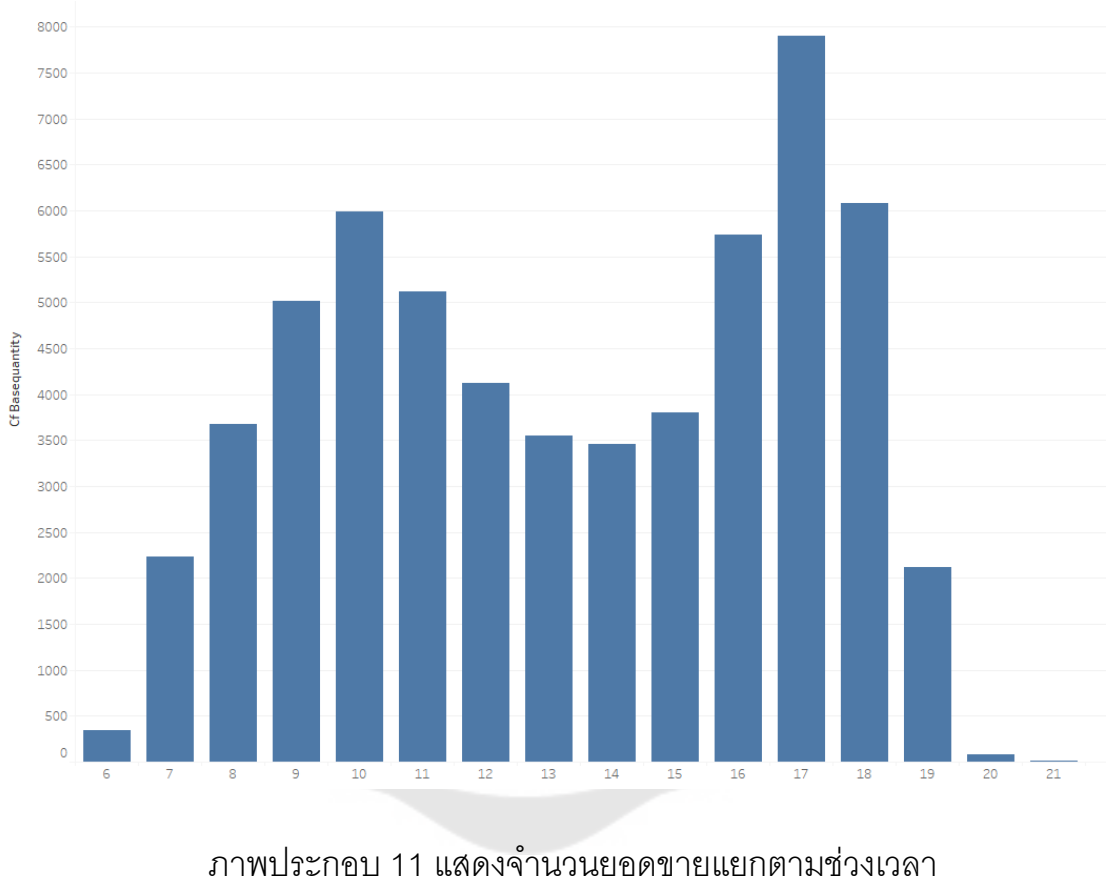
ภาพประกอบ 9 แสดงจำนวนยอดขายแยกรายเดือน

ทำการวิเคราะห์จำนวนยอดขายภาพรวมแยกรายวัน ตามจำนวนหน่วยขายเป็นจำนวนชิ้น พบว่า จำนวนยอดขายในแต่ละวันค่อนข้างใกล้เคียงกัน ซึ่งวันที่มีจำนวนยอดขายสูงสุด คือ วันอาทิตย์ คาดว่าเนื่องมาจากเป็นวันหยุดของทั้งทางราชการและเอกชน ทำให้ลูกค้าซื้อสินค้าในวันอาทิตย์สูงกว่าวันอื่น ๆ และวันที่มีจำนวนยอดขายต่ำสุด คือ วันจันทร์ ดังภาพประกอบที่ 10



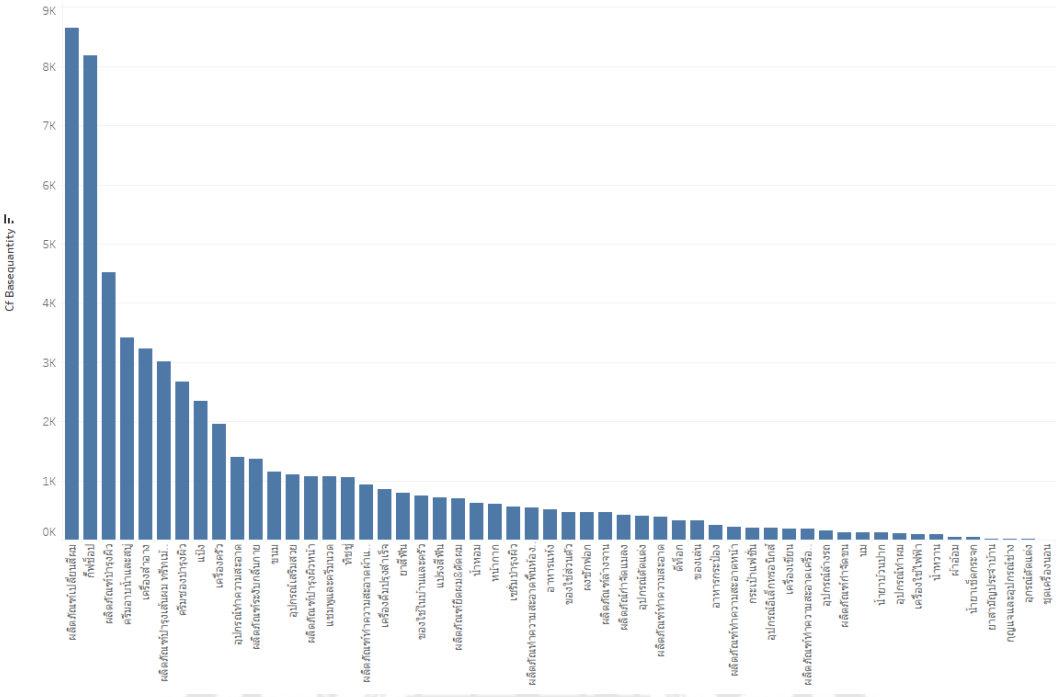
ภาพประกอบ 10 แสดงจำนวนยอดขายแยกรายวัน

ทำการวิเคราะห์จำนวนยอดขายภาพรวมแยกตามช่วงเวลา ตามจำนวนหน่วยขายเป็นจำนวนชิ้น พบว่า ช่วงเวลาที่มีจำนวนยอดขายสูงสุดของช่วงเช้า คือ เวลา 10.00 น. และของช่วงเย็น คือ เวลา 17.00 น. ซึ่งเป็นเวลาที่มีจำนวนยอดขายสูงสุดของวันด้วย และช่วงเวลาที่มีจำนวนยอดขายต่ำสุด คือ 21.00 น. ดังภาพประกอบที่ 11



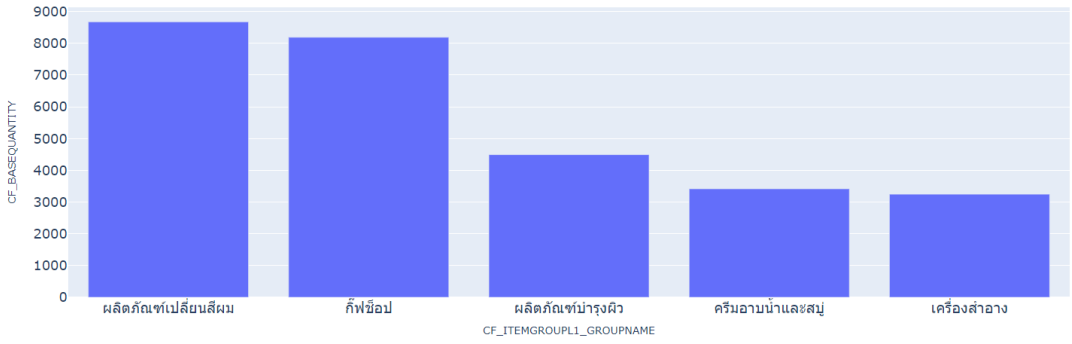
ภาพประกอบ 11 แสดงจำนวนยอดขายแยกตามช่วงเวลา

ทำการวิเคราะห์จำนวนยอดขายภาพรวมในแต่ละหมวดหมู่จากหมวดหมู่ทั้งหมด 54 หมวดหมู่ ตามจำนวนหน่วยขายเป็นจำนวนชิ้น พบว่า หมวดหมู่ผลิตภัณฑ์เปลี่ยนสีผมขายดีที่สุด และหมวดหมู่ที่มียอดขายต่ำสุด คือ หมวดหมู่ชุดเครื่องนอน ดังภาพประกอบที่ 12



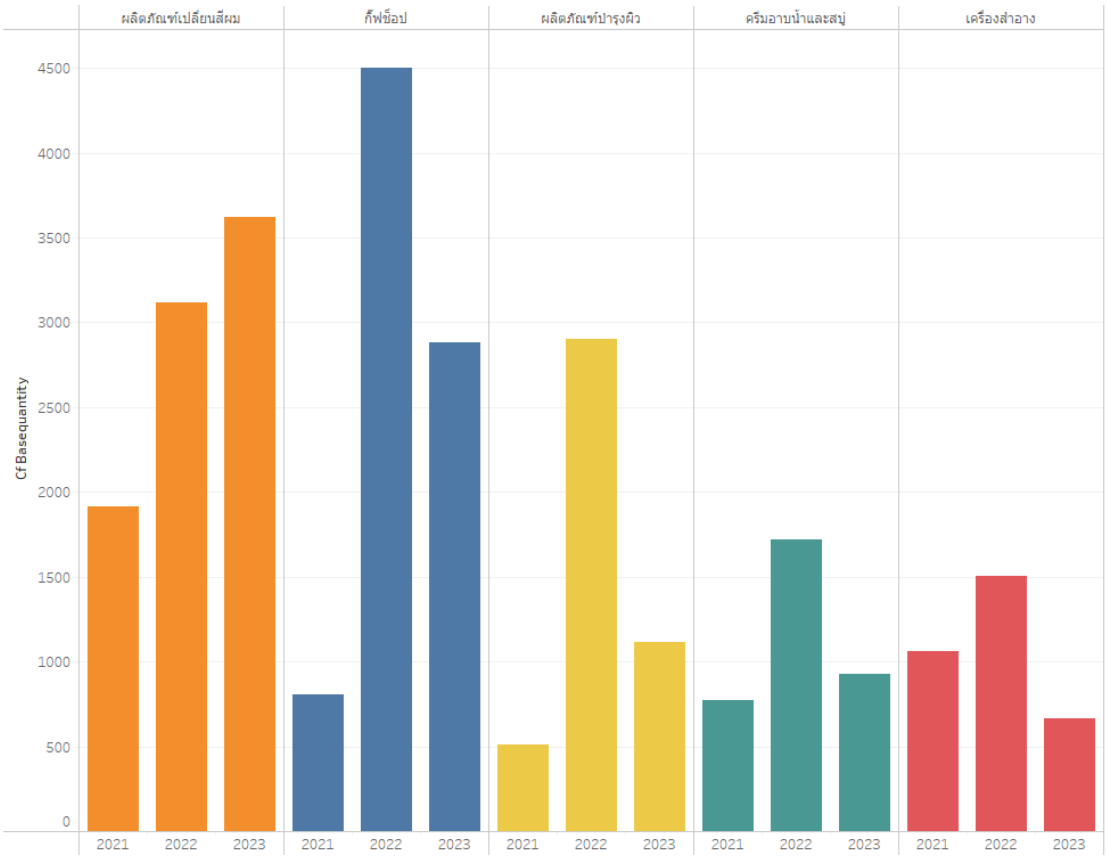
ภาพประกอบ 12 แสดงจำนวนยอดขายแยกตามหมวดหมู่

ทำการวิเคราะห์ยอดขายของหมวดหมู่ที่ขายดี 5 อันดับแรกตามจำนวนหน่วยขายเป็นจำนวนชิ้น คือ ผลิตภัณฑ์เปลี่ยนสีผม, กิ๊ฟช็อป, ผลิตภัณฑ์บำรุงผิว, ครีมอาบน้ำและสบู่ และเครื่องสำอางตามลำดับ ดังภาพประกอบที่ 13



ภาพประกอบ 13 แสดงจำนวนยอดขาย 5 อันดับแรกแยกตามหมวดหมู่

วิเคราะห์ยอดขายของหมวดหมู่ที่ขายดี 5 อันดับแรก แยกรายปี ตามจำนวนหน่วยขาย เป็นจำนวนชิ้น ตั้งแต่ปี ค.ศ. 2021–2023 พบว่า หมวดหมู่ผลิตภัณฑ์เปลี่ยนสีผมมียอดขายสูงสุดในปี ค.ศ. 2023 หมวดหมู่กิ๊ฟช็อปมียอดขายสูงสุดในปี ค.ศ. 2022 ผลิตภัณฑ์บำรุงผิวมียอดขายสูงสุดในปี ค.ศ. 2022 ครีมอาบน้ำและสบู่มียอดขายสูงสุดในปี ค.ศ. 2022 และเครื่องสำอางมียอดขายสูงสุดในปี ค.ศ. 2022 จากข้อสังเกตพบว่าหมวดหมู่ส่วนใหญ่ขายได้สูงสุดในปี ค.ศ. 2022 ดังภาพประกอบที่ 14



ภาพประกอบ 14 แสดงจำนวนยอดขายแยกตามหมวดหมู่และแยกตามปี 5 อันดับแรก

สรุปผลลัพธ์ที่ได้จากการสำรวจข้อมูล Exploratory Data Analysis (EDA)

1. หมวดหมู่ผลิตภัณฑ์เปลี่ยนสีผมเป็นหมวดหมู่สินค้าที่ขายดีที่สุด
2. เดือนเมษายนมีจำนวนยอดขายสูงสุด
3. วันอาทิตย์เป็นวันที่มีจำนวนยอดขายสูงสุด
4. ช่วงเช้ามีจำนวนยอดขายสูงสุด คือ เวลา 10.00 น. และช่วงเย็น คือ เวลา 17.00 น.

3.4 **การเตรียมข้อมูล** (Data Preparation)

การเตรียมข้อมูลสำหรับการฝึกฝนของแบบจำลอง มีขั้นตอนดังนี้

1. ศึกษาและวิเคราะห์ Feature ใน Dataset

2. ทำการ Clean ข้อมูล ด้วยการรวมหมวดหมู่สินค้าที่ใกล้เคียงกันไว้ด้วยกัน จากนั้นทำการลบแถวที่มีผลรวมยอดขายเป็น 0 จำนวน 4 แถว ดังภาพประกอบที่ 15

CF_BASEQUANTITY	CF_UNITNAME	CF_PRICE	CF_AMOUNT	CF_AVGCOST	TRANDATE
1	Pcs.	0	0	58.00	2022-09-14 18:17:34
1	Pcs.	0	0	0.00	2022-09-07 07:14:15
0	Pcs.	49	0	40.00	2021-06-07 16:36:11
0	Pcs.	249	0	134.46	2021-06-07 16:36:11

ภาพประกอบ 15 แสดงแถวที่ทำการ Clean ข้อมูล

และทำการลบคอลัมน์ที่มีค่าที่ขาดหายไป จำนวน 8 คอลัมน์ ดังนี้

คอลัมน์ที่ 1 คือ CF_SYSITEMGROUPL2ID

คอลัมน์ที่ 2 คือ CF_SYSITEMGROUPL3ID

คอลัมน์ที่ 3 คือ CF_SYSITEMGROUPL4ID

คอลัมน์ที่ 4 คือ CF_SYSITEMGROUPL5ID

คอลัมน์ที่ 5 คือ CF_ITEMGROUPL2_GROUPNAME

คอลัมน์ที่ 6 คือ CF_ITEMGROUPL3_GROUPNAME

คอลัมน์ที่ 7 คือ CF_ITEMGROUPL4_GROUPNAME

คอลัมน์ที่ 8 คือ CF_ITEMGROUPL5_GROUPNAME

ดังภาพประกอบที่ 16

CF_SYSITEMGROUPL2ID	CF_SYSITEMGROUPL3ID	CF_SYSITEMGROUPL4ID	CF_SYSITEMGROUPL5ID	CF_ITEMGROUPL2_GROUPNAME	CF_ITEMGROUPL3_GROUPNAME	CF_ITEMGROUPL4_GROUPNAME	CF_ITEMGROUPL5_GROUPNAME
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

ภาพประกอบ 16 แสดงคอลัมน์ที่ทำการ Clean ข้อมูล

จากนั้นลบคอลัมน์ที่มีค่าเริ่มต้นเหมือนกันทั้งหมด 2 คอลัมน์ เนื่องจากไม่มีความจำเป็นในการนำมาวิเคราะห์ข้อมูล เพื่อหาข้อมูลเชิงลึก ได้แก่ คอลัมน์ CF_VOUCHER_INCOME และ CF_VOUCHER_VOUCHERTYPE ดังภาพประกอบที่ 17

CF_VOUCHER_INCOME	CF_VOUCHER_VOUCHERTYPE
1	NN
1	NN
1	NN
1	NN
1	NN
1	NN
1	NN
1	NN
1	NN
1	NN

ภาพประกอบ 17 แสดงคอลัมน์ที่มีค่าเริ่มต้นเหมือนกัน

3. ทำการวิเคราะห์ข้อมูลจำนวนยอดขายของสินค้าทั้งหมดตามหน่วยขายเป็นจำนวนชิ้น แยกในแต่ละช่วงเวลา วัน เดือน ปี และหมวดหมู่

4. ทำการเพิ่มคอลัมน์ first_sale_date เพื่อหาวันเดือนปีที่สินค้ามียอดขายออกครั้งแรก เพื่อนำข้อมูลคอลัมน์นี้มาใช้ในการแบ่งข้อมูลสินค้าปัจจุบันและสินค้าใหม่ จากนั้นเพิ่มคอลัมน์ first_sale_year เพื่อหาปีที่สินค้ามียอดขายออกครั้งแรก ดังภาพประกอบที่ 18

CF_TRANDATE	Year	Month	first_sale_date	first_sale_year
2022-06-19	2022	6	2022-02-01	2022
2022-06-05	2022	6	2022-02-01	2022
2022-06-19	2022	6	2022-02-01	2022

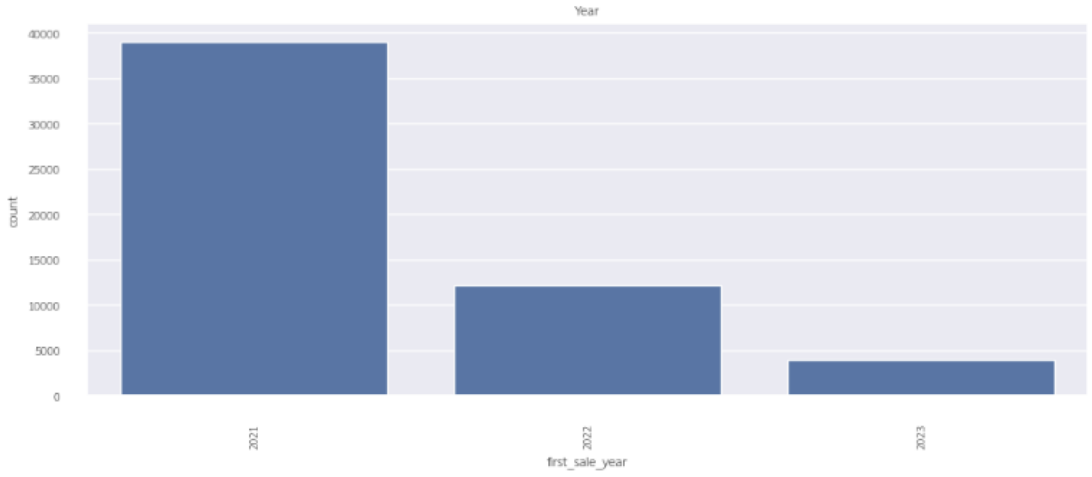
ภาพประกอบ 18 แสดงคอลัมน์ที่ทำการเพิ่ม

ทำการแบ่งข้อมูลออกเป็น 2 ชุด ดังนี้

ข้อมูลชุดที่ 1 กำหนดให้สินค้าที่มียอดขายออกครั้งแรกในปี ค.ศ. 2021-2022 เป็นสินค้าปัจจุบัน จากนั้นทำการลบยอดขายในปี ค.ศ. 2023 ของสินค้าปัจจุบันออก มีจำนวน 1,381 SKU

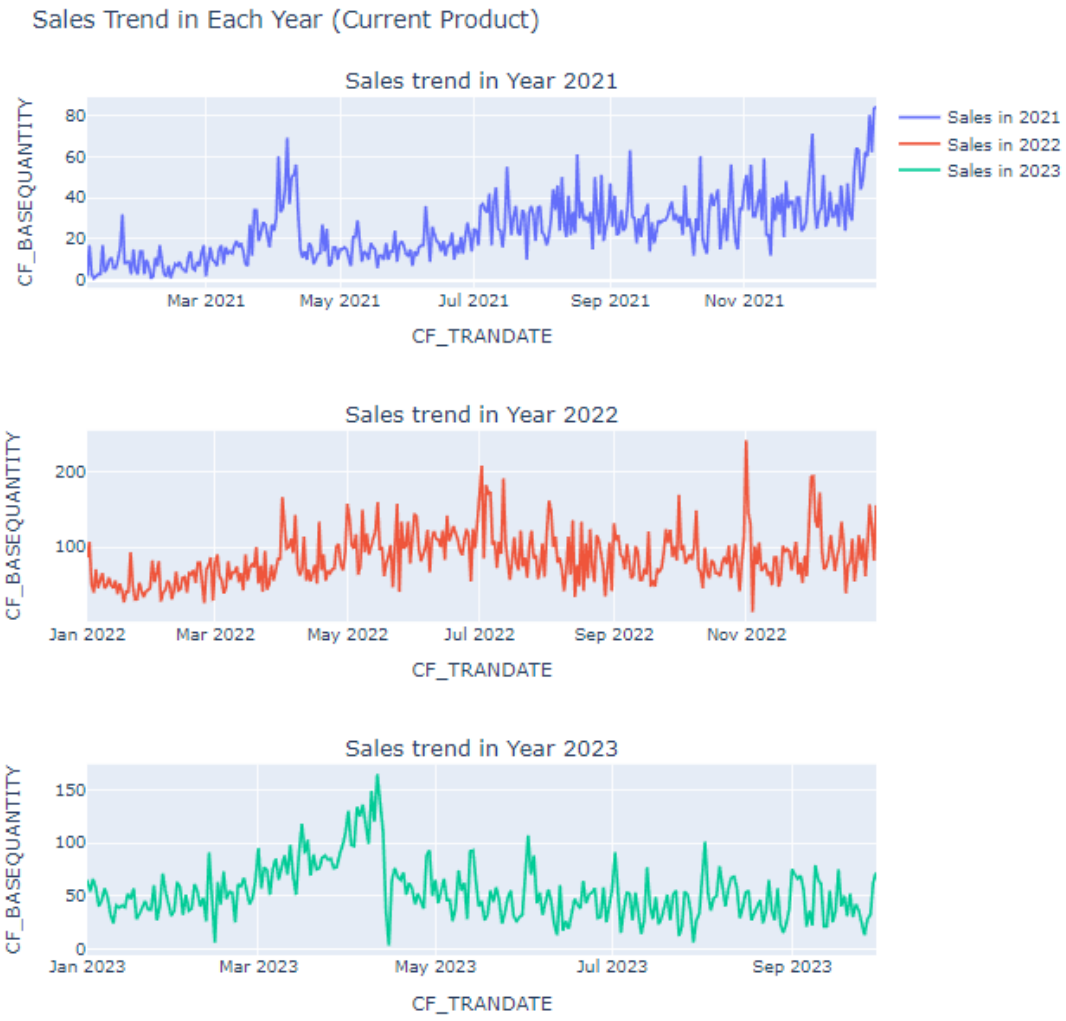
ข้อมูลชุดที่ 2 กำหนดให้สินค้าที่มียอดขายออกครั้งแรกในปี ค.ศ. 2023 เป็นสินค้าใหม่ มีจำนวน 395 SKU

จากนั้นทำการวิเคราะห์จำนวนธุรกรรมการขายของสินค้าที่เริ่มขายในปีนั้น ๆ โดยสินค้าที่มียอดขายออกครั้งแรกในปี ค.ศ. 2021 มีจำนวนธุรกรรมการขายสูงที่สุด รองลงมา คือ สินค้าที่มียอดขายออกครั้งแรกปี ค.ศ. 2022 และ ค.ศ. 2023 ตามลำดับ ดังภาพประกอบที่ 19



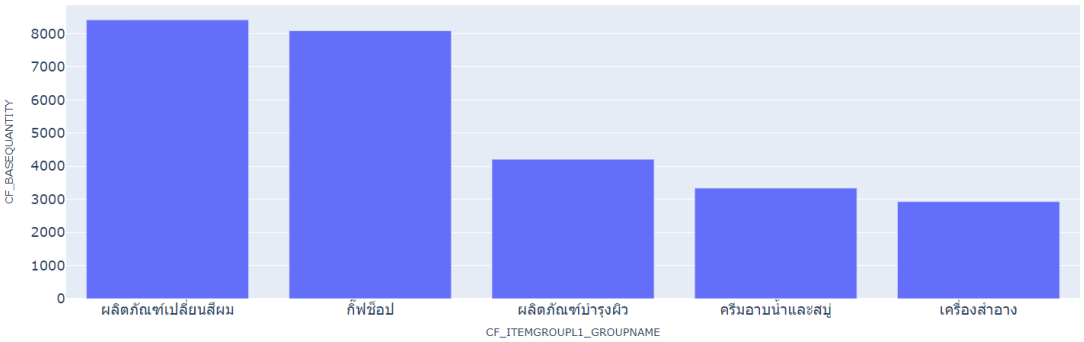
ภาพประกอบ 19 แสดงจำนวนธุรกรรมการขายของสินค้าแยกตามปีที่สินค้าเริ่มวางขาย

ทำการวิเคราะห์ข้อมูลยอดขายสินค้าปัจจุบันทั้งหมดแยกรายปี พบว่า ปี ค.ศ. 2021-2023 มีแนวโน้มยอดขายสูงช่วงเดือนเมษายนของทุกปี และช่วงปลายปี ค.ศ. 2021 มีแนวโน้มยอดขายสูงขึ้น ดังภาพประกอบที่ 20



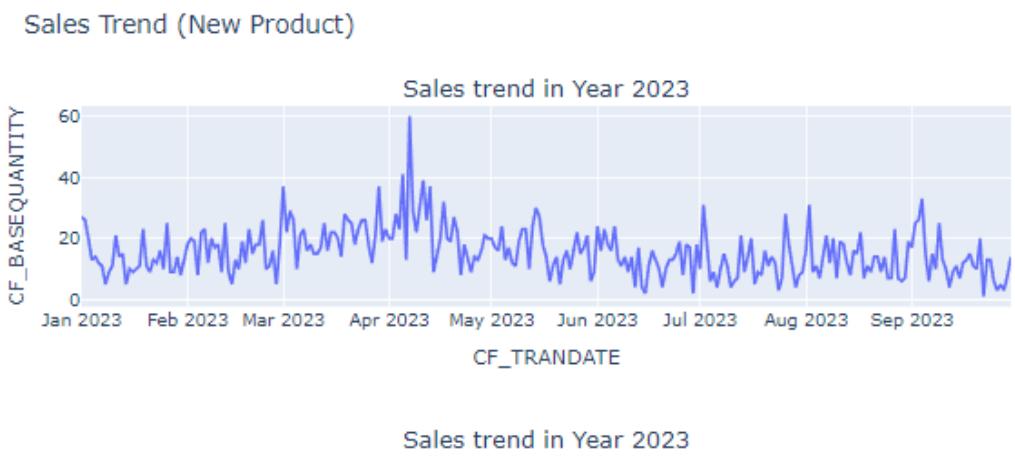
ภาพประกอบ 20 แสดงแนวโน้มจำนวนยอดขายของสินค้าปัจจุบัน

ทำการวิเคราะห์หมวดหมู่สินค้าที่ขายดี 5 อันดับแรกของสินค้าปัจจุบันทั้งหมด ที่มียอดขายสูงสุดตามจำนวนหน่วยขายเป็นจำนวนชิ้น คือ หมวดหมู่ผลิตภัณฑ์เปลี่ยนสีผม, หมวดหมู่กิ๊ฟช็อป, หมวดหมู่ผลิตภัณฑ์บำรุงผิว, หมวดหมู่ครีมอาบน้ำและสบู่ และหมวดหมู่เครื่องสำอาง ตามลำดับ ดังภาพประกอบที่ 21



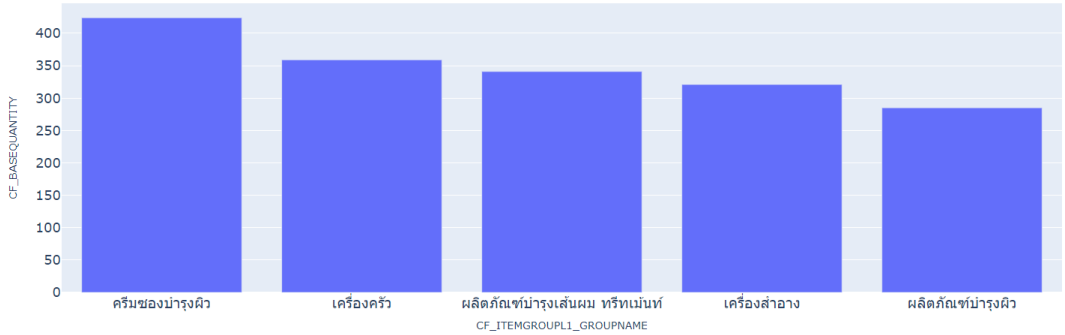
ภาพประกอบ 21 แสดงหมวดหมู่สินค้าที่ขายดีที่สุด 5 อันดันแรก ของสินค้าปัจจุบัน

ทำการวิเคราะห์ข้อมูลยอดขายสินค้าใหม่ทั้งหมด ที่เริ่มวางขายในปี ค.ศ. 2023 มีแนวโน้มยอดขายสูงช่วงเดือนเมษายน และหลังจากนั้นยอดขายค่อนข้างคงที่ไปจนถึงเดือนกันยายน ดังภาพประกอบที่ 22



ภาพประกอบ 22 แสดงแนวโน้มจำนวนยอดขายของสินค้าใหม่

ทำการวิเคราะห์หมวดหมู่สินค้าที่ขายดี 5 อันดับแรกของสินค้าใหม่ทั้งหมด ที่มียอดขายสูงสุดตามจำนวนหน่วยขายเป็นจำนวนชิ้น คือ หมวดหมู่ครีมซองบำรุงผิว, หมวดหมู่เครื่องครัว, หมวดหมู่ผลิตภัณฑ์บำรุงเส้นผม ทรีทเม้นท์, หมวดหมู่ผลิตภัณฑ์บำรุงผิว, หมวดหมู่เครื่องสำอาง และหมวดหมู่ผลิตภัณฑ์บำรุงผิว ตามลำดับ ดังภาพประกอบที่ 23



ภาพประกอบ 23 แสดงหมวดหมู่สินค้าที่ขายดีที่สุด 5 อันดันแรก ของสินค้าใหม่

5. หาสินค้าที่มีความคล้ายคลึงกันระหว่างสินค้าปัจจุบันกับสินค้าใหม่ ที่อยู่ในหมวดหมู่เดียวกัน จำนวน 4 คู่ รวมเป็น 8 SKU โดยทำการแบ่งสินค้าดังนี้

– สินค้าปัจจุบัน จำนวน 4 SKU

– สินค้าใหม่ จำนวน 4 SKU

จากนั้นทำการพยากรณ์ความต้องการสินค้าใหม่จากข้อมูลของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายกัน โดยพยากรณ์ความต้องการสินค้าใหม่ตามจำนวนหน่วยขายเป็นจำนวนชิ้น

ขั้นตอนในการหาสินค้าที่มีความคล้ายคลึงกัน

การหาสินค้าที่มีความคล้ายคลึงกันระหว่างสินค้าปัจจุบันกับสินค้าใหม่ ผู้วิจัยทำการเลือกสินค้าด้วยวิธี Manual โดยหลักเกณฑ์ที่ใช้ในการคัดเลือกสินค้าปัจจุบันที่คล้ายคลึงกับสินค้าใหม่มากที่สุด เพื่อหาตัวแทนของสินค้าใหม่ ดำเนินการตามขั้นตอนต่อไปนี้

ขั้นตอนในการหาสินค้าที่มีความคล้ายคลึงกัน มีขั้นตอนดังต่อไปนี้

1. เลือกสินค้าที่อยู่ในหมวดหมู่เดียวกัน

2. กรณีที่ร้านมีสินค้าปัจจุบันที่เป็นแบรนด์เดียวกันกับสินค้าใหม่ จะพิจารณาสินค้าในอยู่ในแบรนด์เดียวกันก่อน จากนั้นดำเนินการตามขั้นตอนดังต่อไปนี้

2.1 เลือกสินค้าที่การใช้งานเหมือนกัน

2.2 เลือกสินค้าที่มีขนาดใกล้เคียงกันมากที่สุด

2.3 เลือกสินค้าที่มีราคาใกล้เคียงกันมากที่สุด

3. กรณีที่ร้านไม่มีสินค้าปัจจุบันที่เป็นแบรนด์เดียวกันกับสินค้าใหม่ จะพิจารณาสินค้าจากแบรนด์อื่นที่ร้านมีขายอยู่ หากมีมากกว่า 1 แบรนด์จะพิจารณาแบรนด์ที่อยู่ในระดับใกล้เคียงกันมากที่สุด จากนั้นดำเนินการตามขั้นตอนดังต่อไปนี้

3.1 เลือกสินค้าที่การใช้งานเหมือนกัน

3.2 เลือกสินค้าที่มีขนาดใกล้เคียงกันมากที่สุด

3.3 เลือกสินค้าที่มีราคาใกล้เคียงกันมากที่สุด

จากการดำเนินการคัดเลือกสินค้าปัจจุบันที่คล้ายคลึงกับสินค้าใหม่มากที่สุด ด้วยวิธีที่กล่าวข้างต้นนั้น ได้สินค้าปัจจุบันที่เป็นตัวแทนสินค้าใหม่จำนวน 4 SKU ซึ่งสินค้าใหม่และสินค้าปัจจุบันในแต่ละคู่เป็นสินค้าที่อยู่ในหมวดหมู่และแบรนด์เดียวกัน และอยู่ในหมวดหมู่ที่มียอดขายเป็นจำนวนชิ้นสูงสุด 5 หมวดหมู่แรกของสินค้าทั้งหมด แต่สินค้าต่างคู่กันนั้นจะต่างแบรนด์กัน เนื่องจากไม่สามารถเปิดเผยรายชื่อสินค้าเต็ม แบรนด์ และราคาของสินค้าได้ ผู้วิจัยจึงใช้รหัสแทนชื่อสินค้า และแบรนด์ของสินค้าที่คัดเลือกมาว่ามีความคล้ายคลึงกันรายละเอียดดังตารางที่ 4 อธิบายรายละเอียดของสินค้าปัจจุบันและสินค้าใหม่ ได้ดังนี้

– สินค้าปัจจุบันและสินค้าใหม่คู่ที่ 1 เป็นสินค้าที่อยู่ในหมวดหมู่ผลิตภัณฑ์เปลี่ยนสีผมและเป็นสินค้าแบรนด์เดียวกัน (แบรนด์ A) ซึ่งราคาใกล้เคียง

– สินค้าปัจจุบันและสินค้าใหม่คู่ที่ 2 เป็นสินค้าที่อยู่ในหมวดหมู่ผลิตภัณฑ์เปลี่ยนสีผมและเป็นสินค้าแบรนด์เดียวกัน (แบรนด์ B) ซึ่งราคาใกล้เคียง

– สินค้าปัจจุบันและสินค้าใหม่คู่ที่ 3 เป็นสินค้าที่อยู่ในหมวดหมู่ครีมอาบน้ำและสบู่ และเป็นสินค้าแบรนด์เดียวกัน (แบรนด์ C) ซึ่งราคาเท่ากัน

– สินค้าปัจจุบันและสินค้าใหม่คู่ที่ 4 เป็นสินค้าที่อยู่ในหมวดหมู่เครื่องสำอางและเป็นสินค้าแบรนด์เดียวกัน (แบรนด์ D) ซึ่งราคาเท่ากัน

ตาราง 4 แสดงข้อมูลสินค้าปัจจุบัน และสินค้าใหม่ที่มีความคล้ายคลึงกัน

คู่ที่	สินค้าปัจจุบัน	สินค้าใหม่	หมวดหมู่	แบรนด์
1	ครีมย้อมผม C1	ครีมย้อมผม N1	ผลิตภัณฑ์เปลี่ยนสีผม	A
2	ครีมย้อมผม C2	ครีมย้อมผม N2	ผลิตภัณฑ์เปลี่ยนสีผม	B
3	สบู่วิตามิน C3	สบู่วิตามิน N3	ครีมอาบน้ำและสบู่	C
4	ดินสอเขียนคิ้ว C4	ดินสอเขียนคิ้ว N4	เครื่องสำอาง	D

3.5 การสร้างแบบจำลองพยากรณ์

ในงานวิจัยนี้สร้างแบบจำลองในการพยากรณ์ข้อมูล ด้วยภาษา Python และทำการปรับค่าพารามิเตอร์ให้เหมาะสม จากนั้นสร้างแบบจำลองโดยใช้อัลกอริทึม Long Short-Term Memory (LSTM), แบบจำลองอนุกรมเวลาแบบ Autoregressive Integrated Moving Average (ARIMA) และวิธีการพยากรณ์อนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA) โดยเครื่องมือที่ใช้วัดประสิทธิภาพของแบบจำลอง คือ Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) การวัดประสิทธิภาพแบบจำลองแบ่งเป็น 2 ส่วนด้วยกัน คือ วัดประสิทธิภาพในการพยากรณ์ยอดขายใน 3 เดือนแรก ซึ่งค่าพยากรณ์ยอดขายใน 3 เดือนแรกใช้สำหรับประกอบการตัดสินใจในการพิจารณาเลือกสินค้าใหม่เข้ามาวางขาย และวัดประสิทธิภาพในการพยากรณ์ยอดขายใน 1 เดือนแรก ซึ่งค่าพยากรณ์ใน 1 เดือนแรกใช้สำหรับการสั่งซื้อสินค้าใหม่ครั้งแรกมาสต๊อกก่อนเริ่มวางขาย ซึ่งข้อมูลใน Dataset มีข้อมูลยอดขายรายวันเป็นรายการธุรกรรมการขาย ตั้งแต่เดือนมกราคม ค.ศ. 2021 ถึงเดือนกันยายน ค.ศ. 2023 โดยใช้ข้อมูลยอดขายรายเดือนของสินค้าปัจจุบันตั้งแต่ปี ค.ศ. 2021-2022 เป็นข้อมูลฝึกสอน และใช้ข้อมูลยอดขายรายเดือนของสินค้าใหม่ 3 เดือนแรกในปี ค.ศ. 2023 เป็นข้อมูลทดสอบ ในงานวิจัยนี้ใช้ข้อมูล Feature CF_BASEQUANTITY (จำนวนยอดขายเป็นจำนวนชิ้น) เป็นข้อมูล input เพื่อพยากรณ์ความต้องการสินค้าใหม่จากข้อมูลของสินค้าปัจจุบันที่มีความคล้ายคลึงกัน โดยพยากรณ์ความต้องการสินค้าใหม่ตามจำนวนหน่วยขายเป็นจำนวนชิ้น

แบบจำลอง Long Short-Term Memory (LSTM)

การพยากรณ์ความต้องการสินค้าของแบบจำลอง Long Short-Term Memory (LSTM) นั้นทำการแบ่งข้อมูลออกเป็น 2 ชุด ซึ่งการแบ่งชุดข้อมูล Train และ Test เพื่อใช้ในการสร้างและทดสอบแบบจำลอง LSTM โดยใช้ข้อมูลยอดขายรายเดือนของสินค้าปัจจุบันตั้งแต่ปี ค.ศ. 2021-2022 เป็นข้อมูลฝึกสอน และใช้ข้อมูลยอดขายรายเดือนของสินค้าใหม่ 3 เดือนแรกในปี ค.ศ.2023 เป็นข้อมูลทดสอบ มีขั้นตอนดังนี้

1. ใช้ข้อมูลยอดขายสินค้ารายเดือนของสินค้าปัจจุบันตั้งแต่ปี ค.ศ. 2021-2022 เป็นข้อมูลฝึกสอน และใช้ข้อมูลยอดขายรายเดือนของสินค้าใหม่ 3 เดือนแรกในปี ค.ศ. 2023 เป็นข้อมูลทดสอบ

2. ทำการแปลงชุดข้อมูล Train และ Test เป็นรูปแบบของอาร์เรย์ 3 มิติที่เหมาะสำหรับการใช้กับแบบจำลอง LSTM โดยใช้ฟังก์ชัน Reshape

3. สร้างและฝึกโมเดล LSTM โดยใช้ชุดข้อมูล Train ที่ถูกแปลงเป็นรูปแบบของอาร์เรย์

4. ทดลอง Tuning Parameter เพื่อหาค่าพารามิเตอร์ที่ดีที่สุด

4.1 ทดลองกำหนด Timestep ซึ่งเป็นการกำหนดขนาดของช่วงเวลาที่ใช้ในการสร้างข้อมูล input ของแบบจำลอง เพื่อใช้ในการทำนายค่าต่อเนื่องของชุดข้อมูลตามลำดับเวลา เมื่อต้องการทำนายค่าในอนาคตของลำดับเวลา ใช้ข้อมูลจากช่วงเวลาก่อนหน้า (Past Time Steps) เป็นข้อมูล Input เพื่อทำนายค่าในช่วงเวลาถัดไป (Future Time Step) โดยใช้ 3 เดือนก่อนหน้ามาพยากรณ์เดือนถัดไป

4.2 ทดลองปรับแต่งจำนวน Node โดยทดลองใช้จำนวน Node ที่มีค่าตั้งแต่ 8, 16, 32, 50 และ 64

4.3 ทดลองปรับแต่งจำนวน Epoch ที่มีค่าตั้งแต่ 100, 200 และ 1000

4.4 กำหนดค่า Optimize เท่ากับ Adam

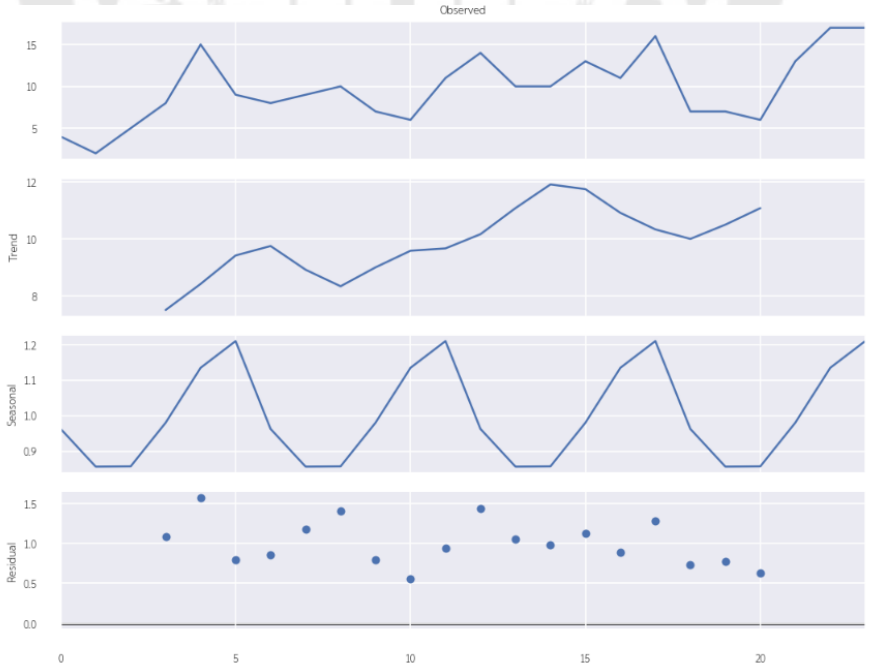
4.5 ข้อมูล Output กำหนดค่าเป็น 3 เพราะต้องการพยากรณ์ความต้องการสินค้าใน 3 เดือนข้างหน้า

จากการทดลองปรับแต่งค่าพารามิเตอร์ในแบบจำลอง LSTM จึงได้ค่าพารามิเตอร์ที่เหมาะสมในการสร้างแบบจำลอง คือ Timestep เท่ากับ 3, Hidden Neuron (n_node) ของ LSTM Layer แรก เท่ากับ 50 และ LSTM Layer ที่สอง เท่ากับ 64, Dense 3 Layer เท่ากับ 32, 16 และ1 ตามลำดับ และ Epoch เท่ากับ 100 จากนั้นนำค่าพารามิเตอร์ที่ได้ไป Fit กับแบบจำลองเพื่อพยากรณ์ยอดขายสินค้าปัจจุบันใน 3 เดือนถัดไป จากนั้นนำค่าพยากรณ์ยอดขายนั้นมาใช้เป็น

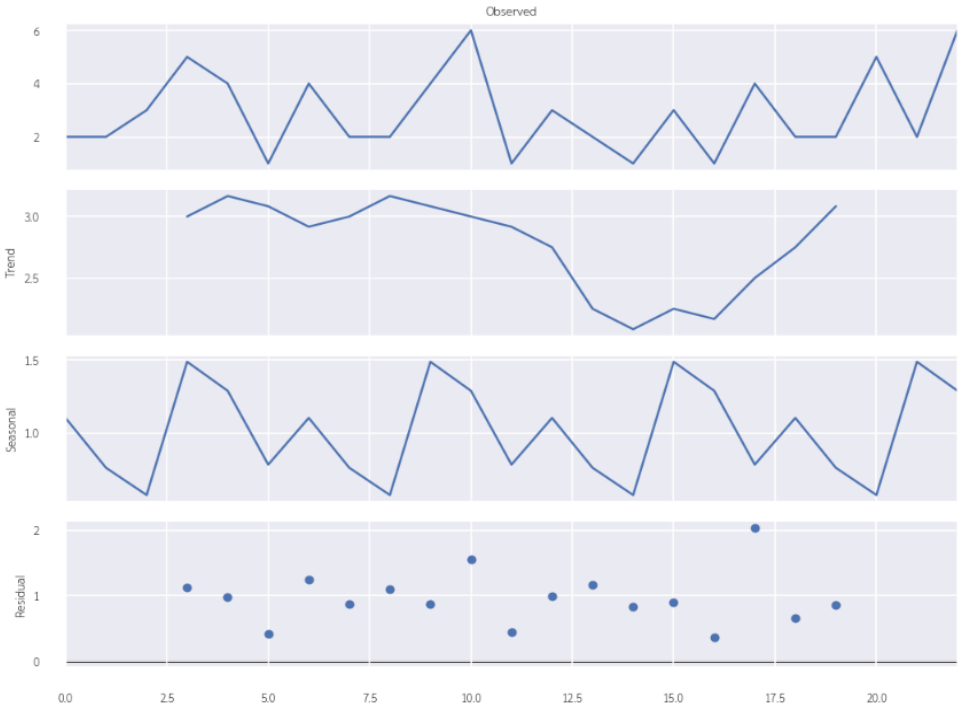
ตัวแทนยอดขาย 3 เดือนแรกของสินค้าใหม่ แล้วนำค่าที่ได้จากการพยากรณ์มาเปรียบเทียบกับข้อมูลทดสอบ (ข้อมูลจริงหรือยอดขาย 3 เดือนแรกของสินค้าใหม่) และทำการหาค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) เพื่อใช้ในการวัดประสิทธิภาพแบบจำลอง

แบบจำลองอนุกรมเวลาแบบ Autoregressive Integrated Moving Average (ARIMA)

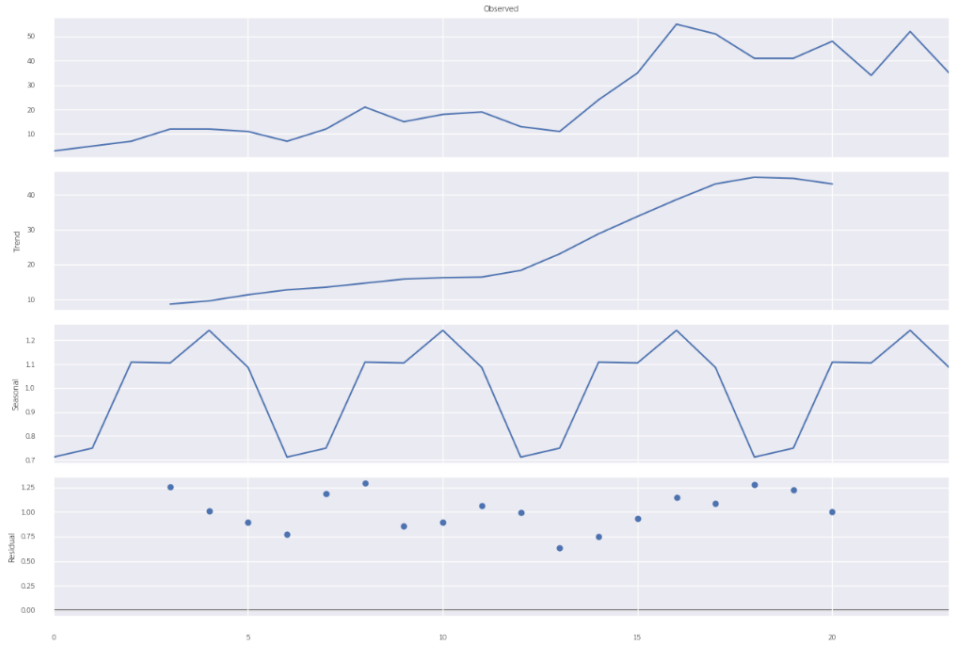
การพยากรณ์ความต้องการสินค้าของแบบจำลอง Autoregressive Integrated Moving Average (ARIMA) นั้นทำการแบ่งข้อมูลออกเป็น 2 ชุด โดยใช้ข้อมูลยอดขายรายเดือนของสินค้าปัจจุบันตั้งแต่ปี ค.ศ. 2021-2022 เป็นข้อมูลฝึกสอน และใช้ข้อมูลยอดขายรายเดือนของสินค้าใหม่ 3 เดือนแรกในปี ค.ศ. 2023 เป็นข้อมูลทดสอบ การทำงานของ ARIMA เหมาะกับข้อมูลอนุกรมเวลาที่มีความคงที่ (Stationary) และนำข้อมูลที่ต้องการพยากรณ์มาใช้ในการ Train จากนั้นนำข้อมูลยอดขาย (หน่วยเป็นจำนวนชิ้น) ที่ทำการแบ่งเป็น Train Set ของสินค้าปัจจุบันที่ตัวแทนของสินค้าใหม่ คือ ครีมย้อมผม C1, ครีมย้อมผม C2, สบู่วิตามิน C3 และดินสอเขียนคิ้ว C4 มาตรวจสอบดูค่า Trend Seasonal และ Residual ดังภาพประกอบที่ 24-27



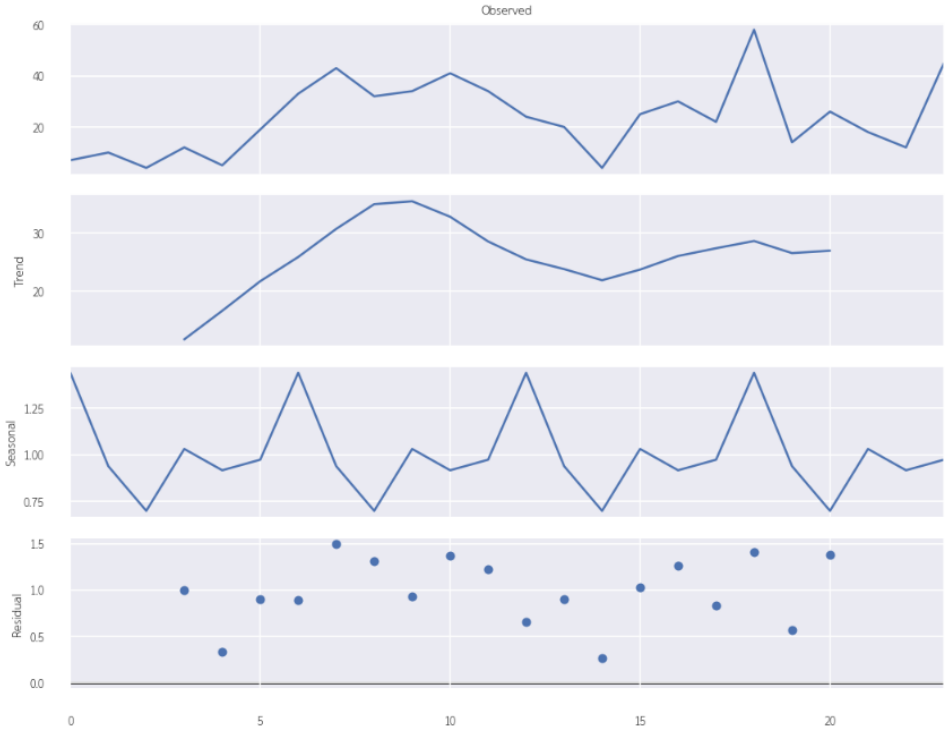
ภาพประกอบ 24 แสดงค่า Trend Seasonal และ Residual ของ Train Data ครีมย้อมผม C1



ภาพประกอบ 25 แสดงค่า Trend Seasonal และ Residual ของ Train Data ครีมย้อมผม C2



ภาพประกอบ 26 แสดงค่า Trend Seasonal และ Residual ของ Train Data สบู่วิตามิน C3



ภาพประกอบ 27 แสดงค่า Trend Seasonal และ Residual ของ Train Data ดินสอเขียนคิ้ว C4

ในงานวิจัยนี้ใช้สถิติทดสอบ Augmented Dickey-Fuller Test (ADF-Test) เทียบกับค่าวิกฤต = 0.05 หากค่า P-value < 0.05 แสดงว่าข้อมูลมีความคงที่ (Stationary) และการหาค่า Parameter p,d,q ใช้ฟังก์ชัน auto_arima() ในการหาค่า Parameter p,d,q ที่ดีที่สุดสำหรับชุดข้อมูลที่กำหนด ที่ให้ค่า AIC ที่น้อยที่สุดเป็น Best Model ค่าพารามิเตอร์ที่ใช้ของแต่ละสินค้าแสดงดังตารางที่ 5

ตาราง 5 แสดงค่าพารามิเตอร์ที่ใช้ของแต่ละสินค้า

Products	ARIMA (p,d,q)
ครีมย้อมผม N1 (ครีมย้อมผม C1)	(1,0,0)
ครีมย้อมผม N2 (ครีมย้อมผม C2)	(1,0,0)
สบู่วิตามิน N3 (สบู่วิตามิน C3)	(1,0,0)
ดินสอเขียนคิ้ว N4 (ดินสอเขียนคิ้ว C4)	(1,0,0)

จากนั้นนำค่าพารามิเตอร์ที่ได้ไป Fit กับแบบจำลอง เพื่อพยากรณ์ความต้องการสินค้าใน 3 เดือนถัดไปของสินค้าปัจจุบัน เพื่อใช้เป็นตัวแทนยอดขาย 3 เดือนแรกของสินค้าใหม่ แล้วนำค่าที่ได้จากการพยากรณ์มาเปรียบเทียบกับข้อมูลทดสอบ (ข้อมูลจริงหรือยอดขาย 3 เดือนแรกของสินค้าใหม่) และทำการหาค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) เพื่อใช้ในการวัดประสิทธิภาพแบบจำลอง

วิธีการพยากรณ์แบบอนุกรมเวลาแบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages)

การพยากรณ์โดยวิธีค่าเฉลี่ยเคลื่อนที่นั้นทำการพยากรณ์โดยทดลองใช้ข้อมูลยอดขายจริงในอดีตของสินค้าปัจจุบันตั้งแต่ปี ค.ศ. 2021-2022 ที่เป็นตัวแทนของสินค้าใหม่ จำนวน 3, 6, 12 และ 24 เดือน มาคำนวณหาค่าเฉลี่ยของยอดขาย เพื่อพยากรณ์ความต้องการสินค้าในช่วงเวลา 3 เดือนถัดไปของสินค้าปัจจุบัน และใช้ค่าพยากรณ์ยอดขายนั้นเป็นตัวแทนยอดขายของสินค้าใหม่ที่เริ่มวางขายใน 3 เดือนแรก

พบว่าการใช้ค่าเฉลี่ยของยอดขายจำนวน 24 เดือนในอดีตในการพยากรณ์ยอดขายอีก 3 เดือนข้างหน้านั้นมีประสิทธิภาพในการพยากรณ์มากที่สุด ดังนั้นการพยากรณ์โดยวิธีค่าเฉลี่ยเคลื่อนที่นั้นใช้ยอดขายในอดีตจำนวน 24 เดือนในการพยากรณ์ (ยกเว้นสินค้าครีมย้อมผม C2 ที่ใช้ยอดขายในอดีตจำนวน 23 เดือน เนื่องจากมียอดขายในปี ค.ศ. 2021–2022 จำนวน 23 เดือน) จากนั้นนำค่าที่ได้จากการพยากรณ์มาเปรียบเทียบกับข้อมูลทดสอบ (ข้อมูลจริงหรือยอดขาย 3 เดือนแรกของสินค้าใหม่) และทำการหาค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) เพื่อใช้ในการวัดประสิทธิภาพแบบจำลอง

บทที่ 4
ผลการดำเนินการวิจัย

ในการวิจัยการพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรกที่เริ่มวางขายจากข้อมูลของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายคลึงกัน โดยพยากรณ์ความต้องการสินค้าใหม่ตามจำนวนหน่วยขายเป็นจำนวนชิ้น โดยใช้เทคนิคการเรียนรู้ของเครื่อง ซึ่งใช้ข้อมูลยอดขายจากร้านค้าปลีกแห่งหนึ่งที่มีการขายสินค้าผ่านช่องทางออนไลน์ควบคู่ไปกับการขายผ่านหน้าร้าน ผู้วิจัยได้ดำเนินการวิจัยโดยการศึกษาตามขบวนการและขั้นตอนต่าง ๆ ตลอดจนการวัดประสิทธิภาพ เพื่อให้บรรลุจุดประสงค์ของการวิจัยที่ได้กำหนดไว้ ดังนี้

1. ผลลัพธ์ของการสร้างแบบจำลอง Long Short-Term Memory (LSTM)
2. ผลลัพธ์ของการสร้างแบบจำลอง Autoregressive Integrated Moving Average (ARIMA)
3. ผลลัพธ์ของการพยากรณ์แบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA)
4. ผลลัพธ์จากการเปรียบเทียบการพยากรณ์ความต้องการสินค้าระหว่างแบบจำลอง Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA) และ Moving Averages (MA)

4.1 ผลลัพธ์ของการสร้างแบบจำลอง Long Short-Term Memory (LSTM)

จากการสร้างแบบจำลอง LSTM ในการพยากรณ์ความต้องการสินค้าใหม่ จากข้อมูลของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายคลึงกัน โดยใช้ข้อมูลยอดขายรายเดือนของสินค้าปัจจุบันเป็นข้อมูลฝึกสอนมีการกำหนด Time Step เป็น 3 เดือน เพื่อให้โมเดลมองข้อมูลเป็นลำดับเวลาที่ยาวขึ้นและช่วยให้โมเดลสามารถจับความสัมพันธ์ที่ซับซ้อนและลำดับเหตุการณ์ได้ดีขึ้น และกำหนดข้อมูลยอดขายของสินค้าปัจจุบันที่ต้องการพยากรณ์ในอีก 3 เดือนข้างหน้า เพื่อใช้เป็นตัวแทนยอดขายของสินค้าใหม่ใน 3 เดือนแรกที่เริ่มวางขาย โดยใช้ค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) ในการวัดประสิทธิภาพแบบจำลอง ซึ่งค่าพยากรณ์ยอดขายใน 3 เดือนแรกใช้สำหรับประกอบการตัดสินใจในการพิจารณาเลือกสินค้าใหม่เข้ามาวางขาย และในส่วนของค่าพยากรณ์ใน 1 เดือนแรกใช้สำหรับการสั่งซื้อสินค้าใหม่ครั้งแรกมาสต๊อกก่อนเริ่มวางขาย จึงทำการวัดประสิทธิภาพแบบจำลองเป็น 2 ส่วนด้วยกัน คือ ประสิทธิภาพในการพยากรณ์ยอดขายใน 3 เดือนแรก และ 1 เดือนแรก สรุปผลการทดลองได้ดังนี้

พบว่าสินค้าใหม่ครีมย้อมผม N2 ทีใช้ข้อมูลยอดขายของสินค้าปัจจุบันครีมย้อมผม C2 เป็นข้อมูลฝึกสอน เพื่อพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรกมีค่า RMSE และ MAE น้อยที่สุด ซึ่งค่า RMSE และ MAE ที่ต่ำ แสดงว่า Predictions ยิ่งใกล้เคียงกับ Actual Values มากขึ้น รองลงมา คือ สินค้าดินสอเขียนคิ้ว N4 ที่ใช้ข้อมูลยอดขายของสินค้าปัจจุบันดินสอเขียนคิ้ว C4 เป็นข้อมูลฝึกสอน, ครีมย้อมผม N1 ที่ใช้ข้อมูลยอดขายของสินค้าปัจจุบันครีมย้อมผม C1 เป็นข้อมูลฝึกสอน และ สบู่วิตามิน N3 ที่ใช้ข้อมูลยอดขายของสินค้าปัจจุบันสบู่วิตามิน C3 เป็นข้อมูลฝึกสอนตามลำดับ

เมื่อพิจารณาในส่วนของค่าพยากรณ์ยอดขายในเดือนแรก พบว่า ครีมย้อมผม N1 และครีมย้อมผม N2 มีค่า RMSE และ MAE น้อยกว่าค่า RMSE และ MAE ในการพยากรณ์ยอดขาย 3 เดือนแรก แสดงว่ามีประสิทธิภาพในการพยากรณ์เดือนแรกได้แม่นยำมากกว่า ส่วนสินค้าสบู่วิตามิน N3 และดินสอเขียนคิ้ว N4 นั้นมีค่า RMSE และ MAE มากกว่าค่า RMSE และ MAE ในการพยากรณ์ยอดขาย 3 เดือนแรก แสดงว่ามีประสิทธิภาพในการพยากรณ์เดือนแรกได้แม่นยำน้อยกว่า ดังตารางที่ 6

ตาราง 6 แสดงผลการพยากรณ์ความต้องการสินค้าของสินค้าใหม่ โดยใช้แบบจำลอง LSTM

Products	Forecast	RMSE	MAE
ครีมย้อมผม N1	1 เดือนแรก	0.22	0.22
	3 เดือนแรก	3.93	3.25
ครีมย้อมผม N2	1 เดือนแรก	0.83	0.83
	3 เดือนแรก	1.37	1.06
สบู่วิตามิน N3	1 เดือนแรก	31.36	31.36
	3 เดือนแรก	21.25	18.83
ดินสอเขียนคิ้ว N4	1 เดือนแรก	2.66	2.66
	3 เดือนแรก	1.99	1.79

4.2 ผลลัพธ์ของการสร้างแบบจำลอง Autoregressive Integrated Moving Average (ARIMA)

จากการสร้างแบบจำลอง ARIMA ในการพยากรณ์ความต้องการสินค้าใหม่ จากข้อมูลของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายคลึงกัน โดยใช้ข้อมูลยอดขายของสินค้าปัจจุบันเป็นข้อมูลฝึกสอน เพื่อพยากรณ์ความต้องการสินค้าในอีก 3 เดือนข้างหน้า และใช้เป็นตัวแทนยอดขายของสินค้าใหม่ใน 3 เดือนแรก โดยใช้ค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) ในการวัดประสิทธิภาพแบบจำลอง ซึ่งค่าพยากรณ์ยอดขายใน 3 เดือนแรกใช้สำหรับประกอบการตัดสินใจในการพิจารณาเลือกสินค้าใหม่เข้ามาวางขาย และในส่วนของค่าพยากรณ์ใน 1 เดือนแรกใช้สำหรับการสั่งซื้อสินค้าใหม่ครั้งแรกมาสต๊อกก่อนเริ่มวางขาย จึงทำการวัดประสิทธิภาพแบบจำลองเป็น 2 ส่วนด้วยกัน คือ ประสิทธิภาพในการพยากรณ์ยอดขายใน 3 เดือนแรก และ 1 เดือนแรก สรุปผลการทดลองได้ดังนี้

พบว่าสินค้าใหม่ครีมย้อมผม N2 ที่ใช้ข้อมูลยอดขายของสินค้าปัจจุบันครีมย้อมผม C2 เป็นข้อมูลฝึกสอน เพื่อพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรกมีค่า RMSE และ MAE น้อยที่สุด ซึ่งค่า RMSE และ MAE ที่ต่ำ แสดงว่า Predictions ยิ่งใกล้เคียงกับ Actual Values มากขึ้น รองลงมา คือ ครีมย้อมผม N1 ที่ใช้ข้อมูลยอดขายของสินค้าปัจจุบันครีมย้อมผม C1 เป็นข้อมูลฝึกสอน, ดินสอเขียนคิ้ว N4 ที่ใช้ข้อมูลยอดขายของสินค้าปัจจุบันดินสอเขียนคิ้ว C4 เป็นข้อมูลฝึกสอน และ สบู่วิตามิน N3 ที่ใช้ข้อมูลยอดขายของสินค้าปัจจุบันสบู่วิตามิน C3 เป็นข้อมูลฝึกสอน ตามลำดับ

เมื่อพิจารณาในส่วนของค่าพยากรณ์ยอดขายในเดือนแรก พบว่า ครีมย้อมผม N1, ครีมย้อมผม N2 และดินสอเขียนคิ้ว N4 มีค่า RMSE และ MAE น้อยกว่าค่า RMSE และ MAE ในการพยากรณ์ยอดขาย 3 เดือนแรก แสดงว่ามีประสิทธิภาพในการพยากรณ์เดือนแรกได้แม่นยำมากกว่า ส่วนสินค้าสบู่วิตามิน N3 นั้นมีค่า RMSE และ MAE มากกว่าค่า RMSE และ MAE ในการพยากรณ์ยอดขาย 3 เดือนแรก แสดงว่ามีประสิทธิภาพในการพยากรณ์เดือนแรกได้แม่นยำน้อยกว่า ดังตารางที่ 7

ตาราง 7 แสดงผลการพยากรณ์ความต้องการสินค้า 3 เดือนแรกของสินค้าใหม่ โดยใช้แบบจำลอง ARIMA

Products	Forecast	RMSE	MAE
ครีมย้อมผม N1	1 เดือนแรก	0.15	0.15
	3 เดือนแรก	4.86	3.88
ครีมย้อมผม N2	1 เดือนแรก	1.11	1.11
	3 เดือนแรก	1.39	1.13
สบู่วิตามิน N3	1 เดือนแรก	12.60	12.60
	3 เดือนแรก	7.60	5.76
ดินสอเขียนคิ้ว N4	1 เดือนแรก	0.90	0.90
	3 เดือนแรก	5.36	4.65

4.3 ผลลัพธ์ของการพยากรณ์แบบค่าเฉลี่ยเคลื่อนที่ (Moving Averages: MA)

จากการพยากรณ์แบบค่าเฉลี่ยเคลื่อนที่ (MA) ในการพยากรณ์ความต้องการสินค้าใหม่จากข้อมูลของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายคลึงกัน โดยใช้ข้อมูลยอดขายในอดีตของสินค้าปัจจุบันมาคำนวณหาค่าเฉลี่ยของยอดขาย เพื่อพยากรณ์ความต้องการสินค้าในอีก 3 เดือนข้างหน้า และใช้เป็นตัวแทนยอดขายของสินค้าใหม่ใน 3 เดือนแรก โดยใช้ค่า Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) ในการวัดประสิทธิภาพแบบจำลอง ซึ่งค่าพยากรณ์ยอดขายใน 3 เดือนแรกใช้สำหรับประกอบการตัดสินใจในการพิจารณาเลือกสินค้าใหม่เข้ามาวางขาย และในส่วนของค่าพยากรณ์ใน 1 เดือนแรกใช้สำหรับการสั่งซื้อสินค้าใหม่ครั้งแรกมาสต๊อกก่อนเริ่มวางขาย จึงทำการวัดประสิทธิภาพแบบจำลองเป็น 2 ส่วนด้วยกัน คือ ประสิทธิภาพในการพยากรณ์ยอดขายใน 3 เดือนแรก และ 1 เดือนแรก สรุปผลการทดลองได้ดังนี้

พบว่าสินค้าใหม่ครีมย้อมผม N2 ที่ใช้ข้อมูลยอดขายเฉลี่ยของสินค้าปัจจุบันครีมย้อมผม C2 ในอดีต ตั้งแต่ปี ค.ศ. 2021-2022 เพื่อพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรก มีค่า RMSE และ MAE น้อยที่สุด ซึ่งค่า RMSE และ MAE ที่ต่ำ แสดงว่า Predictions ยิ่งใกล้เคียงกับ Actual Values มากขึ้น รองลงมา คือ สินค้าดินสอเขียนคิ้ว N4 ที่ใช้ข้อมูลยอดขายเฉลี่ยของสินค้าปัจจุบันดินสอเขียนคิ้ว C4 มาพยากรณ์, ครีมย้อมผม N1 ที่ใช้ข้อมูลยอดขายเฉลี่ยของสินค้า

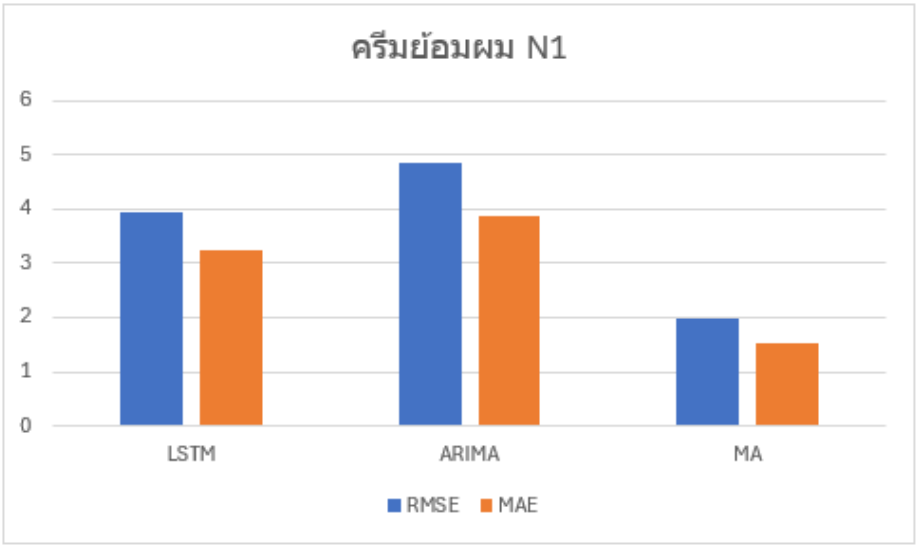
ปัจจุบันครีมย้อมผม C1 มาพยากรณ์ และสบู่วิตามิน N3 ที่ใช้ข้อมูลยอดขายเฉลี่ยของสินค้าปัจจุบันสบู่วิตามิน C3 มาพยากรณ์ ตามลำดับ

เมื่อพิจารณาในส่วนของค่าพยากรณ์ยอดขายในเดือนแรก พบว่า ครีมย้อมผม N1, ครีมย้อมผม N2, สบู่วิตามิน N3 และดินสอเขียนคิ้ว N4 มีค่า RMSE และ MAE น้อยกว่าค่า RMSE และ MAE ในการพยากรณ์ยอดขาย 3 เดือนแรก แสดงว่ามีประสิทธิภาพในการพยากรณ์เดือนแรกได้แม่นยำมากกว่า ดังตารางที่ 8

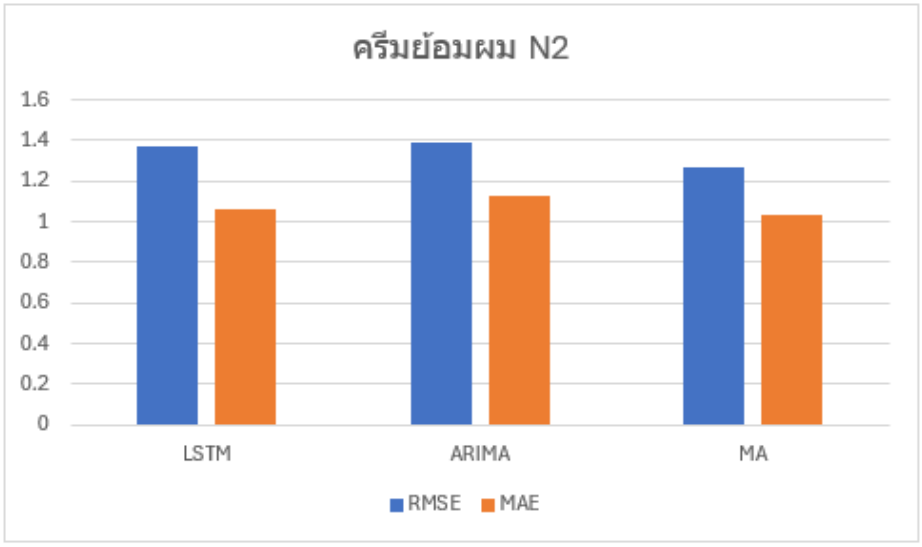
ตาราง 8 แสดงผลการพยากรณ์ความต้องการสินค้า 3 เดือนแรกของสินค้าใหม่ โดยวิธี MA

Products	Forecast	RMSE	MAE
ครีมย้อมผม N1	1 เดือนแรก	0.21	0.21
	3 เดือนแรก	1.98	1.54
ครีมย้อมผม N2	1 เดือนแรก	1.09	1.09
	3 เดือนแรก	1.27	1.03
สบู่วิตามิน N3	1 เดือนแรก	14.25	14.25
	3 เดือนแรก	17.42	17.25
ดินสอเขียนคิ้ว N4	1 เดือนแรก	0.83	0.83
	3 เดือนแรก	1.84	1.72

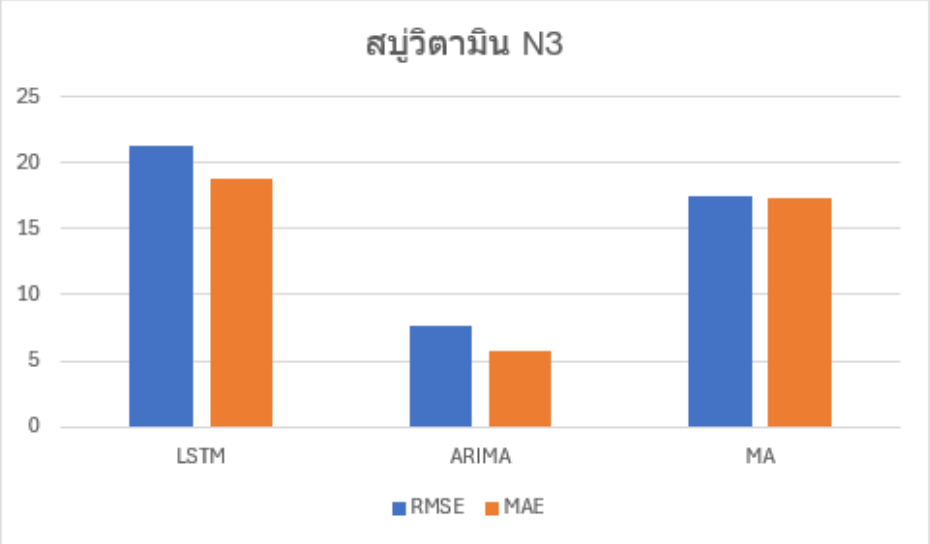
จากนั้นทำการเปรียบเทียบค่า RMSE และ MAE ของแบบจำลอง LSTM, ARIMA และวิธี Moving Averages ในการพยากรณ์ยอดขาย 3 เดือนแรกด้วยกราฟ เพื่อดูค่า Error ของแต่ละแบบจำลอง ดังภาพประกอบที่ 28-31



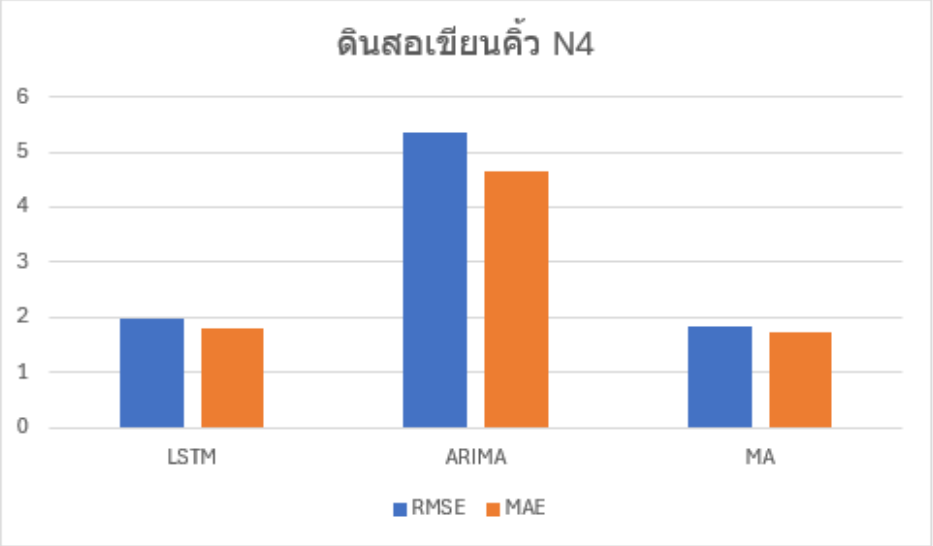
ภาพประกอบ 28 กราฟแสดงค่า RMSE และ MAE เปรียบเทียบระหว่าง LSTM, ARIMA และ MA ของสินค้าครีมย้อมผม N1



ภาพประกอบ 29 กราฟแสดงค่า RMSE และ MAE เปรียบเทียบระหว่าง LSTM, ARIMA และ MA ของสินค้าครีมย้อมผม N2



ภาพประกอบ 30 กราฟแสดงค่า RMSE และ MAE เปรียบเทียบระหว่าง LSTM, ARIMA และ MA ของสินค้าสบู่วิตามิน N3

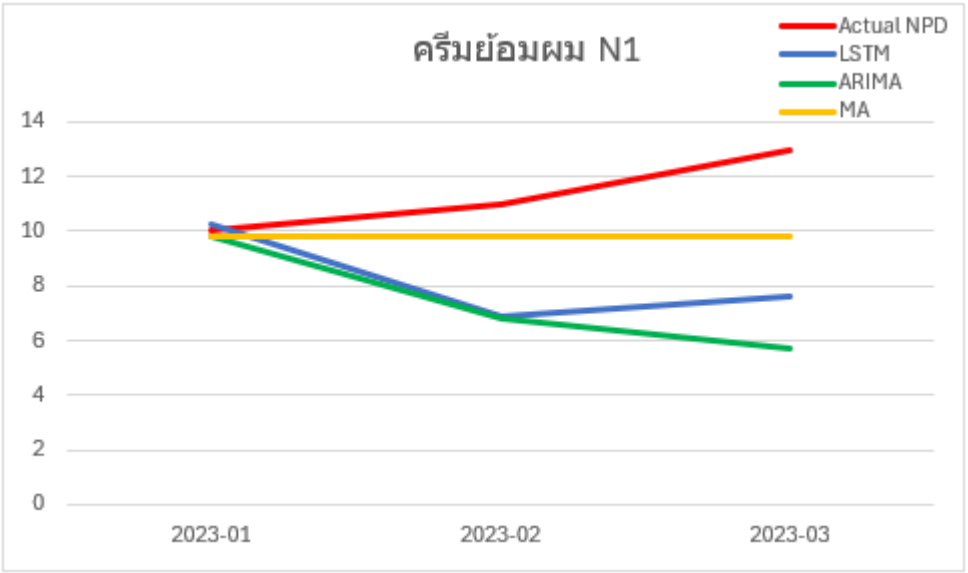


ภาพประกอบ 31 กราฟแสดงค่า RMSE และ MAE เปรียบเทียบระหว่าง LSTM, ARIMA และ MA ของสินค้าดินสอเขียนคิ้ว N4

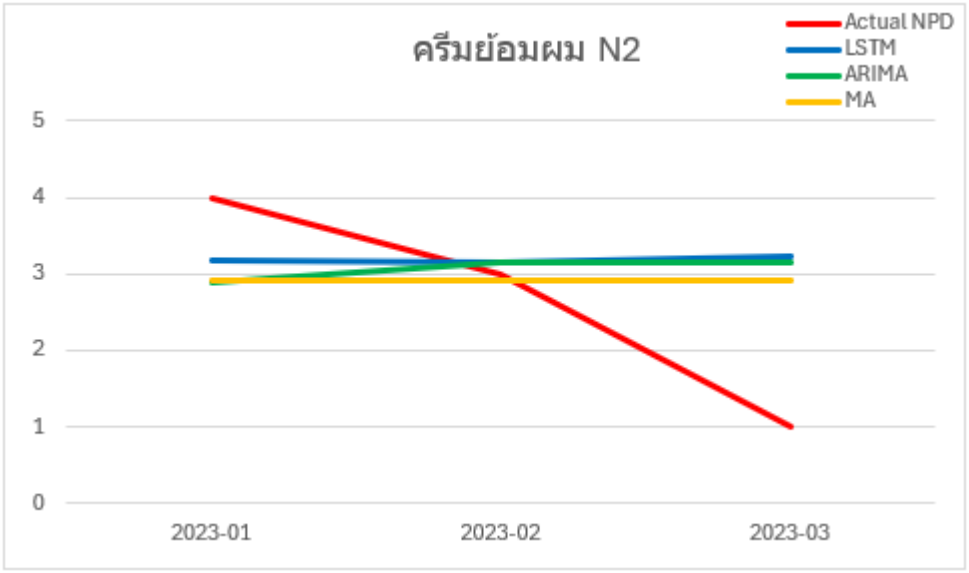
จากกราฟจะเห็นได้ว่าการพยากรณ์ด้วยวิธี MA นั้น มีค่า RMSE และ MAE น้อยที่สุด จำนวน 3 สินค้า คือ สินค้าครีมย้อมผม N1, สินค้าครีมย้อมผม N2 และดินสอเขียนคิ้ว N4 ส่วนแบบจำลอง ARIMA มีค่า RMSE และ MAE น้อยที่สุดในสินค้าสบู่วิตามิน N3 ซึ่งหากค่า RMSE และ MAE ยิ่งต่ำ แสดงว่าค่าที่พยากรณ์ได้ใกล้เคียงกับค่าจริงมากขึ้น

4.4 ผลลัพธ์จากการเปรียบเทียบการพยากรณ์ความต้องการสินค้าระหว่างแบบจำลอง Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA) และ Moving Averages (MA)

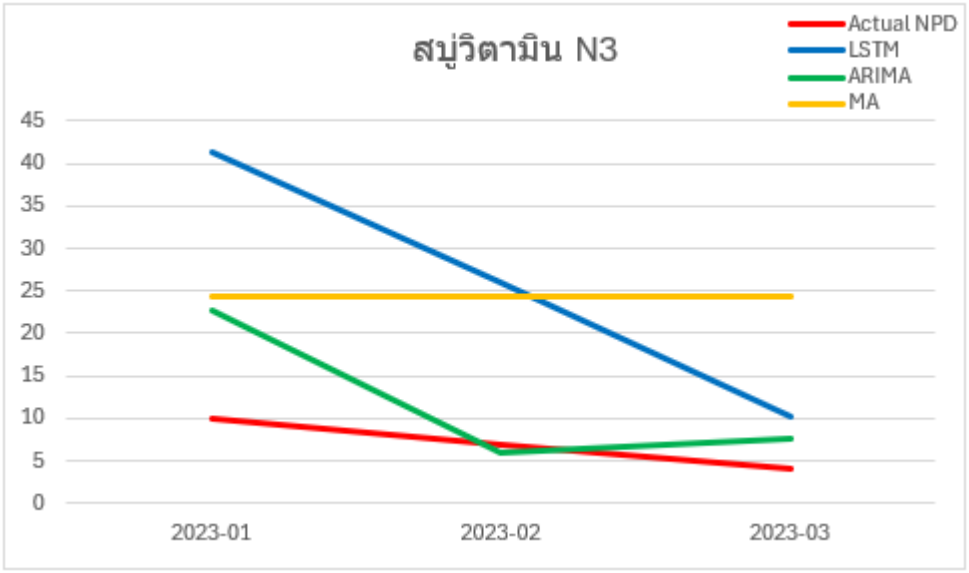
ผลลัพธ์ที่ได้จากการทดลองด้วยแบบจำลอง LSTM, ARIMA และ MA ของสินค้าใหม่ทั้ง 4 SKU นำค่าพยากรณ์ที่ได้มาเปรียบเทียบกับยอดขายจริง ด้วยการพล็อตกราฟ ดังภาพประกอบที่ 32-35



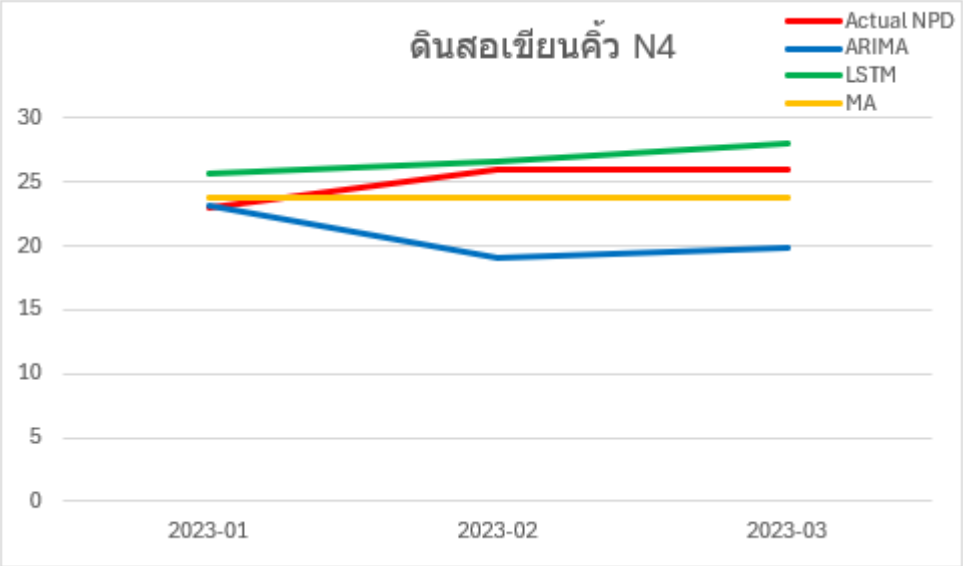
ภาพประกอบ 32 กราฟแสดงผลการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าครีมย้อมผม N1 ใน 3 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA



ภาพประกอบ 33 กราฟแสดงผลการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าครีมย้อมผม N2 ใน 3 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA



ภาพประกอบ 34 กราฟแสดงผลการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าสบู่วิตามิน N3 ใน 3 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA



ภาพประกอบ 35 กราฟแสดงผลการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าดินสอเขียนคิ้ว N4 ใน 3 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA

สรุปผลการพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรกของสินค้าทั้ง 4 SKU ได้ดังนี้

สินค้าที่ 1 : สินค้าครีมย้อมผม N1 พบว่า ผลการพยากรณ์ความต้องการสินค้าใน 3 เดือนแรกด้วยวิธี MA นั้นมีประสิทธิภาพในการพยากรณ์ที่แม่นยำมากว่าแบบจำลอง LSTM และ ARIMA เมื่อพิจารณาจากค่า RMSE และ MSE

เมื่อพิจารณาที่ยอดขายจริงของสินค้าใหม่เฉพาะในเดือนแรกเทียบกับค่าที่พยากรณ์ทั้ง 3 แบบจำลองนั้น แบบจำลอง ARIMA พยากรณ์ได้แม่นยำและใกล้เคียงกับยอดขายจริงอย่างมาก อย่างไรก็ตามแม้ว่าแบบจำลอง ARIMA มีความแม่นยำมากกว่าในเดือนแรก แต่แบบจำลอง LSTM และ MA ก็พยากรณ์ยอดขายเดือนแรกได้ใกล้เคียงกับยอดขายจริงเช่นกัน ซึ่งค่าต่างจากวิธี MA เพียงเล็กน้อยเท่านั้น ซึ่งในการนำค่าพยากรณ์ไปใช้งานจริง ต้องทำการปัดเลขเป็นจำนวนเต็ม เนื่องจากในการสั่งซื้อสินค้านั้นต้องสั่งซื้อเป็นเลขจำนวนเต็มเท่านั้น

สินค้าที่ 2 : สินค้าครีมย้อมผม N2 พบว่า ผลการพยากรณ์ความต้องการสินค้าใน 3 เดือนแรกด้วยวิธี MA นั้นมีประสิทธิภาพในการพยากรณ์ที่แม่นยำมากว่า LSTM และ ARIMA เมื่อพิจารณาจากค่า RMSE และ MSE

เมื่อพิจารณาที่ยอดขายจริงของสินค้าใหม่เฉพาะในเดือนแรกเทียบกับค่าที่พยากรณ์ทั้ง 3 แบบจำลองนั้น วิธี MA พยากรณ์ได้แม่นยำและใกล้เคียงกับยอดขายจริงมากกว่า อย่างไรก็ตาม แม้ว่าวิธี MA มีความแม่นยำมากกว่าในเดือนแรก แต่แบบจำลอง LSTM และ ARIMA ก็พยากรณ์ยอดขายเดือนแรกได้ใกล้เคียงกับยอดขายจริงเช่นกัน ซึ่งค่าต่างจากวิธี MA เพียงเล็กน้อยเท่านั้น ซึ่งในการนำค่าพยากรณ์ไปใช้งานจริง ต้องทำการปัดเลขเป็นจำนวนเต็ม เนื่องจากในการสั่งซื้อสินค้านั้นต้องสั่งซื้อเป็นเลขจำนวนเต็มเท่านั้น

สินค้าที่ 3 : สินค้าสบู่วิตามิน N3 พบว่า ผลการพยากรณ์ความต้องการสินค้าใน 3 เดือนแรกด้วยแบบจำลอง ARIMA นั้นมีประสิทธิภาพในการพยากรณ์ที่แม่นยำมากว่า LSTM และ MA เมื่อพิจารณาจากค่า RMSE และ MSE

เมื่อพิจารณาที่ยอดขายจริงของสินค้าใหม่เฉพาะในเดือนแรกเทียบกับค่าที่พยากรณ์ทั้ง 3 แบบจำลองนั้น แบบจำลอง ARIMA พยากรณ์ได้ใกล้เคียงกับยอดขายจริงมากกว่า LSTM และ MA แต่เมื่อเทียบกับยอดขายจริงในเดือนที่ 2-3 นั้น ARIMA พยากรณ์ได้แม่นยำมากกว่าเดือนแรก

สินค้าที่ 4 : สินค้าดินสอเขียนคิ้ว N4 พบว่า ผลการพยากรณ์ความต้องการสินค้าใน 3 เดือนแรกด้วยวิธี MA นั้นมีประสิทธิภาพในการพยากรณ์ที่แม่นยำมากว่า LSTM และ ARIMA เมื่อพิจารณาจากค่า RMSE และ MSE

เมื่อพิจารณาที่ยอดขายจริงของสินค้าใหม่เฉพาะในเดือนแรกเทียบกับค่าที่พยากรณ์ทั้ง 3 แบบจำลองนั้น วิธี MA พยากรณ์ได้แม่นยำและใกล้เคียงกับยอดขายจริงอย่างมาก แต่แบบจำลอง ARIMA พยากรณ์ยอดขายเดือนแรกได้ใกล้เคียงกับยอดขายจริงเช่นกัน ซึ่งค่าต่างจากวิธี MA เพียงเล็กน้อยเท่านั้น

จากนั้นทำการเปรียบเทียบผลการพยากรณ์ความต้องการสินค้าใหม่ 1 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA เทียบกับจำนวนสั่งซื้อที่ทำการปัดค่าพยากรณ์เป็นจำนวนเต็ม ดังตารางที่ 9

ตาราง 9 แสดงผลการพยากรณ์ความต้องการสินค้าใหม่ 1 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA เทียบกับค่าจริงและจำนวนสั่งซื้อที่ทำการปัดค่าพยากรณ์เป็นจำนวนเต็ม

Products	Actual	Forecast			**จำนวนสั่งซื้อ**		
		LSTM	ARIMA	MA	LSTM	ARIMA	MA
ครีมย้อมผม N1	10	10.22	9.85	9.79	10	10	10
ครีมย้อมผม N2	4	3.17	2.89	2.91	3	3	3
สบู่วิตามิน N3	10	41.36	22.60	24.25	41	23	24
ดินสอเขียนคิ้ว N4	23	25.66	23.09	23.83	26	23	24

สรุปผลการพยากรณ์ความต้องการสินค้าใหม่ใน 1 เดือนแรกเทียบกับจำนวนสั่งซื้อที่ทำการปัดค่าพยากรณ์เป็นจำนวนเต็มของสินค้าทั้ง 4 SKU ได้ดังนี้

สินค้าครีมย้อมผม N1 เมื่อนำค่าพยากรณ์เดือนแรกทั้ง 3 วิธี มาทำการปัดเลขเป็นจำนวนเต็ม เพื่อใช้ในการสั่งซื้อสินค้าใหม่ครั้งแรกมาสต๊อกก่อนสินค้าใหม่เริ่มวางขาย พบว่าได้จำนวนชิ้นเท่ากัน คือ 10 ชิ้น ส่วนสินค้าครีมย้อมผม N2 ทั้ง 3 วิธีก็ได้จำนวนชิ้นเท่ากัน คือ 3 ชิ้น เมื่อพิจารณาที่สินค้าดินสอเขียนคิ้ว N4 พบว่าแม้แต่ละวิธีนั้นพยากรณ์ได้ค่าที่ต่างกันแต่ค่าพยากรณ์ที่ได้ค่อนข้างใกล้เคียงกันทั้ง 3 วิธี และใกล้เคียงกับยอดขายจริงด้วย

สรุปภาพรวมการพยากรณ์ความต้องการสินค้าใหม่ สามารถสรุปได้ว่าการพยากรณ์ด้วยวิธี MA มีประสิทธิภาพในการพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรก มากกว่า LSTM และ ARIMA โดยพยากรณ์ได้แม่นยำมากกว่า 3 SKU จากทั้งหมด 4 SKU คือสินค้าครีมย้อมผม N1, สินค้าครีมย้อมผม N2 และ และสินค้าดินสอเขียนคิ้ว N4 ในขณะที่วิธี ARIMA พยากรณ์ได้แม่นยำที่สุด 1 SKU คือ สินค้าสบู่วิตามิน N3 ส่วนวิธี LSTM นั้นไม่มีสินค้าที่พยากรณ์ได้มีประสิทธิภาพดีที่สุด คาดว่าเนื่องมาจากข้อมูลมีปริมาณน้อยและไม่มีความซับซ้อน อีกทั้งมี 2 SKU ที่ค่าพยากรณ์ความต้องการสินค้าในเดือนแรก เมื่อนำไปใช้งานจริงต้องทำการปัดเลขเป็นจำนวนเต็ม เนื่องจากในการสั่งซื้อสินค้านั้นต้องซื้อเป็นเลขจำนวนเต็มเท่านั้น ซึ่งพบว่าทั้ง 3 วิธี ได้จำนวนชิ้นเท่ากัน คือ สินค้าสินค้าครีมย้อมผม N1 และสินค้าครีมย้อมผม N2

บทที่ 5
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

ในการวิจัยเรื่องการพยากรณ์ความต้องการสินค้าใหม่ของลูกค้าในธุกิจค้าปลีกออนไลน์ โดยใช้การเรียนรู้ของเครื่อง ผู้วิจัยได้วัดประสิทธิภาพของแบบจำลองแต่ละอัลกอริทึม เพื่อเปรียบเทียบและสรุปผล โดยสามารถแบ่งหัวข้อในการสรุปผลได้ดังต่อไปนี้

1. สรุปผลการวิจัย
2. อภิปรายผลการวิจัย
3. ข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การดำเนินธุรกิจให้มีความหลากหลายของสินค้าและบริการ เพื่อเพิ่มยอดขาย ตอบสนองความต้องการของลูกค้า และเพื่อรักษาฐานลูกค้าหรือขยายฐานลูกค้าก็ตาม อาจทำให้ผู้ประกอบการธุรกิจมีปัญหาในการบริหารจัดการสินค้าที่ยุ่งยากมากขึ้น โดยเฉพาะกับสินค้าใหม่ที่เตรียมนำเข้ามาวางขาย เนื่องจากยังไม่มีข้อมูลปริมาณการขายในอดีตมาช่วยในการพยากรณ์ความต้องการของลูกค้าได้ ซึ่งอาจส่งผลทำให้เกิดปัญหาตามมา เช่น ปัญหาการสั่งซื้อสินค้ามากจนเกินไปเกิดความเสี่ยงในการขายสินค้าไม่ทันก่อนที่จะหมดอายุ ปัญหาพื้นที่จัดเก็บ ค่าใช้จ่ายในการเก็บรักษา และการหมุนเวียนเงินทุน เป็นต้น ดังนั้นหากมีวิธีการที่สามารถช่วยพยากรณ์ความต้องการสินค้าใหม่ได้ และนำข้อมูลที่ได้จากการพยากรณ์มาเป็นข้อมูลพื้นฐานในการวางแผนการสั่งซื้อสินค้าใหม่ได้มีประสิทธิภาพมากยิ่งขึ้น

ในงานวิจัยนี้ ผู้วิจัยทำการศึกษาและได้วิเคราะห์เปรียบเทียบประสิทธิภาพแบบจำลอง Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA) และ Moving Averages (MA) โดยเลือกศึกษาชุดข้อมูลธุรกรรมการขายของร้านค้าปลีกแห่งหนึ่งที่มีการขายสินค้าผ่านช่องทางออนไลน์ควบคู่กับการขายผ่านหน้าร้าน และนำชุดข้อมูลธุรกรรมการขายดังกล่าวใช้ในการสร้างแบบจำลองพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรกที่เริ่มวางขาย จากข้อมูลการขายของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายคลึงกัน โดยผลการวัดประสิทธิภาพแบบจำลองแต่ละอัลกอริทึมเป็น 2 ส่วนด้วยกัน คือ ประสิทธิภาพในการพยากรณ์ยอดขายใน 3 เดือนแรก และ 1 เดือนแรก ได้ผลดังตารางที่ 10

ตาราง 10 แสดงผลการพยากรณ์ความต้องการสินค้าใหม่ของแต่ละสินค้าระหว่างแบบจำลอง LSTM, ARIMA และ MA

Products	Forecast	LSTM		ARIMA		MA	
		RMSE	MAE	RMSE	MAE	RMSE	MAE
ครีมย้อมผม N1	1 เดือนแรก	0.22	0.22	0.15	0.15	0.21	0.21
	3 เดือนแรก	3.93	3.25	4.86	3.88	1.98	1.54
ครีมย้อมผม N2	1 เดือนแรก	0.83	0.83	1.11	1.11	1.09	1.09
	3 เดือนแรก	1.37	1.06	1.39	1.13	1.27	1.03
สบู่วิตามิน N3	1 เดือนแรก	31.36	31.36	12.60	12.60	14.25	14.25
	3 เดือนแรก	21.25	18.83	7.60	5.76	17.42	17.25
ดินสอเขียนคิ้ว N4	1 เดือนแรก	2.66	2.66	0.90	0.90	0.83	0.83
	3 เดือนแรก	1.99	1.79	5.36	4.65	1.84	1.72

จากผลการทดลองสรุปได้ว่า แบบจำลองที่มีประสิทธิภาพดีที่สุดในการพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรก เพื่อเป็นข้อมูลในการช่วยตัดสินใจว่าจะทำการสั่งซื้อสินค้าใหม่ดังกล่าวเข้ามาวางจำหน่ายในร้านหรือไม่ สามารถสรุปแยกตามแบบจำลองได้ดังนี้

1. การพยากรณ์ด้วยวิธี MA พยากรณ์ได้แม่นยำที่สุด 3 SKU คือ 1) สินค้าครีมย้อมผม N1 มีค่า RMSE เท่ากับ 1.98 และ MAE เท่ากับ 1.54 2) สินค้าครีมย้อมผม N2 มีค่า RMSE เท่ากับ 1.27 และ MAE เท่ากับ 1.03 และ 3) สินค้าดินสอเขียนคิ้ว N4 มีค่า RMSE เท่ากับ 1.84 และ MAE เท่ากับ 1.72 ซึ่งยอดขายของแต่ละสินค้าอยู่ในช่วงใกล้เคียงกันในแต่ละเดือน

2. การพยากรณ์ด้วยวิธี ARIMA พยากรณ์ได้แม่นยำที่สุด 1 SKU คือ สินค้าสบู่วิตามิน N3 มีค่า RMSE เท่ากับ 7.60 และ MAE เท่ากับ 5.76 ซึ่งยอดขาย 3 เดือนแรกมีแนวโน้มลดลงและยอดขายแต่ละเดือนอยู่ในช่วงใกล้เคียงกันน้อยกว่าสินค้าครีมย้อมผม N1, สินค้าครีมย้อมผม N2 และสินค้าดินสอเขียนคิ้ว N4 โดยสินค้าสินค้าสบู่วิตามิน N3 มีแนวโน้มยอดขายที่ลดลงค่อนข้างมากจากเดือน 1 ไป 2 ประมาณ 30% และ เดือน 2 ไป 3 ประมาณ 43% ถึงแม้ว่าพยากรณ์ในระยะเวลาที่สั้นกว่าจะมีความแม่นยำกว่าแต่โอกาสที่แบบจำลองจะพยากรณ์ในระยะเวลาที่ยาวกว่าได้แม่ยำมากกว่าการพยากรณ์ระยะสั้นก็มีโอกาสที่จะเกิดขึ้นได้ แม้โอกาสที่เกิดขึ้นนั้นน้อยกว่า

3. การพยากรณ์ด้วยวิธี LSTM นั้นไม่มีสินค้าที่พยากรณ์ได้มีประสิทธิภาพดีที่สุด คาดว่าเนื่องมาจากข้อมูลมีปริมาณน้อยและไม่มีความซับซ้อน แต่ถ้าหากพิจารณาดูที่ค่า RMSE และ MAE จะเห็นว่ามีค่า Error ไม่ได้แต่ต่างจากวิธี ARIMA และ MA มากนัก

เนื่องจากในการนำไปใช้งานจริงในการสั่งซื้อใหม่ครั้งแรกมาสต๊อก เพื่อเตรียมสินค้าก่อนที่สินค้าใหม่เริ่มวางขายนั้นใช้เฉพาะค่าพยากรณ์เดือนแรก จึงทำการวัดประสิทธิภาพการพยากรณ์ยอดขายเดือนแรกของแต่ละวิธี ซึ่งค่า RMSE และ MAE ของเดือนแรกมีค่าเท่ากัน เนื่องจากคำนวณจากยอดพยากรณ์ 1 เดือนเทียบกับยอดขายจริง 1 เดือน โดยแบบจำลอง LSTM มีประสิทธิภาพในการพยากรณ์ยอดขาย 1 เดือนแรกดีที่สุดในสินค้าครีมย้อมผม N2 มีค่า RMSE และ MAE เท่ากับ 0.83 แบบจำลอง ARIMA มีประสิทธิภาพในการพยากรณ์ยอดขาย 1 เดือนแรกดีที่สุด 2 สินค้าด้วยกัน คือ 1.สินค้าครีมย้อมผม N1 มีค่า RMSE และ MAE เท่ากับ 0.15 และ 2.สินค้าสบู่วิตามิน N3 มีค่า RMSE และ MAE เท่ากับ 12.60 ส่วนวิธี MA มีประสิทธิภาพในการพยากรณ์ยอดขาย 1 เดือนแรกดีที่สุดในสินค้าดินสอเขียนคิ้ว N4 ค่า RMSE และ MAE เท่ากับ 0.83 และเนื่องจากในการสั่งซื้อสินค้านั้นต้องซื้อเป็นเลขจำนวนเต็มเท่านั้น หากทำการปัดเลขค่าพยากรณ์เดือนแรกเป็นจำนวนเต็มพบว่าทั้ง 3 วิธี ได้จำนวนชิ้นเท่ากัน คือ สินค้าสินค้าครีมย้อมผม N1 และสินค้าครีมย้อมผม N2 ในการนำค่าพยากรณ์ยอดขาย 1 เดือนแรกไปใช้ในการสั่งซื้อสินค้าดังตารางที่ 11

ตาราง 11 แสดงผลการพยากรณ์ความต้องการสินค้าใหม่ 1 เดือนแรก ระหว่างแบบจำลอง LSTM, ARIMA และ MA เทียบกับจำนวนสั่งซื้อที่ทำการปัดค่าพยากรณ์เป็นจำนวนเต็ม

Products	Forecast			**จำนวนสั่งซื้อ**		
	LSTM	ARIMA	MA	LSTM	ARIMA	MA
ครีมย้อมผม N1	10.22	9.85	9.79	10	10	10
ครีมย้อมผม N2	3.17	2.89	2.91	3	3	3
สบู่วิตามิน N3	41.36	22.60	24.25	41	23	24
ดินสอเขียนคิ้ว N4	25.66	23.09	23.83	26	23	24

5.2 อภิปรายผลการวิจัย

ผู้วิจัยได้ทำการศึกษาและวิเคราะห์เปรียบเทียบประสิทธิภาพแบบจำลอง LSTM, ARIMA และ MA โดยเลือกใช้ค่า RMSE และ MAE เป็นตัววัดประสิทธิภาพของแบบจำลองทั้ง 3 อัลกอริทึมภายในสินค้าเดียวกัน ซึ่งค่า RMSE ที่ได้จากการพยากรณ์มีค่ามากกว่าค่า MAE เนื่องมาจาก RMSE ให้ความสำคัญกับ Outlier มากกว่า MAE สาเหตุมากจาก RMSE ยกกำลังสองของ Error ก่อนหาค่าเฉลี่ย ทำให้ค่า Error ที่มีขนาดใหญ่ (Outlier) ส่งผลต่อค่า RMSE มากกว่า MAE ส่วน MAE หาค่าเฉลี่ยของ Absolute Error โดยไม่ต้องยกกำลังสอง ทำให้ค่า Error ทุกตัวมีผลต่อค่า MAE เท่ากัน ผลที่ตามมา คือ RMSE มีความไวต่อ Outlier มากกว่า MAE หมายความว่า Outlier เพียงตัวเดียวสามารถทำให้ค่า RMSE สูงขึ้นได้มาก ซึ่ง MAE มีความทนทานต่อ Outlier มากกว่า RMSE หมายความว่า Outlier ไม่ส่งผลต่อค่า MAE มากนัก สรุปได้ว่า RMSE เหมาะกับกรณีที่ต้องการวัดความแม่นยำโดยรวม โดยไม่คำนึงถึง Outlier ส่วน MAE เหมาะกับกรณีที่ต้องการวัดความแม่นยำโดยไม่ให้ Outlier ส่งผล จึงใช้ MAE ร่วมกับ RMSE เพื่อดูว่า Outlier ส่งผลต่อโมเดลมากน้อยแค่ไหน และทำการหาผลต่างระหว่างค่า RMSE และ MAE เพื่อดูค่า Error ของแต่ละแบบจำลองในการพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรก สามารถอธิบายผลการทดลองแยกตามรายสินค้าได้ดังนี้

1. แบบจำลองการพยากรณ์ความต้องการสินค้าครีมย้อมผม N1 โดยวิธี LSTM, ARIMA และ MA ทำการหาผลต่างระหว่างค่า RMSE และ MAE ที่ได้จากการวัดประสิทธิภาพแบบจำลองในการพยากรณ์ความต้องการสินค้าใน 3 เดือนแรก ได้ดังนี้ LSTM เท่ากับ 0.68, ARIMA เท่ากับ 0.98 และ MA เท่ากับ 0.44 ซึ่งเห็นได้ว่า ค่า Error ของแบบจำลอง LSTM, ARIMA และ MA ไม่ได้สูงมากนัก แต่ MA มีค่า Error ที่น้อยที่สุด รองลงมา คือ LSTM และ ARIMA ดังนั้นการพยากรณ์ด้วยวิธี MA จึงมีประสิทธิภาพในการพยากรณ์ได้มีความแม่นยำมากกว่าแบบจำลอง LSTM และ ARIMA

2. แบบจำลองการพยากรณ์ความต้องการสินค้าครีมย้อมผม N2 โดยวิธี LSTM, ARIMA และ MA ทำการหาผลต่างระหว่างค่า RMSE และ MAE ที่ได้จากการวัดประสิทธิภาพแบบจำลองในการพยากรณ์ความต้องการสินค้าใน 3 เดือนแรก ได้ดังนี้ LSTM เท่ากับ 0.31, ARIMA เท่ากับ 0.26 และ MA เท่ากับ 0.24 เห็นได้ว่า ค่า Error ของแบบจำลอง LSTM, ARIMA และ MA ไม่ได้สูงมากนัก แต่ MA มีค่า Error ที่น้อยว่าสุด รองลงมา คือ ARIMA และ LSTM ดังนั้นการพยากรณ์ด้วยวิธี MA จึงมีประสิทธิภาพในการพยากรณ์ได้มีความแม่นยำมากกว่าแบบจำลอง LSTM และ ARIMA

3. แบบจำลองการพยากรณ์ความต้องการสินค้าสบู่วิตามิน N3 โดยวิธี LSTM, ARIMA และ MA ทำการหาผลต่างระหว่างค่า RMSE และ MAE ที่ได้จากการวัดประสิทธิภาพแบบจำลองในการพยากรณ์ความต้องการสินค้าใน 3 เดือนแรก ได้ดังนี้ LSTM เท่ากับ 2.42, ARIMA เท่ากับ 1.84 และ MA เท่ากับ 0.17 เห็นได้ว่า ค่า Error ของแบบจำลอง LSTM, ARIMA และ MA ไม่ได้สูงมากนัก แต่ ARIMA มีค่า Error ที่น้อยสุด รองลงมาคือ MA และ LSTM ดังนั้นแบบจำลอง ARIMA จึงมีประสิทธิภาพในการพยากรณ์ได้มีความแม่นยำมากกว่าแบบจำลอง MA และ LSTM

4. แบบจำลองการพยากรณ์ความต้องการสินค้าดินสอเขียนคิ้ว N4 โดยวิธี LSTM, ARIMA และ MA ทำการหาผลต่างระหว่างค่า RMSE และ MAE ที่ได้จากการวัดประสิทธิภาพแบบจำลองในการพยากรณ์ความต้องการสินค้าใน 3 เดือนแรก ได้ดังนี้ LSTM เท่ากับ 0.20, ARIMA เท่ากับ 0.71 และ MA เท่ากับ 0.12 เห็นได้ว่า ค่า Error ของ MA ไม่ได้สูงมากนัก รองลงมา คือ LSTM และ ARIMA ดังนั้นการพยากรณ์ด้วยวิธี MA จึงมีประสิทธิภาพในการพยากรณ์ได้มีความแม่นยำมากกว่าแบบจำลอง LSTM และ ARIMA

จากผลลัพธ์ที่ได้ของแบบจำลองทั้ง 3 วิธีนั้น ค่า RMSE และ MAE ค่อนข้างต่ำ แสดงว่ามีประสิทธิภาพในการพยากรณ์ได้ค่อนข้างแม่นยำสำหรับชุดข้อมูลนี้ แต่การพยากรณ์ด้วยวิธี MA ให้ผลการพยากรณ์ที่มีความแม่นยำมากกว่า เมื่อเปรียบเทียบกับสินค้าใหม่ทั้งหมด 4 SKU นั้น แม่นยำกว่าถึง 3 SKU ดังนั้นการพยากรณ์ด้วยวิธี MA เป็นวิธีที่น่าสนใจหากนำเอาวิธีดังกล่าวไปประยุกต์ใช้งานจริงกับการพยากรณ์ความต้องการสินค้าใหม่ใน 3 เดือนแรก เพราะเป็นวิธีการที่ง่าย ใช้เวลาน้อยในการพยากรณ์ และเหมาะกับข้อมูลที่ไม่ซับซ้อนและมีความคงที่ แม้เป็นวิธีการดั้งเดิมแต่ก็ให้ค่าความแม่นยำสูงเมื่อเทียบกับอีก 2 วิธีที่ใช้เวลาและมีวิธีการที่ซับซ้อนมากกว่า ซึ่งผลการพยากรณ์ที่ได้นั้นไม่ได้แตกต่างจากวิธี MA มากนัก โดยเฉพาะค่าพยากรณ์ยอดขายเดือนแรก ซึ่งทั้ง 3 วิธีแตกต่างกันเพียงเล็กน้อยเท่านั้น เนื่องจากการนำไปใช้งานจริงต้องทำการปัดเลขเป็นจำนวนเต็ม เพื่อนำค่าที่ได้ไปใช้งานในการสั่งซื้อใหม่ครั้งแรกมาสต๊อกก่อนสินค้าใหม่เริ่มวางขาย และในการนำไปใช้งานจริงในการสั่งซื้อสินค้าใหม่นั้นใช้เพียงค่าพยากรณ์ยอดขายในเดือนแรก เนื่องจากเมื่อสินค้าใหม่เริ่มวางขายไปได้ 1 เดือนก็สามารถใช้ยอดขายจริงในเดือนแรกของสินค้าใหม่ในการพยากรณ์ความต้องการในเดือนถัดไปเองได้ โดยไม่ต้องพึ่งพาข้อมูลการขายของสินค้าปัจจุบันที่มีคุณสมบัติคล้ายคลึงกันแล้ว ผู้วิจัยทำการพยากรณ์ไปถึง 3 เดือน เพื่อเป็นข้อมูลในการช่วยตัดสินใจในการพิจารณาเลือกสินค้าใหม่ดังกล่าวเข้ามาวางจำหน่ายในร้านหรือไม่ และปริมาณเท่าใดถึงเหมาะสม จึงต้องดูค่าพยากรณ์ยอดขายในเดือนที่ 1-3 ประกอบด้วยเพื่อพิจารณาความคุ้มค่ากับการนำมาจำหน่าย และคาดการณ์ระยะเวลาในการขายสินค้า จากผล

การทดลองสามารถนำมาเป็นข้อมูลพื้นฐานในการช่วยตัดสินใจในการสั่งซื้อสินค้าใหม่ เพื่อลดปัญหาการสั่งซื้อสินค้าใหม่เข้ามามากเกินความต้องการของลูกค้าได้ ในการศึกษาวิจัยนี้ผู้วิจัยทำการศึกษาเปรียบเทียบประสิทธิภาพของ 3 แบบจำลองและชุดข้อมูลธุรกรรมการขายจากเพียงร้านเดียวเท่านั้น ซึ่งอาจมีแบบจำลองอื่น ๆ ที่มีประสิทธิภาพที่ดีกว่าและหากใช้ข้อมูลจากร้านค้าหลายร้าน อาจช่วยให้แบบจำลองเรียนรู้ข้อมูลได้มากขึ้น และมีประสิทธิภาพในการพยากรณ์ความต้องการได้แม่นยำมากยิ่งขึ้น

5.3 ข้อเสนอแนะ

1. ในงานวิจัยนี้ใช้เพียงยอดขายเป็นจำนวนชิ้นเพียงอย่างเดียวในการฝึกสอนแบบจำลอง ควรลองใช้ตัวแปรอื่น ๆ เพิ่มเติม เช่น ราคา โปรโมชั่น ข้อมูลทางเศรษฐกิจ หรือปัจจัยอื่น ๆ ที่ส่งผลต่อการซื้อสินค้าของลูกค้า

2. ในงานวิจัยนี้ใช้ข้อมูลเพียง 24 เดือนในการฝึกสอนแบบจำลอง ควรเพิ่มปริมาณข้อมูลให้มากขึ้น เพื่อช่วยให้แบบจำลองเรียนรู้ข้อมูลได้มากขึ้นและมีประสิทธิภาพในการพยากรณ์ความต้องการได้แม่นยำมากยิ่งขึ้น

3. ในงานวิจัยนี้เลือกสินค้าจาก 3 หมวดหมู่ ในการศึกษาการสร้างแบบจำลองการพยากรณ์ความต้องการสินค้าใหม่ ควรลองศึกษาเพิ่มเติมในสินค้าหมวดหมู่อื่น ๆ ที่หลากหลายมากยิ่งขึ้น

4. ในงานวิจัยนี้ใช้ข้อมูลศึกษาจากร้านค้าเพียงแห่งเดียว หากใช้ข้อมูลจากร้านค้าหลายร้าน อาจช่วยให้แบบจำลองเรียนรู้ข้อมูลได้มากขึ้น และมีประสิทธิภาพในการพยากรณ์แม่นยำขึ้น

บรรณานุกรม

Ali, Y., & Nakti, S. (2023, 15-17 March 2023). *Sales Forecasting: A Comparison of Traditional and Modern Times-Series Forecasting Models on Sales Data with Seasonality*. Paper presented at the 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom).

Dai, Y., & Huang, J. (2021, 24-26 April 2021). *A Sales Forecast Method for Products with No Historical Data.* Paper presented at the 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA).

Dave, E., Leonardo, A., Jeanice, M., & Hanafiah, N. (2021). Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM. *Procedia Computer Science,* 179, 480-487.

Elalem, Y. K., Maier, S., & Seifert, R. W. (2023). A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks. *International Journal of Forecasting,* 39(4), 1874-1894.

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.

Jain, A., Karthikeyan, V., S, B., S, B. R., S, K., & B, S. (2020, 6-8 Nov. 2020). *Demand Forecasting for E-Commerce Platforms.* Paper presented at the 2020 IEEE International Conference for Innovation in Technology (INOCON).

Nithin, S. S. J., Rajasekar, T., Jayanthy, S., Karthik, K., & Rithick, R. R. (2022, 16-18 March 2022). *Retail Demand Forecasting using CNN-LSTM Model.* Paper presented at the 2022 International Conference on Electronics and Renewable Systems (ICEARS).

Olah, C. (2015). Understanding LSTM Networks. Retrieved from. Retrieved from http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Tang, T. (2023, 26-29 May 2023). *Analysis and Demand Forecasting Based On e-Commerce Data.* Paper presented at the 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD).

Tangruamsub, S. (2017). Long Short-Term Memory (LSTM). Retrieved from https://medium.com/@sinart.t/long-short-term-memory-lstm-e6cb23b494c6

Tunpaiboon, N. (2562). แนวโน้มธุรกิจ/อุตสาหกรรมปี 2562-2564: ธุรกิจร้านค้าปลีกสมัยใหม่. https://www.krungsri.com/th/research/industry/industry-outlook/wholesale-retail/modern-trade/io/io-modern-trade-20

van Steenbergen, R. M., & Mes, M. R. K. (2020). Forecasting demand profiles of new products. *Decision Support Systems,* 139, 113401.

Vithitsoontorn, C., & Chongstitvatana, P. (2022, 9-11 March 2022). *Demand Forecasting in Production Planning for Dairy Products Using Machine Learning and Statistical Method.* Paper presented at the 2022 International Electrical Engineering Congress (iEECON).

แนวทางการสั่งซื้อสินค้า. (2022). สืบค้นจาก https://www.businessplus.co.th/Activities/%E0%B8%82%E0%B9%88%E0%B8%B2%E0%B8%A7%E0%B8%AA%E0%B8%B2%E0%B8%A3-pos-c020

ถนอม บริคุต. (2014). ผลิตภัณฑ์ใหม่ (New Product Planning). https://fifathanom.wordpress.com/2014/12/07/%E0%B8%9A%E0%B8%97%E0%B8%97%E0%B8%B5%E0%B9%88-10-%E0%B8%9C%E0%B8%A5%E0%B8%B4%E0%B8%95%E0%B8%A0%E0%B8%B1%E0%B8%93%E0%B8%91%E0%B9%8C%E0%B9%83%E0%B8%AB%E0%B8%A1%E0%B9%88-new-product-planning/

ผู้ช่วยศาสตราจารย์ ดร. เฉลิมพล จตุพรม. (2562). การพยากรณ์ทางอนุกรมเวลา. In. https://cj007blog.files.wordpress.com/2020/04/07-time-series-forecasting.pdf

มหาวิทยาลัยเชียงใหม่. (2553). ทฤษฎีข้อมูลอนุกรมเวลา, 12-22. https://archive.lib.cmu.ac.th/full/T/2553/econ0553wp_ch2.pdf

วีรภัทร สาธิตคณิตกุล. (2023). วิธีวัดผลความแม่นยำโมเดล. https://bdi.or.th/big-data-101/mape-evaluation/

สินค้า SKU. (2020). สืบค้นจาก https://pospos.co/article/detail/get-to-know-sku

ประวัติผู้เขียน

ชื่อ-สกุล	สุพรรษา วัฒนบุตร
วัน เดือน ปี เกิด	25 มกราคม พ.ศ. 2536
สถานที่เกิด	กรุงเทพมหานคร
วุฒิการศึกษา	พ.ศ. 2555 วิทยาศาสตรบัณฑิต สาขาสถิติประยุกต์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง พ.ศ. 2565 วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูล มหาวิทยาลัยศรีนครินทรวิโรฒ
ที่อยู่ปัจจุบัน	506/8 หมู่ 6 ตำบลเชียงรากน้อย อำเภอบางปะอิน จังหวัดพระนครศรีอยุธยา 13180