การพยากรณ์จำนวนการขายสินค้าของร้านขายเบเกอรี่ กรณีศึกษาร้านเบเกอรี่แห่งหนึ่ง

FORECASTING BAKERY PRODUCT SALES : A CASE STUDY OF A BAKERY SHOP

ธนกฤต เปาวสันต์

สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาตรบัณฑิต(สถิติประยุกต์)
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา2566

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

FORECASTING BAKERY PRODUCT SALES : A CASE STUDY OF A BAKERY SHOP

Thanakrit Paowasan

A COOPARATIVE EDUCATION SUBMITTED IN
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS)
DEPARTMENT OF STATISTICS, SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2023

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา	การพยากรณ์จำนวนการขายสินค้าของร้านขายเบเกอรี่ กรณีศึกษาร้านเบเกอรี่แห่งหนึ่ง Forecasting Bakery Product Sales: A Case Study of a Bakery Shop
ชื่อนักศึกษา	นาย ธนกฤต เปาวสันต์ รหัสนักศึกษา 63050625
ปริญญา	วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา	สถิติ
ปีการศึกษา	2566
อาจารย์ที่ปรึกษา	รศ.ดร.วลัยลักษณ์ อัตธีรวงศ์

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้สหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา วิทยาศาสตรบัณฑิต (สถิติประยุกต์) ประจำปีการศึกษา 2566

คณะกรรมการสอบ	ลายมือชื่อ
ดร.สกุณา ศรีอโนมัย ประธานกรรมการ	
คุณ สมิตา ยงไพศาล กรรมการ	
รศ.ดร.วลัยลักษณ์ อัตธีรวงศ์ กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อสหกิจศึกษา การพยากรณ์จำนวนการขายสินค้าของร้านขายเบเกอรี่ กรณีศึกษาร้านเบเกอรี่แห่งหนึ่ง
Forecasting Bakery Product Sales: A Case Study of a Bakery Shop
ชื่อนักศึกษา นาย ธนกฤต เปาวสันต์ รหัสนักศึกษา 63050625
ปริญญา วิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชา สถิติ
คณะ วิทยาศาสตร์
มหาวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.)
ปีการศึกษา 2566
อาจารย์ที่ปรึกษา รศ.ดร.วลัยลักษณ์ อัตธีรวงศ์

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์ในการวิเคราะห์หาตัวแบบที่เหมาะสมในการพยากรณ์ยอดขายรายวันของสินค้าทั้งหมดที่มีการขายมากกว่า 60 วันในร้านเบเกอรี่สาขาหนึ่งเพื่อเป็นแนวทางในการผลิตสินค้าในแต่ละวันให้เหมาะสมกับความต้องการของลูกค้าโดยใช้ข้อมูลการขายสินค้าย้อนหลัง 1 ปี โดยวิธีการคัดเลือกตัวแปรที่ใช้ในงานวิจัยนี้มี 3 รูปแบบคือ 1) ใช้ตัวแปรอิสระทั้งหมด 2) ใช้ตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามโดยใช้สมการถดถอยเชิงเส้นพหุคูณในการคัดเลือกตัวแปร 3) ใช้ตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามโดยใช้สมการถดถอยเชิงเส้นพหุคูณที่แปลงข้อมูลยอดขายรายวันด้วยลอการิทึมธรรมชาติ และนำตัวแปรทั้ง 3 รูปแบบให้ตัวแบบการเรียนรู้ของเครื่อง 3 ตัวแบบ คือ 1) ตัวแบบต้นไม้ตัดสินใจ 2) ตัวแบบป่าสุ่ม 3) ตัวแบบ XGBoost เรียนรู้และปรับค่าไฮเปอร์พารามิเตอร์ด้วยการค้นหาแบบกริด จากนั้นคำนวณค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยของแต่ละเมนูแล้วนำค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสองของทุกเมนูที่ได้จากแต่ละตัวแบบมาคิดเป็นค่าเฉลี่ยเพื่อนำมาเปรียบเทียบหาตัวแบบและวิธีการคัดเลือกตัวแปรที่ค่าเฉลี่ยค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสองน้อยที่สุด เพื่อนำไปพยากรณ์ยอดขายรายวันของสินค้าทั้ง 89 เมนู โดยผลการศึกษาพบว่า ตัวแบบป่าสุ่มที่ใช้การคัดเลือกตัวแปรรูปแบบที่ 2 ให้ค่าเฉลี่ยค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของทั้ง 89 เมนูน้อยที่สุดเท่ากับ 3.14 จึงนำตัวแบบป่าสุ่มที่ใช้การคัดเลือกตัวแปรรูปแบบที่ 2 ไปใช้ในการพยากรณ์ยอดขายรายวันของสินค้าทั้ง 89 เมนูเพื่อเป็นแนวทางในการตัดสินใจในการผลิตสินค้าของร้านค้ากรณีศึกษา

คำสำคัญ : พยากรณ์ เบเกอรี่ ป่าสุ่ม ต้นไม้ตัดสินใจ สมการถดถอยเชิงเส้นพหุคูณ XGBoost

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Forecasting Bakery Product Sales: A Case Study of a Bakery Shop
Students	Mr. Thanakrit Paowasan Student ID 63050625
Degree	Bachelor of Science (Applied Statistics)
Department	Statistics
School	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2023
Advisor	Assoc.Prof.Dr. Walailak Atthirawong

Abstract

The aim of this study is to examine and identify appropriate models for predicting daily sales of all products sold at a bakery for over 60 days, with the purpose of directing daily production in accordance with customer demand via sales data. The variable selection approaches employed in this study encompass three distinct forms : 1) Utilize all independent variables. 2) Employ variables that are correlated with the dependent variable through multiple linear regression to select variables. 3) Apply the multiple linear regression technique to identify variables correlated with the dependent variable and transform daily sales data using a natural logarithm algorithm. These three forms for variables are subsequently supplied to three machine learning models, i.e. 1) Decision Tree model, 2) Random Forest model, and 3) XGBoost model, all of which utilize grid search to learn and modify hyperparameters. Subsequently, the root mean squared error (RMSE) is computed for every menu derived from each model. The mean squared error of these values is then compared to determine which model and variable selection method yields the smallest average RMSE. The analysis revealed that the Random Forest model, utilizing variable selection technique 2, achieved the lowest average root mean squared error of 3.14 across all 89 menus. Therefore, to help with decision-making about the manufacturing of goods for the case study bakery, the Random Forest use variable selection technique 2 was employed to forecast daily sales of all 89 menus.

Keywords : Bakery, Decision Tree, Forecasting, Multiple Linear Regression, Random Forest, XGBoost

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

สหกิจศึกษาฉบับนี้สำเร็จเสร็จสิ้นได้โดยได้รับความอนุเคราะห์จากอาจารย์ที่ปรึกษาสหกิจศึกษา รศ.ดร. วลัยลักษณ์ อัตธีรวงศ์ ที่เสียสละเวลาในการให้คำปรึกษาและให้คำแนะนำในการทำสหกิจศึกษานี้ตั้งแต่เริ่มจนกระทั่งเสร็จสิ้นการทำสหกิจศึกษาด้วยความเป็นห่วงและเอาใจใส่เป็นอย่างดี ทั้งยังให้กำลังใจตลอดมาผู้วิจัยทราบซึ้งในความเอาใจใส่ของอาจารย์และขอขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอขอบพระคุณ ดร. สกุณา ศรีอโนมัย ที่ให้ความเมตตากรุณาเป็นประธานกรรมการในการสอบการทำสหกิจศึกษานี้ โดยเสียสละเวลาในการให้คำแนะนำในการแก้ไขข้อบกพร่องในรายงานสหกิจฉบับนี้จนกระทั่งเสร็จสิ้นการทำสหกิจศึกษา และขอขอบพระคุณคณาจารย์และบุคลากรประจำภาควิชาสถิติ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่านที่มอบความรู้และประสบการณ์ให้กับผู้วิจัย

ขอขอบพระคุณ คุณ สมิตา ยงไพศาล ในความเมตตาให้การดูแล ช่วยเหลือ มอบความรู้และประสบการณ์ในการทำสหกิจครั้งนี้แก่ผู้ทำวิจัยตั้งแต่เริ่มทำสหกิจศึกษาจนกระทั่งแล้วเสร็จการทำสหกิจศึกษา ขอขอบพระคุณบุคลากรท่านอื่นๆในบริษัทที่ให้คำแนะนำและมอบประสบการณ์ในการสหกิจศึกษาครั้งนี้แก่ผู้วิจัย

สุดท้ายนี้ขอขอบพระคุณ บิดา มารดา และคนรอบตัวที่สนับสนุนในทุกด้านและให้กำลังใจแก่ผู้ทำวิจัยตลอดมาจนกระทั่งงานวิจัยนี้เสร็จสิ้น โดยผู้วิจัยหวังเป็นอย่างยิ่งว่างานวิจัยนี้จะเป็นประโยชน์กับหน่วยงานที่เกี่ยวข้องและผู้ที่เข้ามาศึกษาเพื่อนำไปพัฒนาต่อยอดต่อไป

ธนกฤต เปาวสันต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ก
บทคัดย่อภาษาอังกฤษ..... ข
กิตติกรรมประกาศ..... ค
สารบัญ..... ง
สารบัญตาราง..... ฉ
สารบัญรูป..... ช
บทที่ 1 บทนำ..... 1
1.1 ความเป็นมาและความสำคัญ..... 1
1.2 วัตถุประสงค์..... 2
1.3 ขอบเขต..... 2
1.4 ประโยชน์ที่คาดว่าจะได้รับ..... 2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง..... 3
2.1 ทฤษฎีและหลักการที่เกี่ยวข้อง..... 3
2.1.1 นิยามของการพยากรณ์..... 3
2.1.2 ประเภทของการพยากรณ์..... 3
2.1.2.1 การพยากรณ์เชิงคุณภาพ..... 3
2.1.2.2 การพยากรณ์เชิงปริมาณ..... 4
2.1.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ..... 4
2.1.4 การเรียนรู้ของเครื่อง..... 5
2.1.4.1 การเรียนรู้แบบมีผู้สอน..... 5
2.1.4.2 การเรียนรู้แบบไม่มีผู้สอน..... 6
2.1.4.3 การเรียนรู้แบบเสริมแรง..... 6
2.1.5 ต้นไม้ตัดสินใจ..... 6
2.1.6 ป่าสุ่ม..... 7
2.1.7 eXtreme Gradient Boosting..... 9
2.1.8 การปรับไฮเปอร์พารามิเตอร์..... 9
2.1.9 ตัววัดประสิทธิภาพตัวแบบ..... 9
2.2 งานวิจัยที่เกี่ยวข้อง..... 9
บทที่ 3 วิธีการดำเนินงานวิจัย..... 12
3.1 ขั้นตอนการดำเนินงาน..... 12
3.1.1 ลักษณะข้อมูล..... 13
3.1.2 ทำความสะอาดข้อมูล..... 13
3.1.3 การเตรียมข้อมูล..... 14
3.1.3.1 การแปลงข้อมูล..... 14
3.1.3.2 การแบ่งข้อมูล..... 14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.1.4 คัดเลือกตัวแปร 14
3.1.5 ตัวแบบที่ใช้ในงานวิจัย 29
3.1.5.1 ต้นไม้ตัดสินใจ 29
3.1.5.2 ป่าสุ่ม 29
3.1.5.3 eXtreme Gradient Boosting 30
3.1.6 การวัดประสิทธิภาพของตัวแบบพยากรณ์ 30
3.2 ตัวแปรที่ใช้ในงานวิจัย 31
3.2.1 ตัวแปรอิสระ 31
3.2.2 ตัวแปรตาม 33
3.3 เครื่องมือที่ใช้ในงานวิจัย 33
บทที่ 4 ผลการวิจัยและการอภิปรายผล 34
4.1 ผลการศึกษาตัวแบบต้นไม้ตัดสินใจ 34
4.2 ผลการศึกษาตัวแบบป่าสุ่ม 35
4.3 ผลการศึกษาตัวแบบ eXtreme Gradient Boosting 36
4.4 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบ 37
4.5 ผลการพยากรณ์ยอดขาย 7 วันข้างหน้า 37
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 39
5.1 สรุปผลการวิจัย 39
5.2 ข้อเสนอแนะ 41
5.3 ข้อจำกัดในงานวิจัย 41
เอกสารอ้างอิง 42
ภาคผนวก 43
ภาคผนวก ก 44
ภาคผนวก ข 80

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ **หน้า**

3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2..... 15
3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3..... 22
3.3 ไฮเปอร์พารามิเตอร์และค่าที่ใช้ในการปรับของตัวแบบต้นไม้ตัดสินใจ..... 29
3.4 ไฮเปอร์พารามิเตอร์และค่าที่ใช้ในการปรับของตัวแบบป่าสุ่ม..... 29
3.5 ไฮเปอร์พารามิเตอร์และค่าที่ใช้ในการปรับของตัวแบบ Xgboost..... 30
3.6 ตัวแปรอิสระที่ใช้ในงานวิจัย..... 31
3.7 ตัวแปรตามที่ศึกษา..... 33
4.1 ค่าเฉลี่ย RMSE ทั้ง 3 รูปแบบการคัดเลือกตัวแปรที่ได้จากตัวแบบทั้ง 3 ตัวแบบ..... 37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

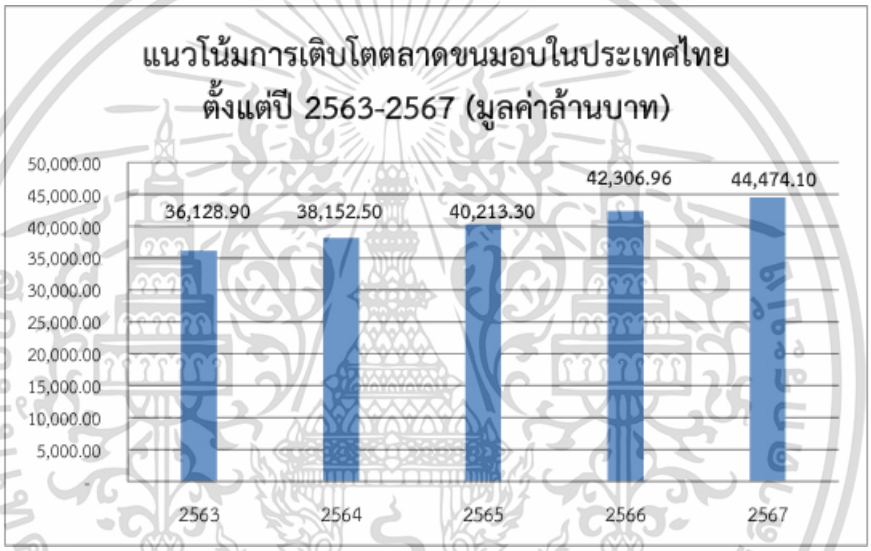
รูปที่		หน้า
1.1	มูลค่าการเติบโตตลาดขนมอบในประเทศไทยปี 2563-2567	1
2.1	ประเภทของ Machine Learning	5
2.2	ตัวอย่างแผนผังการตัดสินใจ CART	7
2.3	ตัวอย่างการทำงาน Random Forest Classification	8
2.4	ตัวอย่างการทำงาน Random Forest Regression	8
2.5	ตัวอย่างการทำงานของ eXtreme Gradient Boosting (XGboost)	9
3.1	ขั้นตอนการดำเนินงาน	12
3.2	ข้อมูลการขายสินค้ารายวันของแต่ละเมนู	13
3.3	ค่าสูญหายของแต่ละคอลัม	13
3.4	ตัวอย่างการใช้ One-hot Encoder ในการแปลงข้อมูลให้อยู่ในรูปตัวเลข	14
3.5	หน้าแรกของ Visual Studio Code	33
4.1	ค่าเฉลี่ย RMSE ของทั้ง 3 รูปแบบการคัดเลือกตัวแปรที่ได้จากตัวแบบต้นไม้ตัดสินใจ	34
4.2	ค่าเฉลี่ย RMSE ของทั้ง 3 รูปแบบการคัดเลือกตัวแปรที่ได้จากตัวแบบป่าสุ่ม	35
4.3	ค่าเฉลี่ย RMSE ของทั้ง 3 รูปแบบการคัดเลือกตัวแปรที่ได้จากตัวแบบXGBoost	36
4.4	ตัวอย่างการทำนายยอดขายสินค้า 7 วันข้างหน้าเมนู A	38
4.5	ตัวอย่างการทำนายยอดขายสินค้า 7 วันข้างหน้าเมนู B	38

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ธุรกิจเบเกอรี่ในสังคมปัจจุบันเป็นที่นิยมอย่างแพร่หลายในคนทุกเพศทุกวัยเนื่องมาจากไลฟ์สไตล์ที่เร่งรีบและชอบความสะดวกสบายร้านเบเกอรี่ที่มีขนมหลากหลายมีสถานที่ตั้งทำเลค่อนข้างดีจึงได้รับความนิยมอย่างต่อเนื่อง ทำให้มีธุรกิจเบเกอรี่เปิดใหม่มากขึ้นและส่งผลให้การแข่งขันสูงตามไปด้วย ธุรกิจเบเกอรี่ที่เปิดมานานจึงควรติดตามเทคโนโลยีต่างๆที่สามารถนำมาพัฒนาธุรกิจของตนเองให้ดีขึ้นเพื่อให้ธุรกิจอยู่รอดรักษาฐานลูกค้าไว้และยังเติบโตในการหาฐานลูกค้าใหม่ได้เพิ่มขึ้น



รูปที่ 1.1 มูลค่าการเติบโตตลาดขนมอบในประเทศไทยปี 2563-2567
(ที่มา : ศูนย์อัจฉริยะเพื่ออุตสาหกรรมอาหาร สถาบันอาหาร,2563)

การนำข้อมูลมาใช้ในธุรกิจถือว่ามีความสำคัญมากในการทำธุรกิจจนมีประโยคที่ว่า “Data is the New Oil” นั้นคือเปรียบว่า “ข้อมูลมีค่าเหมือนเป็นแหล่งน้ำมันแหล่งใหม่” เพราะข้อมูลคือสิ่งที่สามารถทำให้ทราบถึงความต้องการของลูกค้าและรูปแบบของทำเลที่ตั้งแต่ละแบบมีความต้องการเป็นอย่างไรทำให้สามารถนำเสนอสินค้าหรือจัดสินค้าที่จะขายในร้านได้ตรงกับความต้องการของผู้ซื้อส่งผลให้ยอดขายและกำไรเพิ่มขึ้น และยังสามารถทำให้รู้ว่าความต้องการของตลาดเป็นอย่างไรเนื่องจากปัจจุบันความต้องการของตลาดเปลี่ยนไปอย่างรวดเร็ว การที่รู้ว่าตลาดเบเกอรี่มีความสนใจอย่างไรจะช่วยให้ปรับสินค้าแต่ละประเภทตามความต้องการของตลาดได้ทันและยังสามารถพัฒนาประสิทธิภาพของกระบวนการผลิตเพื่อลดการขาดทุนได้อีกด้วย

ในสาขาที่ผู้วิจัยทำการศึกษามีปัญหาในการที่ขายสินค้าได้จำนวนมากแต่กำไรเฉลี่ยของสาขาน้อยมาจากการผลิตสินค้ามากเกินไปทำให้ต้องมีการลดราคาและทิ้งสินค้าที่ขายไม่หมด ผู้วิจัยจึงสร้างตัวแบบในการพยากรณ์หาจำนวนสินค้าที่ต้องผลิตในแต่ละวันและเปรียบเทียบประสิทธิภาพในการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พยากรณ์เพื่อหาตัวแบบที่เหมาะสมกับการพยากรณ์จำนวนสินค้าที่ต้องผลิตเพื่อลดการลดราคาและการทิ้งสินค้า

1.2 วัตถุประสงค์ของงานวิจัย

1) เพื่อวิเคราะห์หาตัวแบบที่เหมาะสมในการทำนายยอดขายรายวันของสินค้าทั้ง 89 เมนูที่ขายในร้านค้าด้วยวิธีการเรียนรู้ของเครื่องโดยใช้เทคนิค Decision Tree, Random Forest และ XGboots

2) เพื่อเปรียบเทียบประสิทธิภาพของตัวแบบและหาตัวแบบที่เหมาะสมในการทำนายสินค้าทั้ง 89 เมนู โดยใช้ค่าเฉลี่ย RMSE ของทุกเมนูจากแต่ละตัวแบบการเรียนรู้ของเครื่องในการเปรียบเทียบ

3) เพื่อพยากรณ์จำนวนยอดขายรายวันล่วงหน้า 7 วันด้วยตัวแบบที่มีค่าเฉลี่ย RMSE น้อยที่สุด

1.3 ขอบเขตงานวิจัย

1) ขอบเขตด้านข้อมูล

งานวิจัยนี้ใช้ข้อมูลการขายสินค้ารายวันของร้านเบเกอรี่สาขา A ในเครือห้างสรรพสินค้าแห่งหนึ่งที่มีการขายแต่ละรายการมากกว่า 60 วันโดยเก็บรวบรวมข้อมูลเป็นระยะเวลา 1 ปี เริ่มตั้งแต่ปี 2566 ถึง 2567

2) ขอบเขตด้านระยะเวลา

เก็บข้อมูลตั้งแต่ 1 มกราคม 2566 ถึงวันที่ 21 กุมภาพันธ์ 2567

3) ขอบเขตด้านเครื่องมือ

ใช้โปรแกรม Visual Studio Code ในการเขียนภาษา Python

ใช้โปรแกรม Microsoft Excel ในการใช้เป็นฐานข้อมูลในการสร้างตาราง

4) ขอบเขตด้านความปลอดภัยของข้อมูล

เนื่องจากข้อมูลที่ผู้วิจัยใช้เป็นข้อมูลที่มีข้อจำกัดด้านการเผยแพร่ทำให้ไม่สามารถเปิดเผยข้อมูลที่ใช้โดยตรงได้จึงต้องทำการสมมุติตัวแปรและปิดบังข้อมูลบางส่วน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) ใช้ค่าที่พยากรณ์ที่ได้เป็นแนวทางในการช่วยตัดสินใจผลิตสินค้าในแต่ละวัน

2) ช่วยเพิ่มกำไรด้วยการลดสินค้าที่จำเป็นต้องทิ้งเนื่องจากขายสินค้าไม่หมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ผู้วิจัยได้รวมรวมทฤษฎีและงานวิจัยที่เกี่ยวข้องกับหัวข้อที่ผู้วิจัยทำการศึกษาเพื่อเป็นพื้นฐานในการพัฒนาตัวแบบที่ใช้ในงานวิจัย โดยมีรายละเอียดดังนี้

2.1 ทฤษฎีและหลักการที่เกี่ยวข้อง

2.1.1 นิยามของการพยากรณ์

2.1.2 ประเภทของการพยากรณ์

2.1.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ

2.1.4 การเรียนรู้ของเครื่อง

2.1.5 ต้นไม้ตัดสินใจ

2.1.6 ป่าสุ่ม

2.1.7 eXtreme Gradient Boosting

2.1.8 การปรับไฮเปอร์พารามิเตอร์

2.1.9 ตัววัดประสิทธิภาพของตัวแบบ

2.2 งานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีและหลักการที่เกี่ยวข้อง

2.1.1 นิยามของการพยากรณ์

พิภพ (2549) กล่าวว่าการพยากรณ์ หมายถึง การประมาณการถึงค่าและคุณลักษณะที่เกิดขึ้นในอนาคต การพยากรณ์ในกิจกรรมการค้าปลีกนั้นมีความสำคัญมาก มีจุดประสงค์เพื่อรู้ถึงความต้องการที่จะเกิดขึ้นในอนาคต เพื่อใช้ในการวางแผนและจัดสรรสินค้าและบริการได้อย่างมีประสิทธิภาพ

สมศรี (2559) กล่าวว่าการพยากรณ์ หมายถึง การคาดคะเนหรือทำนายเหตุการณ์หรือสภาพการณ์ต่าง ๆ ในอนาคต โดยการพยากรณ์จะทำการศึกษาแนวโน้มและรูปแบบการเกิดของเหตุการณ์หรือสภาพการณ์จากข้อมูลในอดีต หรือใช้ความรู้ความสามารถ ประสบการณ์ และวิจารณญาณของผู้พยากรณ์

2.1.2 ประเภทของการพยากรณ์

การพยากรณ์ในปัจจุบันแบ่งออกเป็น 2 กลุ่มคือ การพยากรณ์เชิงคุณภาพ (Qualitative Forecasting) และการพยากรณ์เชิงปริมาณ (Quantitative Forecasting) โดยมีรายละเอียดดังนี้

2.1.2.1 การพยากรณ์เชิงคุณภาพ (Qualitative Forecasting)

สมศรี (2559) กล่าวว่า การพยากรณ์เชิงคุณภาพ เป็นการพยากรณ์ที่ผู้พยากรณ์ต้องอาศัยความรู้ความสามารถ ประสบการณ์ และวิจารณญาณในเรื่องที่จะพยากรณ์ การพยากรณ์รูปแบบนี้ไม่มีรูปแบบ กฎเกณฑ์หรือสูตรในการคำนวณที่แน่นอน

พิภพ (2549) กล่าวว่า การพยากรณ์เชิงคุณภาพเป็นวิธีที่อาศัยวิจารณญาณของผู้เชี่ยวชาญ หรือความเห็นของผู้เชี่ยวชาญในการคาดคะเนถึงเหตุการณ์ที่จะเกิดขึ้นในอนาคต วิธีดัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งไม่อนุญาตให้นำไปเผยแพร่ แจกจ่าย ทำซ้ำ หรือดัดแปลงเนื้อหาในเอกสารนี้

กล่าวจะเป็นประโยชน์กับการพยากรณ์ในระยะกลางและระยะยาว การใช้วิธีการพยากรณ์เชิงคุณภาพนี้ไม่มีหลักเกณฑ์ที่แน่นอน เพียงแต่ใช้ความรู้และประสบการณ์ของผู้เชี่ยวชาญหรือผู้บริหารเป็นเกณฑ์ ทั้งนี้อาจจะเพราะผู้พยากรณ์มีข้อมูลไม่เพียงพอที่จะนำมาใช้การพยากรณ์เชิงปริมาณหรืออาจเป็นเพราะสถานการณ์ในอดีตและอนาคตมีความแตกต่างกันมากจนทำให้ข้อมูลในอดีตไม่สามารถนำมาเป็นแนวทางในการพิจารณาได้ วิธีการพยากรณ์เชิงคุณภาพมีหลากหลายวิธีโดยจะยกตัวอย่างวิธีการพยากรณ์เชิงคุณภาพดังที่จะกล่าวต่อไปนี้

1) วิธีเดลฟลาย เป็นวิธีการพยากรณ์โดยสร้างกลุ่มของผู้เชี่ยวชาญขึ้น และสร้างแบบสอบถามแบบต่อเนื่องนั้นคือคำตอบของแบบสอบถามหนึ่งจะถูกนำมาใช้เป็นข้อมูลสำหรับแบบสอบถามต่อไป

2) วิธีวิจัยการตลาด เป็นวิธีที่ใช้ข้อมูลการตลาดที่ได้มาจากการส่งแบบสอบถาม การสำรวจทางโทรศัพท์ การอภิปรายกลุ่ม และการสัมภาษณ์ ในการพยากรณ์

3) วิธี Panel Consensus เป็นการรวบรวมผู้เชี่ยวชาญเพื่ออภิปรายปัญหาต่างๆจนกว่าจะได้รับข้อสรุปอันถือว่าเป็นการพยากรณ์ได้

4) วิธี Grass-Roots Forecasting ใช้วิธีสอบถามบุคลากรที่อยู่ใกล้ชิดปัญหา เพื่อให้การพยากรณ์อยู่ในขอบเขตที่รับผิดชอบ แล้วจึงนำการพยากรร์ของแต่ละคนมารวมกันเป็นการพยากรณ์รวม

2.1.2.2 การพยากรณ์เชิงปริมาณ (Quantitative Forecasting)

การพยากรณ์เชิงปริมาณ เป็นวิธีพยากรณ์ที่ใช้ข้อมูลในอดีตมาเป็นหลักในการพิจารณาถึงสถานการณ์ที่จะเกิดขึ้นในอนาคตโดยอาศัยหลักสถิติและคณิตศาสตร์ โดยจุดประสงค์ของการพยากรณ์เชิงปริมาณคือต้องการเห็นรูปแบบข้อมูลในอดีตและทำการตีความรูปแบบข้อมูลดังกล่าวเพื่อทราบถึงทิศทางของข้อมูลเหล่านี้ในอนาคต แบ่งออกเป็น 2 แบบคือ

1) การวิเคราะห์อนุกรมเวลา เป็นการศึกษาแบบแผนในการเคลื่อนไหวของอนุกรมเวลาที่กำหนดด้วยรูปแบบอนุกรมเวลาโดยแบบที่ได้จะนำไปใช้ประโยชน์ในการพยากรณ์โดยมีข้อสมมติว่าอนุกรมเวลาในอนาคตจะไม่ต่างไปจากแผนการเคลื่อนไหวในอดีต

2) การศึกษาความสัมพันธ์ระหว่างปัจจัยหรือตัวแปรต่าง ๆ ที่มีผลต่อเหตุการณ์ โดยปัจจัยหรือตัวแปรที่นำมาศึกษาจะต้องมีมากกว่าหนึ่งปัจจัยหรือหนึ่งตัวแปร (สมศรี, 2559)

2.1.3 การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression)

พรรณทิพา (2564) กล่าวว่า การวิเคราะห์การถดถอยเชิงเส้นพหุคูณเป็นการวิเคราะห์การถดถอยที่มีตัวแปรที่เกี่ยวข้องหลายตัวแปร โดยกำหนดตัวแปรที่ต้องการศึกษาเป็นตัวแปรตาม (Dependent Variable) และตัวแปรที่เหลือแทนปัจจัยที่มีอิทธิพลต่อตัวแปรที่เราสนใจศึกษาเป็นตัวแปรอิสระ (Independent Variable) โดยการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระจะทำตามรูปแบบการถดถอยที่กำหนดขึ้นโดยสามารถเขียนเป็นสมการได้ดังนี้

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_i X_i + \varepsilon \quad (1)$$

โดย

Y แทนค่าของตัวแปรตามที่ศึกษา

β_0 แทนค่าคงที่ของตัวแปรตามเมื่อตัวแปรอิสระทั้งหมดมีค่าเป็นศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$X_1,X_2,...,X_i$ แทนค่าตัวแปรอิสระ i

$\beta_1,\beta_2,..,\beta_i$ แทนค่าสัมประสิทธิ์การถดถอยของ $X_1,X_2,...,X_i$

ε แทนค่าคลาดเคลื่อน

โดยการประมาณค่าสัมประสิทธิ์ของสมการถดถอยที่ใช้ในงานวิจัยนี้คือ วิธีกำลังสองน้อยที่สุด (Least Squares Method) นั้นคือการทำให้ผลรวมความคลาดเคลื่อนกำลังสองมีค่าน้อยที่สุด

2.1.4 การเรียนรู้ของเครื่อง (Machine Learning : ML)

วิโรจน์ (2566) กล่าวว่า การเรียนรู้ของเครื่อง (Machine Learning :ML) เป็นส่วนหนึ่งของการพัฒนาระบบปัญญาประดิษฐ์ (Artificial Intelligent: AI) ที่สามารถทำการเรียนรู้ จดจำ คำนวณ และปรับปรุงตัวแบบได้ด้วยตัวเอง เพื่อให้ได้ผลลัพธ์ตามวัตถุประสงค์ที่ต้องการ โดยที่ไม่ต้องมีการกำหนด วิธีคำนวณ ตรรกะเพื่อให้ได้ผลลัพธ์ที่ต้องการและเมื่อมีการเปลี่ยนแปลงของข้อมูลก็ไม่จำเป็นต้องกำหนดการทำงานใหม่โดยการเรียนรู้ของเครื่องจะทำงานด้วยการสร้างระบบที่เรียนรู้ความสัมพันธ์ระหว่างข้อมูลนำเข้าและผลลัพธ์ เพื่อหาตัวแบบการคำนวณที่เหมาะสมต่อจุดประสงค์ที่ต้องการ โดยการเรียนรู้ของเครื่องแบ่งออกเป็น 3 ประเภทดังแสดงในรูปที่ 2.1

รูปที่ 2.1 ประเภทของ Machine Learning
(ที่มา Thanyavuth Akarasomcheep,2565)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.4.1 การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การเรียนรู้แบบมีผู้สอน เป็นการเรียนรู้โดยที่ผู้สอนจะให้การเรียนรู้ของเครื่องเรียนรู้ทั้งข้อมูลนำเข้า (input) และ ผลลัพธ์ (output) (วิโรจน์, 2566) จากนั้นเครื่องจะทำการหาความสัมพันธ์ของข้อมูลนำเข้า (input) และ ผลลัพธ์ (output) และเก็บไว้ในรูปแบบที่เหมาะสมเพื่อใช้ในการปรับปรุงและหาผลลัพธ์เมื่อมีการนำเข้าข้อมูลใหม่ โดยการเรียนรู้แบบมีผู้สอนแบ่งได้สองประเภทดังนี้

การจำแนก (Classification)

การจำแนกเป็นการแบ่งประเภทข้อมูลหรือทำนายข้อมูลที่ไม่มีความต่อเนื่องกันเช่น เป็นเด็กหรือคนแก่ ผู้หญิงหรือผู้ชาย หรือการทำนายว่าเป็นผู้ป่วยมะเร็งหรือไม่จากข้อมูลนำเข้าที่เป็นปัจจัยของผู้ที่เป็นมะเร็งและผลเฉลยว่าชุดข้อมูลไหนเป็นของผู้ป่วยมะเร็ง โดยเครื่องจะทำการเรียนรู้รูปแบบเพื่อคำนวณความน่าจะเป็นของข้อมูลที่ได้รับใหม่และแยกประเภทว่าข้อมูลชุดนี้เป็นของผู้มีโอกาสเป็นมะเร็งหรือไม่ นอกจากนี้การจำแนกยังสามารถนำไปใช้งานอื่นๆได้อีกเช่น การตรวจจับวัตถุ เป็นต้น

การถดถอย (Regression)

การถดถอยมักใช้ในการทำนายช่วงเวลาถัดไปของข้อมูลที่มีความต่อเนื่องกัน เช่น การทำนายจำนวนคนที่จะใช้บริการรถไฟฟ้า (Y) จากปัจจัยที่ส่งผลให้คนใช้บริการรถไฟฟ้า (X) รถติดหรือไม่ ฝนตกหรือไม่ มีการปรับราคาน้ำมันขึ้นหรือไม่ โดยเครื่องจะสร้างสมการพยากรณ์จากข้อมูลที่นำเข้าและผลลัพธ์ที่ให้เครื่องเรียนรู้ จากนั้นเมื่อมีข้อมูลใหม่ถูกนำเข้าเครื่องจะทำการพยากรณ์และปรับปรุงเพื่อให้ได้ผลลัพธ์ที่ดีที่สุด

2.1.4.2 การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอน มีความต่างจากการเรียนรู้แบบมีผู้สอนคือข้อมูลที่ให้เรียนรู้จะไม่มีผลลัพธ์ (วิโรจน์, 2566) และเครื่องจะทำการประมวลผลจัดกลุ่มข้อมูลตามจำนวนกลุ่มที่ผู้ใช้ตั้งค่าไว้จากนั้นผู้ใช้จึงนำข้อมูลที่ผ่านการจัดกลุ่มแล้วไปวิเคราะห์เพื่อหาผลลัพธ์ต่อไป

2.1.4.3 การเรียนรู้แบบเสริมแรง (Reinforcement Learning)

การเรียนรู้แบบเสริมแรง เป็นการเรียนรู้ด้วยการลองผิดลองถูกเพื่อให้ได้รับรางวัลมากที่สุดในสภาพแวดล้อมขณะนั้น (วิโรจน์, 2566) โดยการเรียนรู้แบบเสริมแรงสามารถยกตัวอย่างเปรียบเทียบกับการทดลองทางจิตวิทยา เช่น ทดลองกับอีกาโดยวางวัตถุที่ปล่อยอาหารออกมาเมื่อหยอดเหรียญพร้อมวางเหรียญและวัตถุอื่นๆไว้ในบริเวณนั้น จากนั้นอีกาจะทดลองกระทำการหลายอย่างในสภาพแวดล้อมนั้นจนเมื่ออีกานำเหรียญไปใส่ในวัตถุใส่อาหารก็จะได้รับรางวัลคืออาหาร จะทำให้อีการเรียนรู้หลังจากนั้นว่าการนำเหรียญไปใส่จะได้รับรางวัล

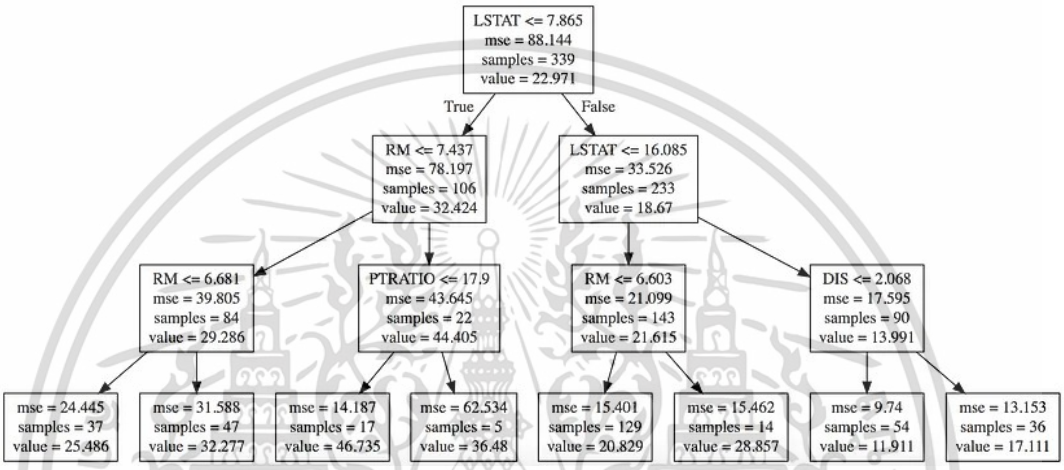
2.1.5 ต้นไม้ตัดสินใจ (Decision Tree)

Brieman (1984) กล่าวว่าต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้ของเครื่องรูปแบบหนึ่งที่นำข้อมูลมาสร้างแบบตัวแบบเพื่อพยากรณ์ในรูปแบบโครงสร้างต้นไม้จากบนลงล่างและมีเป้าหมายในการสร้างตัวแบบที่พยากรณ์ตัวแปรตามโดยเรียนรู้การตัดสินใจจากคุณลักษณะของข้อมูลที่มี โดยเริ่มจากตัวแปรที่มีอิทธิพลต่อการจำแนกข้อมูลมากที่สุด (Root Node) และต่อกิ่งออกไปยังโหนดถัดไปจนกว่าจะไม่สามารถแยกโหนดได้หรือไม่มีเงื่อนไขในการแยกโหนดแล้วจึงจะแสดงผลลัพธ์ของข้อมูลนั้นคือใบ (Leaf Node) โดยในงานวิจัยนี้ผู้วิจัยใช้อัลกอริทึม ต้นไม้ตัดสินใจแบบจำแนกและแบบ

เอกสารนี้ [illegible] ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถดถอย (Classification And Regression Tree : CART) ในการสร้างตัวแบบต้นไม้ตัดสินใจแสดงตัวอย่างการทำงานในรูปที่ 2.2

วิโรจน์ (2566) ได้กล่าวว่าอัลกอรึทึม ต้นไม้ตัดสินใจแบบจำแนกและแบบถดถอย (Classification And Regression Tree : CART) เป็นอัลกอริทึมที่แยกชุดฝึกออกเป็นสองส่วนย่อย โดยใช้ฟีเจอร์เดียว เมื่อแยกชุดฝึกออกเป็นสองชุดสำเร็จ ก็จะแยกชุดย่อยนั้นออกไปเป็นสองโหนดย่อย โดยจะหยุดวนซ้ำเมื่อถึงความลึกสูงสุด (max_depth)หรือหากไม่พบการแยกใดๆที่ช่วยลดค่าความผิดพลาดได้อีก

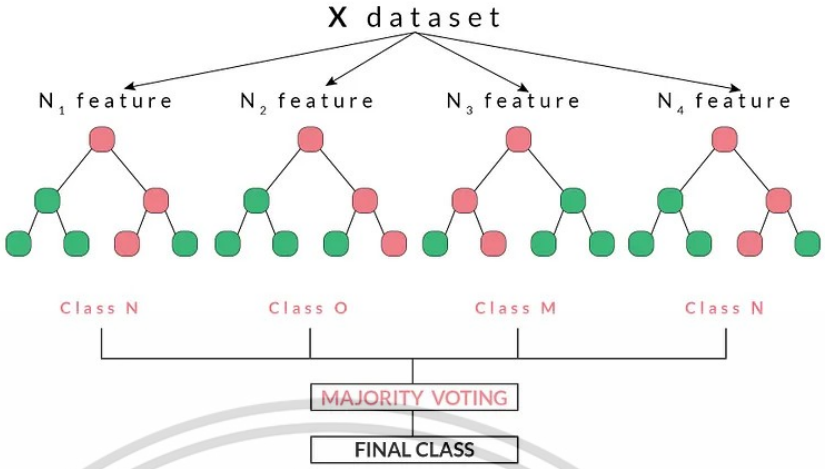


รูปที่ 2.2 ตัวอย่างแผนผังการตัดสินใจ CART
(ที่มา : Witchapong Daroontham,2561)

2.1.6 ป่าสุ่ม (Random Forest)

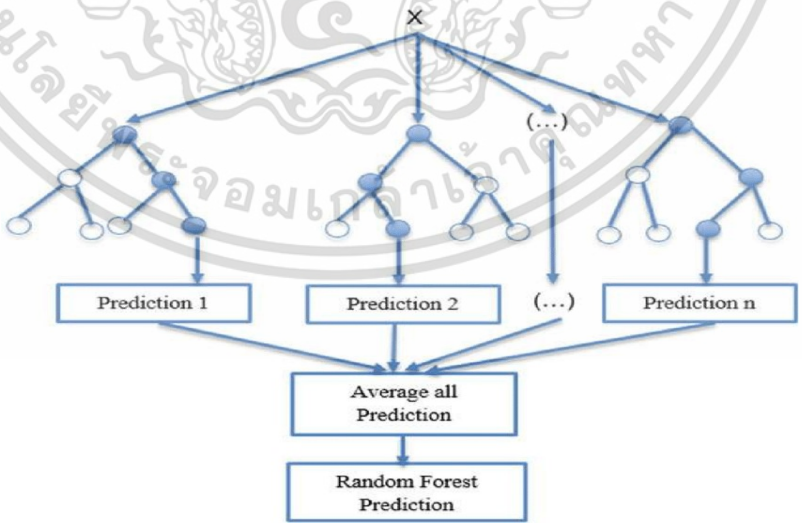
ป่าสุ่มเป็นวิธีการเรียนรู้ของเครื่อง (Machine Learning :ML) รูปแบบหนึ่งที่พัฒนาโดยมีพื้นฐานมาจากต้นไม้ตัดสินใจ (Decision Tree) โดยเทคนิคป่าสุ่ม (Random Forest) จะเป็นการเรียนรู้แบบกลุ่มโมเดล (Ensemble Learning) เป็นการนำต้นไม้ตัดสินใจมาทำการฝึกร่วมกัน โดยที่แต่ละต้นไม้จะได้รับตัวแปรอิสระแบบสุ่ม และผลลัพธ์ของแต่ละต้นไม้ตัดสินใจจะได้รับการโหวต (Vote) ในการจำแนกประเภทและใช้ค่าเฉลี่ยในการถดถอย เพื่อเป็นผลลัพธ์สุดท้ายที่ออกมาจาก ตัวแบบป่าสุ่ม (Random Forest)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 ตัวอย่างการทำงาน Random Forest Classification
(ที่มา Sasiwut Chaiyadecha, 2563)

วิโรจน์ (2566) ได้กล่าวว่าการสุ่มหาคำตอบจากคนหลายพันคน แล้วรวมคำตอบในลักษณะ “ปัญญามหาชน” ย่อมดีกว่าคำตอบจากผู้เชี่ยวชาญคนเดียวในทำนองเดียวกันหากรวมกลุ่มตัวทำนายย่อมได้ผลดีกว่าตัวทำนายตัวเดียว กลุ่มตัวทำนายนี้เรียกว่ากลุ่มโมเดลโดยใช้อัลกอริทึม วิธีการกลุ่มโมเดล (Ensemble Method) โดยกลุ่มโมเดลที่ทำการฝึกด้วย Decision Tree บนชุดข้อมูลย่อยที่สุ่มด้วยวิธีสุ่มแล้วคืนตัวอย่าง (Bagging) จากชุดฝึกรวม เมื่อได้รับผลลัพธ์จากแต่ละ Decision Tree แล้วคลาสที่ได้รับการโหวตมากที่สุดในการจำแนกประเภทและค่าเฉลี่ยที่ได้จากแต่ละ Decision Tree ในการถดถอยจะเป็นผลลัพธ์สุดท้าย กลุ่มโมเดลนี้ถูกเรียกว่า Random Forest

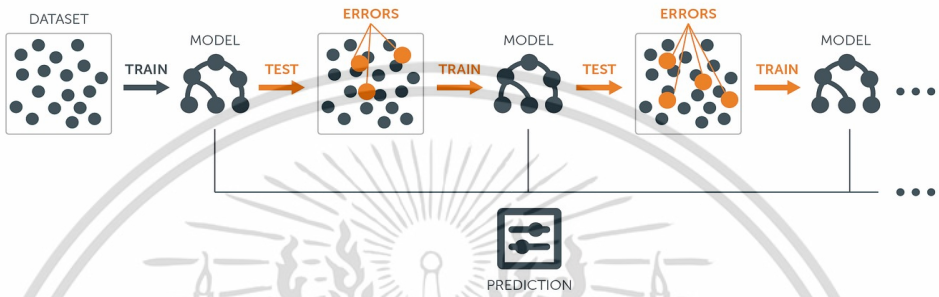


รูปที่ 2.4 ตัวอย่างการทำงาน Random Forest Regression
(ที่มา Md. Mahbubur Rahman ,2565)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.7 eXtreme Gradient Boosting (XGboost)

XGboost เป็นตัวแบบที่พัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) คล้ายกันกับตัวแบบป่าสุ่ม (Random Forest) แตกต่างกันที่ XGboost จะใช้วิธีในการหาคำตอบโดยเรียนรู้จากความผิดพลาดของต้นไม้ก่อนหน้าและปรับปรุงเพื่อสร้างต้นไม้ต้นถัดไปและจะหยุดเมื่อถึงความลึกที่กำหนดหรือไม่เหลือความผิดพลาดจากต้นไม้ก่อนหน้าแล้ว โดยการเรียนรู้จากต้นไม้ก่อนหน้าเรียกว่าเทคนิค Boosting



รูปที่ 2.5 ตัวอย่างการทำงานของ eXtreme Gradient Boosting (XGboost)
(ที่มา Nut Chukamphaeng, 2561)

2.1.8 การปรับไฮเปอร์พารามิเตอร์

ในการสร้างตัวแบบการเรียนรู้ของเครื่อง นอกจากการนำข้อมูลเพื่อให้ตัวแบบเรียนรู้ยังมีการปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบเพื่อให้ตัวแบบทำงานได้มีประสิทธิภาพมากขึ้น (วิโรจน์, 2566) เนื่องจากไม่มีเกณฑ์ใดในการกำหนดค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด ผู้วิจัยจึงต้องทำการทดลองหาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมในการทำนายมากที่สุด โดยในงานวิจัยนี้ผู้วิจัยใช้การค้นหาค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดด้วย GridSearch โดยใช้ GridSearchCV โดยกำหนดค่าไฮเปอร์พารามิเตอร์ที่ต้องการทดสอบจากนั้นทำการเลือกค่าไฮเปอร์พารามิเตอร์ที่ให้ผลดีที่สุดมาใช้กับตัวแบบการเรียนรู้ของเครื่อง

2.1.9 ตัววัดประสิทธิภาพของตัวแบบ

สมศรี (2559) กล่าวว่าความถูกต้องของการพยากรณ์เป็นสิ่งที่ผู้ใช้ค่าพยากรณ์ต้องการ โดยความถูกต้องมากน้อยนั้นแสดงผลออกมาด้วยค่าความคลาดเคลื่อนของการพยากรณ์ (Error) ซึ่งเป็นผลต่างของค่าจริง และค่าพยากรณ์ถ้าความคลาดเคลื่อนมากหมายความว่าค่าจริงห่างจากค่าพยากรณ์มาก นั้นหมายความว่าการพยากรณ์นั้นไม่เหมาะสมกับข้อมูล

โดยในงานวิจัยนี้ผู้วิจัยใช้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error : RMSE) เนื่องจากมีหน่วยเดียวกับค่าที่ต้องการพยากรณ์โดยมีสมการในการคำนวณดังสมการที่ (2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{n} e_t^{\,2}}{n}} \qquad (2)$$

โดย

e_t แทนค่า ความคลาดเคลื่อน ณ เวลา t (ค่าจริง - ค่าพยากรณ์)

n แทนค่า จำนวนข้อมูล

2.2 งานวิจัยที่เกี่ยวข้อง

พรวิทา (2564) ศึกษาตัวแบบในการพยากรณ์ธุรกิจค้าปลีกโดยศึกษา 3 กรณีศึกษาคือ 1.ข้อมูลยอดขายรายวันของร้านกาแฟ 2.ข้อมูลจำนวนผู้ใช้บริการรถไฟฟ้ามหานคร สายฉลองรัชธรรม 3.ข้อมูลผู้ป่วยและผู้เสียชีวิตรายใหม่จากสถานการณ์โควิด-19ในประเทศไทย โดยใช้เทคนิค ARIMA , SARIMA การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) ต้นไม้ตัดสินใจ(Decision Tree) และป่าสุ่ม (Random Forest) โดยทำการแปลงข้อมูลในกรณีศึกษาที่ 1 และ 2 ให้เป็นข้อมูลตัวเลข เนื่องจากเป็นตัวแปรเชิงคุณภาพ ในส่วนกรณีศึกษาที่ 2 เป็นข้อมูลตัวเลขอยู่แล้วจึงไม่ทำการแปลงข้อมูล จากนั้นนำข้อมูลของกรณีศึกษาทั้ง 3 กรณีศึกษาเข้าตัวแบบทั้ง 5 ตัวแบบข้างต้น โดยที่กรณีศึกษาที่ 1 นำเข้าทั้ง 5 ตัวแบบ กรณีศึกษาที่ 2 นำเข้าตัวแบบ ARIMA และ SARIMA กรณีศึกษาที่ 3 นำเข้าตัวแบบ การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) ต้นไม้ตัดสินใจ(Decision Tree) และป่าสุ่ม (Random Forest) พบว่าข้อมูลที่มีอิทธิพลของฤดูกาลเข้ามาเกี่ยวเช่นกรณีศึกษาที่ 2 ตัวแบบ SARIMA ให้ผลดีกว่า ARIMA โดยมีค่า MAPE เท่ากับ 25.90% และ 35.15% ตามลำดับ ส่วนการเปรียบเทียบตัวแบบการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) ต้นไม้ตัดสินใจ(Decision Tree) และป่าสุ่ม (Random Forest) สำหรับกรณีศึกษาที่ 1 และ 3 พบว่าตัวแบบRandom Forest ให้ค่า MAPE ต่ำที่สุดเท่ากับ 18.74%-54.15% เมื่อเทียบกับตัวแบบการวิเคราะห์การถดถอยพหุคูณ (Multiple Regression) และต้นไม้ตัดสินใจ(Decision Tree)

Yang (2023) ศึกษาตัวแบบพยากรณ์ด้วย OLS, Random Forest, XGBoost สำหรับการพยากรณ์ยอดขาย Walmart โดยใช้ค่า MSE, MAE, R^2 โดยกำหนดไฮเปอร์พารามิเตอร์ไว้แล้ว พบว่าตัวแบบ XGBoost ให้ผลลัพธ์ของทั้งสามตัววัดดีที่สุด คือ 0.0655, 0.1246, 0.9838 ตามลำดับ รองลงมาคือ Random Forest 0.0887, 0.1432, 0.9781 และอันดับสุดท้ายคือ OLS

ชยากร (2564) ศึกษาตัวแบบพยากรณ์ด้วย TBATS, Multiple linear regression (MLR), XGBoost, Artificial Neural Network (ANN) และตัวแบบผสม สำหรับการพยากรณ์ยอดขายรายวันของสินค้าในธุรกิจค้าปลีก โดยพิจารณาการพยากรณ์ 4 รูปแบบคือ การพยากรณ์ยอดขายโดยตรงกับตัวแปรอิสระปกติ การพยากรณ์ลอการิทึมธรรมชาติของยอดขายกับตัวแปรอิสระที่เพิ่มข้อมูลยอดขายตัวเองในอดีต การพยากรณ์ลอการิทึมธรรมชาติของยอดขายกับตัวแปรอิสระ และการพยากรณ์ยอดขายโดยตรงกับตัวแปรอิสระที่เพิ่มข้อมูลยอดขายตัวเองในอดีต โดยในตัวแบบ XGBoost, Artificial Neural Network (ANN) และตัวแบบผสม ใช้ตัวแปรอิสระสำหรับให้ตัวแบบเรียนรู้จาก Multiple linear regression (MLR) และค้นหาค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดด้วย GridSearch โดยค่าที่ใช้ในการปรับคือ learning_rate 0.05, 0.1, 0.3 max_depth 3, 6, 10 และ n_estimators 10, 100, 1000 พบว่าตัวแบบที่ให้ค่า MAPE ต่ำที่สุดที่ 16.04% คือตัวแบบ Artificial Neural Network (ANN)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชัยพฤกษ์ (2564) ศึกษาตัวแบบพยากรณ์ด้วย Holt's Exponential Smoothing, ARIMA, Multiple Regression, Artificial Neural Network, eXtreme Gradient Boosting, Random Forest, Support Vector Regression และ Light Gradient Boosting Machine และตัวแบบผสมในการพยากรณ์ปริมาณความต้องการเหล็กรีดร้อนภายในประเทศ โดยสำหรับตัวแบบการเรียนรู้ของเครื่องใช้ตัวแปรที่ผ่านการคัดเลือกจาก Multiple linear Regression และค้นหาไฮเปอร์พารามิเตอร์ที่ดีที่สุดด้วยการค้นหาแบบกริด พบว่าตัวแบบที่มีความแม่นยำมากที่สุดคือตัวแบบผสม Artificial Neural Network, Random Forest และ eXtreme Gradient Boosting โดยใช้ตัวแบบ Artificial Neural Network เป็นตัวผสม มีค่า MAPE เท่ากับ 3.59%

ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงแบ่งตัวแปรอิสระที่ใช้ออกเป็น 3 รูปแบบโดย รูปแบบที่ 1 คือใช้ตังแปรอิสระทั้งหมดที่มี รูปแบบที่ 2 ใช้ Multiple linear Regression ด้วย OLS ในไลบารี่ statsmodels.api เพื่อเลือกตัวแปรที่มีผลกับตัวแปรตามอย่างมีนัยสำคัญของแต่ละเมนู รูปแบบที่ 3 ทำการแปลงลอการิทึมธรรมชาติของยอดขาย จากนั้นใช้ Multiple linear Regression ด้วย OLS ในไลบารี่ statsmodels.api เพื่อเลือกตัวแปรที่มีผลกับตัวแปรตามอย่างมีนัยสำคัญของแต่ละเมนู และนำตัวแปรที่ได้จากทั้ง 3 รูปแบบนำเข้าตัวแบบต้นไม้ตัดสินใจ (Decision Tree) ป่าสุ่ม (Random Forest) และeXtreme Gradient Boosting (XGBoost) โดยตัวแบบการเรียนรู้ของเครื่องแต่ละตัวแบบใช้การค้นหาค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดด้วย GridSearch โดยใช้ค่าในการปรับจากงานวิจัยที่เกี่ยวข้องคือค่า learning_rate, max_depth และ n_estimators จากนั้นเปรียบเทียบประสิทธิภาพของตัวแบบแต่ละตัวด้วยค่าเฉลี่ยค่ารากที่สองของเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error : RMSE) ของทุกเมนูที่มีการขายมากกว่า 60 วันและเลือกตัวแบบที่มีค่าเฉลี่ยค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error : RMSE) ของทุกเมนูที่มีการขายมากกว่า 60 วัน น้อยที่สุดแล้วจึงนำไปทำนายยอดขายใน 7 วันข้างหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินงานวิจัย

งานวิจัยนี้มีจุดประสงค์เพื่อหาตัวแบบที่เหมาะสมในการทำนายจำนวนการขายเบเกอรี่ในสาขาที่ผู้วิจัยทำการศึกษา โดยมีรายละเอียดการดำเนินการดังที่จะกล่าวต่อไปนี้

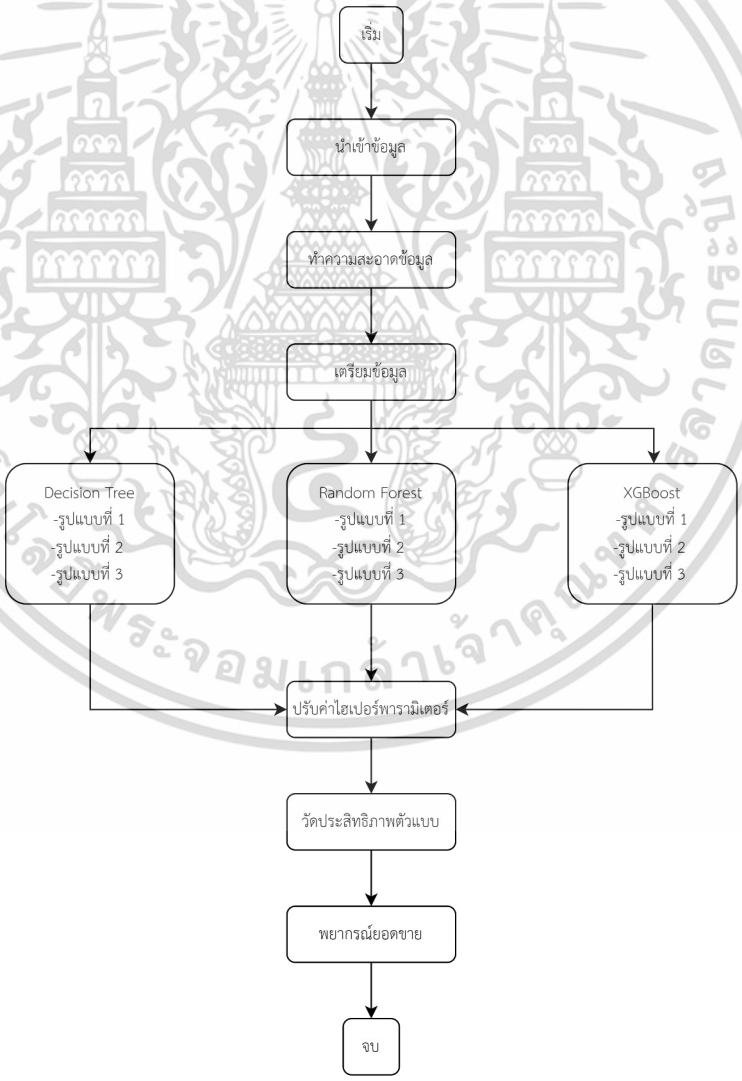
3.1 ขั้นตอนการดำเนินงาน

3.2 ตัวแปรที่ใช้ในงานวิจัย

3.3 เครื่องมือที่ใช้ในงานวิจัย

3.1 ขั้นตอนการดำเนินงาน

การวิจัยเพื่อพยากรณ์จำนวนการขายเบเกอรี่รายวัน มีขั้นตอนการดำเนินงานดังรูปที่ 3.1



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3.1 ขั้นตอนการดำเนินงาน

3.1.1 ลักษณะข้อมูล

ข้อมูลที่ใช้ในตัวแบบการพยากรณ์สินค้าเบเกอรี่ในสาขาที่ผู้วิจัยทำการศึกษาเป็นข้อมูลทุติยภูมิที่ทำการรวบรวมการขายแต่ละสินค้า โดยเก็บข้อมูลการขายเป็นรายวันเริ่มจากวันที่ 1 มกราคม 2566 ถึงวันที่ 21 กุมภาพันธ์ 2567

PR_CODE	ITEM_LONG_ENG	date	day_of_week	month	year	day_type	PromotionType	day_of_month	QUANT	Prev_Day_QUANT
		2023-01-01	Sunday	January	2023	holiday	A	1		
		2023-01-02	Monday	January	2023	normal	A	2		
		2023-01-03	Tuesday	January	2023	normal	A	3		
		2023-01-04	Wednesday	January	2023	normal	A	4		
		2023-01-05	Thursday	January	2023	normal	A	5		
		...	...	...	...	...	...	...		
		2024-02-20	Tuesday	February	2024	normal	A	20		
		2024-02-21	Wednesday	February	2024	normal	A	21		
		2023-04-02	Sunday	April	2023	normal	A	2		
		2023-09-19	Tuesday	September	2023	normal	A	19		
		2023-02-26	Sunday	February	2023	normal	A	26		

รูปที่ 3.2 ข้อมูลการขายสินค้ารายวันของแต่ละเมนู

3.1.2 การทำความสะอาดข้อมูล

ผู้วิจัยได้ทำความสะอาดข้อมูลโดยทำการตรวจสอบข้อมูลสูญหายและทำการตัดข้อมูลสูญหายออกเพื่อเพิ่มประสิทธิภาพในการฝึกสอนตัวแบบ โดยใช้คำสั่ง .isna().sum() ในการตรวจสอบค่าสูญหาย

```
1  df_vis.isna().sum()
✓ 0.0s
PR_CODE           0
ITEM_LONG_ENG     0
date              0
day_of_week       0
month             0
year              0
day_type          0
PromotionType     0
day_of_month      0
QUANT             0
Prev_Day_QUANT    0
```

รูปที่ 3.3 ค่าสูญหายของแต่ละตัวแปร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.3 แสดงให้เห็นว่าไม่มีข้อมูลสูญหายอยู่ในข้อมูลยอดขายสินค้ารายวันของแต่ละเมนู จึงทำการเตรียมข้อมูลเพื่อนำข้อมูลเข้าตัวแบบโดยมีรายละเอียดดังที่จะกล่าวในหัวข้อถัดไป

3.1.3 การเตรียมข้อมูล

ในขั้นตอนการเตรียมข้อมูลผู้วิจัยใช้ฟังก์ชัน One-hot Encoder ในไลบารี่ Scikit-learn ในการแปลงข้อมูล และแบ่งข้อมูลเป็นชุดฝึกสอนและชุดทดสอบ โดยมีรายละเอียดดังนี้

3.1.3.1 การแปลงข้อมูล

ผู้วิจัยทำการแปลงข้อมูลเชิงคุณภาพเป็นข้อมูลเชิงปริมาณเพื่อให้ใช้ในการสร้างตัวแบบโดยตัวแปรที่ทำการแปลงข้อมูลคือ วันในสัปดาห์ เดือน ปี วันหยุด ประเภทของวัน และประเภทของโปรโมชั่น โดยใช้ฟังก์ชัน One-hot Encoder ในไลบารี่ Scikit-learn ในการแปลงข้อมูล

date	day_of_week_Friday	day_of_week_Monday	day_of_week_Saturday	day_of_week_Sunday	day_of_week_Thursday	day_of_week_Tuesday	day_of_week_Wednesday
2023-01-01	0.0	0.0	0.0	1.0	0.0	0.0	0.0
2023-01-01	0.0	0.0	0.0	1.0	0.0	0.0	0.0
2023-01-02	0.0	1.0	0.0	0.0	0.0	0.0	0.0
2023-01-02	0.0	1.0	0.0	0.0	0.0	0.0	0.0
2023-01-03	0.0	0.0	0.0	0.0	0.0	1.0	0.0
...	...	...	...	...	...	...	...
2024-02-17	0.0	0.0	1.0	0.0	0.0	0.0	0.0
2024-02-18	0.0	0.0	0.0	1.0	0.0	0.0	0.0
2024-02-19	0.0	1.0	0.0	0.0	0.0	0.0	0.0
2024-02-20	0.0	0.0	0.0	0.0	0.0	1.0	0.0
2024-02-21	0.0	0.0	0.0	0.0	0.0	0.0	1.0

รูปที่ 3.4 ตัวอย่างการใช้ One-hot Encoder ในการแปลงข้อมูลให้อยู่ในรูปตัวเลข

3.1.3.2 การแบ่งข้อมูล

ผู้วิจัยได้ทำการแบ่งข้อมูลเป็นชุดข้อมูลสำหรับสร้างตัวแบบพยากรณ์ 80% ของข้อมูลทั้งหมด และชุดข้อมูลสำหรับทดสอบประสิทธิภาพของตัวแบบ 20% ของข้อมูลทั้งหมด

3.1.4 คัดเลือกตัวแปร

ผู้วิจัยทำการแบ่งการคัดเลือกตัวแปรออกเป็น 3 รูปแบบโดย รูปแบบที่ 1 จะใช้ตัวแปรอิสระทั้งหมดที่มี รูปแบบที่ 2 ผู้วิจัยใช้ OLS ในไลบารี่ statsmodels.api เพื่อสร้างสมการถดถอยเชิงเส้นพหุคูณ จากนั้นคัดเลือกตัวแปรที่มีค่า P-value ในการทดสอบ t-test น้อยกว่า 0.05 รูปแบบที่ 3 ทำการแปลงข้อมูลยอดขายด้วยค่าลอการิทึมธรรมชาติจากนั้นทำขั้นตอนเดียวกับขั้นตอนที่ 2 โดยตัวแปรที่ผ่านการคัดเลือกจากรูปแบบที่ 2 และ 3 แสดงในตารางที่ 3.1 และ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2

menu	ศ.	จ.	ส.	อา.	อ.	พฤ.	พ.	เม.ย	ส.ค	ธ.ค	ก.พ	ม.ค	ก.ค	มิ.ย	มี.ค
0			X	X							X	X			
1				X								X			
2			X	X											
3			X	X	X			X	X		X	X			X
4			X	X							X	X			
5			X	X							X		X		
6			X	X					X	X	X	X	X		
7				X								X	X		
8							X						X		X
9			X	X											
10	X		X	X											
11				X		X		X				X			X
12			X	X											
13			X	X		X									
14			X	X			X								
15			X				X				X				
16		X							X	X					X
17						X									
18	X		X	X	X	X	X			X					
19	X	X	X	X	X	X	X			X	X	X	X		X
20			X	X											
21			X	X											
22			X								X				
23	X	X	X	X	X					X					
24	X		X	X	X				X						
25			X								X	X	X		
26			X												
27			X		X				X						

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2 (ต่อ)

menu	ศ.	จ.	ส.	อา.	อ.	พฤ.	พ.	เม.ย	ส.ค	ธ.ค	ก.พ	ม.ค	ก.ค	มิ.ย	มี.ค
28	X	X	X								X				
29			X	X		X									
30	X													X	
31					X	X					X				
32			X	X											
33			X			X									
34				X											
35			X	X							X				
36				X											
37			X	X							X				
38			X	X				X		X		X	X	X	
39											X				
40			X	X					X	X		X	X	X	X
41	X		X	X						X		X			
42			X	X								X		X	
43			X	X						X	X	X			X
44			X	X				X		X	X			X	X
45			X	X					X			X		X	X
46		X						X			X				
47				X											
48				X			X								
49				X								X			
50			X											X	
51			X	X	X						X				
52			X	X										X	
53	X		X			X									
54				X							X				
55			X	X							X		X		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2 (ต่อ)

menu	ศ.	จ.	ส.	อา.	อ.	พฤ.	พ.	เม.ย	ส.ค	ธ.ค	ก.พ	ม.ค	ก.ค	มิ.ย	มี.ค
56			X	X							X				
57								X							
58				X											
59		X		X						X					
60											X	X			
61												X			
62			X	X										X	
63										X					
64						X					X	X			
65			X												
66			X												
67		X													
68											X				
69			X		X						X	X			
70				X											
71			X	X						X		X			
72			X	X				X			X	X	X		X
73		X							X		X		X		X
74			X						X			X		X	
75													X		
76													X		
77			X						X	X			X		
78													X		
79					X	X			X			X			
80													X		
81						X									
82												X	X		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2 (ต่อ)

menu	ศ.	จ.	ส.	อา.	อ.	พฤ.	พ.	เม.ย	ส.ค	ธ.ค	ก.พ	ม.ค	ก.ค	มิ.ย	มี.ค
83			X			X									
84			X	X						X					
85			X								X				
86			X	X		X									
87				X									X		
88			X	X				X	X	X	X	X	X	X	X

ตารางที่ 3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2 (ต่อ)

menu	พ.ค.	พ.ย.	ต.ค.	ก.ย.	2023	2024	festival	holiday	normal	ProA	ProB	วันของเดือน	Prev_Quant	ProC	ProD
0						X	X			X	X		X		
1					X					X					
2				X	X	X	X			X					
3				X	X	X				X			X		
4					X	X				X					
5					X	X			X	X			X	X	
6					X			X						X	
7					X	X				X				X	
8						X				X					
9		X			X	X								X	
10	X				X	X			X	X					
11				X	X								X		
12					X	X		X						X	
13					X							X	X	X	
14		X	X	X	X	X			X	X					
15		X			X	X				X	X		X		
16					X	X				X					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2 (ต่อ)

menu	พ.ค.	พ.ย.	ต.ค.	ก.ย.	2023	2024	festival	holiday	normal	Pro A	Pro B	วันของเดือน	Prev_ Quant	Pro C	Pro D
17					X	X		X		X					
18			X		X	X	X	X	X	X	X	X		X	
19					X	X	X	X	X	X	X		X		
20			X		X	X			X	X			X	X	
21	X				X	X		X		X					
22					X	X				X					
23	X			X	X	X	X	X	X	X	X		X	X	
24				X	X	X	X	X	X	X	X		X	X	
25					X	X		X		X					
26					X					X			X		
27					X	X			X	X					
28					X	X	X			X					
29					X	X			X	X			X		
30	X				X	X	X			X					
31					X	X				X			X		
32					X	X		X		X					
33										X		X			
34				X	X	X			X	X					
35				X	X	X		X	X	X					
36						X			X	X					
37			X		X	X		X		X					
38	X		X	X	X	X				X		X			
39									X		X				
40		X	X	X	X	X	X	X	X	X		X		X	
41		X			X	X		X	X	X					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2 (ต่อ)

menu	พ.ค.	พ.ย.	ต.ค.	ก.ย.	2023	2024	festival	holiday	normal	Pro A	Pro B	วันของเดือน	Prev_ Quant	Pro C	Pro D
42					X				X	X		X			
43		X			X					X					
44	X	X	X		X			X					X	X	
45				X	X	X	X	X	X	X					
46					X	X		X		X			X		
47					X	X		X						X	
48	X					X									
49					X			X		X					
50		X		X	X	X				X					
51					X	X				X					
52					X	X				X					
53		X				X				X		X			
54					X	X				X			X		
55		X			X	X	X			X					
56					X	X	X			X					
57					X	X				X					
58					X					X			X		
59			X		X	X				X					
60					X										
61					X	X								X	
62					X	X								X	
63					X					X					
64						X		X	X	X	X		X		
65						X		X		X					
66					X	X								X	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 2 (ต่อ)

menu	พ.ค.	พ.ย.	ต.ค.	ก.ย.	2023	2024	festival	holiday	normal	Pro A	Pro B	วันของเดือน	Prev_ Quant	Pro C	Pro D
67					X		X			X					
68	X				X	X		X						X	
69					X	X		X	X	X	X		X	X	
70					X	X		X		X					
71						X			X				X	X	
72	X				X										X
73		X	X		X	X								X	
74		X	X	X	X	X								X	
75			X		X	X								X	
76					X	X								X	
77					X	X		X		X				X	
78				X	X										
79		X			X	X		X		X			X		
80					X	X				X					
81					X	X		X		X		X			
82					X	X	X							X	
83					X		X			X					
84		X	X	X	X	X	X			X	X		X		
85						X			X	X	X		X		
86					X	X		X	X	X	X		X		
87		X	X		X	X		X	X	X	X	X		X	
88	X	X	X	X	X	X		X	X	X			X		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3

menu	ศ.	จ.	ส.	อา.	อ.	พฤ.	พ.	เม.ย	ส.ค	ธ.ค	ก.พ	ม.ค	ก.ค	มิ.ย	มี.ค
0			X	X							X				X
1												X	X		
2			X	X											
3	X		X	X				X			X	X			X
4			X	X						X	X	X			
5			X	X							X		X		
6			X	X					X		X	X	X		X
7				X								X			
8													X		X
9			X	X					X						
10	X		X	X		X									
11				X		X		X	X			X			
12			X												
13	X		X	X		X									
14			X	X			X								
15			X	X							X			X	
16		X	X						X	X			X		
17						X									
18			X	X			X			X			X	X	
19			X	X					X						
20			X	X											
21			X	X											
22			X												
23		X	X	X				X		X		X			
24	X														
25													X		
26			X												

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3 (ต่อ)

menu	ศ.	จ.	ส.	อา.	อ.	พฤ.	พ.	เม.ย	ส.ค	ธ.ค	ก.พ	ม.ค	ก.ค	มิ.ย	มี.ค
27			X						X						
28	X	X	X			X					X				
29			X	X		X									
30	X			X										X	
31					X	X					X				
32				X								X			
33			X			X		X							
34				X											
35			X	X			X	X							
36				X											
37			X	X											
38	X	X	X	X		X		X				X	X	X	
39									X		X				X
40	X	X	X	X	X	X	X		X	X	X	X	X		
41			X	X						X					
42			X	X								X		X	
43			X	X						X	X	X			X
44			X	X				X		X			X	X	X
45	X		X	X	X	X			X	X			X	X	
46		X						X			X				
47				X											
48				X			X								
49				X								X			
50			X			X				X				X	
51			X	X	X						X				
52	X		X	X										X	
53						X									X

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3 (ต่อ)

menu	ศ.	จ.	ส.	อา.	อ.	พฤ.	พ.	เม.ย	ส.ค	ธ.ค	ก.พ	ม.ค	ก.ค	มิ.ย	มี.ค
54				X	X						X		X		
55				X							X		X		
56			X	X			X				X				
57															
58				X											
59		X								X					
60												X			
61															
62			X												
63													X		
64			X	X	X			X			X		X		
65			X												
66			X												
67		X													
68											X				
69			X	X				X			X				
70				X											
71			X	X						X					
72			X	X				X			X		X		X
73		X							X				X		X
74			X						X					X	
75															
76													X		
77										X			X		
78													X		
79					X	X									
80													X		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3 (ต่อ)

menu	ศ.	จ.	ส.	อา.	อ.	พฤ.	พ.	เม.ย	ส.ค	ธ.ค	ก.พ	ม.ค	ก.ค	มิ.ย	มี.ค
81	X					X									
82												X	X		
83			X			X					X				
84			X	X						X					
85			X	X							X	X			
86			X	X	X						X	X	X		
87			X	X					X						X
88			X	X				X	X		X	X	X	X	X

ตารางที่ 3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3 (ต่อ)

menu	พ.ค.	พ.ย.	ต.ค.	ก.ย.	2023	2024	festival	holiday	normal	Pro A	Pro B	วันของเดือน	Prev_ Quant	Pro C	Pro D
0				X	X	X				X	X		X		
1					X					X					
2				X	X	X			X	X					
3		X			X				X	X			X		
4					X	X	X			X					
5					X	X			X					X	
6				X	X	X		X		X				X	
7															
8						X				X					
9		X			X	X								X	
10	X				X	X		X	X	X					
11	X			X	X										
12					X	X		X							
13					X	X		X		X			X	X	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3 (ต่อ)

menu	พ.ค.	พ.ย.	ต.ค.	ก.ย.	2023	2024	festival	holiday	normal	Pro A	Pro B	วันของเดือน	Prev_ Quant	Pro C	Pro D
14		X	X	X	X	X			X	X					
15					X	X	X		X	X	X		X		
16		X			X	X		X	X	X					
17					X	X				X					
18		X	X		X	X	X	X	X	X	X			X	
19					X	X			X	X	X				
20			X		X	X		X	X	X			X	X	
21	X					X				X					
22					X	X				X					
23		X		X	X	X	X		X	X	X			X	
24						X	X		X	X	X		X		
25					X					X					
26					X					X			X		
27					X					X					
28					X					X					
29						X				X			X		
30	X				X					X					
31					X					X			X		
32								X							
33												X			
34				X						X					
35					X	X				X					
36						X				X			X		
37	X		X		X	X		X	X	X					
38	X		X		X		X	X	X	X		X			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3 (ต่อ)

menu	พ.ค.	พ.ย.	ต.ค.	ก.ย.	2023	2024	festival	holiday	normal	Pro A	Pro B	วันของเดือน	Prev_ Quant	Pro C	Pro D
39					X				X		X	X			
40	X	X	X	X	X	X	X	X	X	X		X		X	
41		X			X	X		X	X	X					
42				X	X	X		X	X	X		X			
43					X			X		X					
44	X	X	X		X			X		X		X	X	X	
45	X			X	X	X	X	X	X	X					
46					X	X		X		X			X		
47					X	X		X		X				X	
48	X					X									
49					X			X	X	X					
50		X			X	X	X			X					
51		X			X	X		X		X					
52					X	X			X	X					
53		X								X		X			
54						X				X			X		
55		X			X	X	X			X					
56					X	X				X					
57			X							X					
58					X	X				X			X		
59					X	X				X					
60					X										
61	X							X						X	
62					X										
63					X		X	X			X				

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ตัวแปรที่ผ่านการคัดเลือกตัวแปรรูปแบบที่ 3 (ต่อ)

menu	พ.ค.	พ.ย.	ต.ค.	ก.ย.	2023	2024	festival	holiday	normal	Pro A	Pro B	วันของเดือน	Prev_ Quant	Pro C	Pro D
64				X		X		X			X		X		
65						X		X		X					
66							X				X				
67					X					X					
68	X				X			X						X	
69		X				X		X		X	X		X		
70					X	X		X		X					
71		X			X	X			X	X			X	X	
72	X	X			X										X
73		X	X		X	X								X	
74		X	X	X	X	X				X				X	
75					X	X		X						X	
76					X	X								X	
77					X									X	
78				X											
79		X			X			X		X			X		
80				X			X		X					X	
81								X				X			
82					X	X	X							X	
83					X					X					
84		X	X	X	X	X	X		X	X	X		X		
85		X	X		X	X		X	X	X	X		X		
86				X	X	X	X	X	X	X	X				
87		X		X	X	X		X	X	X	X	X		X	
88		X			X			X		X			X		

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.5 ตัวแบบที่ใช้ในงานวิจัย

ตัวแบบที่ใช้ในงานวิจัยนี้คือตัวแบบต้นไม้ตัดสินใจ (Decision Tree) ตัวแบบป่าสุ่ม (Random Forest) และตัวแบบ eXtreme Gradient Boosting (XGBoost) โดยการปรับค่าไฮเปอร์พารามิเตอร์ในงานวิจัยนี้ได้มาจากงานวิจัยที่เกี่ยวข้อง (ชยากร,2564) โดยมีค่าไฮเปอร์พารามิเตอร์ที่สนใจศึกษาและมีรายละเอียดดังนี้

3.1.5.1 ต้นไม้ตัดสินใจ (Decision Tree)

เป็นตัวแบบที่พยากรณ์ตัวแปรตามด้วยการสร้างต้นไม้ตัดสินใจ โดยแยกโหนดออกตามคุณลักษณะของตัวแปรตามจนกว่าจะแยกโหนดไม่ได้ ผู้วิจัยได้ใช้ไลบารี่ Scikit-learn ด้วยคำสั่ง DecisionTreeRegressor ในการสร้างตัวแบบต้นไม้ตัดสินใจ (Decision Tree) โดยใช้อัลกอริทึมต้นไม้ตัดสินใจแบบจำแนกและแบบถดถอย (Classification And Regression Tree : CART) และมีการค้นหาค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดด้วย GridSearchCV โดยมีค่าในการค้นหา ดังตารางที่ 3.3

ตารางที่ 3.3 ไฮเปอร์พารามิเตอร์และค่าที่ใช้ในการปรับของตัวแบบต้นไม้ตัดสินใจ

ไฮเปอร์พารามิเตอร์	ค่าในการปรับ
max_depth	1, 3, 5, 10

โดยค่าไฮเปอร์พารามิเตอร์ max_depth เป็นตัวกำหนดระดับสูงสุดของโหนดที่มีในต้นไม้ตัดสินใจ หากไม่กำหนดค่าตัวแบบจะสร้างโหนดจนกว่าจะไม่สามารถแยกโหนดได้อีกหรือถึงค่าไฮเปอร์พารามิเตอร์ที่กำหนด โดยในงานวิจัยนี้ผู้วิจัยทำการค้นหาค่า max_depth ด้วย GridSearchCV โดยมีค่าในการปรับ คือ 1, 3, 5, 10

3.1.5.2 ป่าสุ่ม (Random Forest)

เป็นตัวแบบที่มีพัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) โดยสร้างต้นไม้ตัดสินใจหลายๆต้นจากข้อมูลที่มาจากการสุ่มแบบไม่ใส่คืนจากนั้นทำการรวมผลลัพธ์ของต้นไม้แต่ละต้นเพื่อหาผลลัพธ์สุดท้าย ผู้วิจัยใช้ไลบารี่ Scikit-learn ด้วยคำสั่ง RandomForestRegressor ในการสร้างตัวแบบป่าสุ่ม (Random Forest) และมีการค้นหาค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดด้วย GridSearchCV ดังตารางที่ 3.4

ตารางที่ 3.4 ไฮเปอร์พารามิเตอร์และค่าที่ใช้ในการปรับของตัวแบบป่าสุ่ม

ไฮเปอร์พารามิเตอร์	ค่าในการปรับ
n_estimators	100, 200 ,300 ,500
max_depth	3, 5, 7, 10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

max_depth เป็นตัวกำหนดระดับสูงสุดของโหนดที่มีในต้นไม้ตัดสินใจแต่ละต้น หากไม่กำหนดค่าตัวแบบจะสร้างโหนดจนกว่าจะไม่สามารถแยกโหนดได้อีกหรือถึงค่าไฮเปอร์พารามิเตอร์ที่กำหนด โดยในงานวิจัยนี้ผู้วิจัยทำการปรับค่า max_depth ด้วย GridSearchCV โดยมีค่าในการปรับ คือ 3, 5, 7, 10

n_estimators เป็นตัวกำหนดจำนวนต้นไม้ตัดสินใจที่ใช้ในตัวแบบว่าจะใช้ต้นไม้ตัดสินใจกี่ต้นโดยในงานวิจัยนี้ผู้วิจัยทำการปรับค่า n_estimators ด้วย GridSearchCV โดยมีค่าในการปรับ คือ 100, 200 ,300 ,500

3.1.5.3 eXtreme Gradient Boosting (XGboost)

เป็นตัวแบบที่พัฒนามาจากต้นไม้ตัดสินใจ (Decision Tree) เหมือนกับป่าสุ่ม (Random Forest) แต่ต่างกันที่หลักการทำงานของ Xgboost จะเรียนรู้จากความผิดพลาดของต้นไม้ก่อนหน้าเพื่อหาคำตอบที่ดีที่สุดในการทำนาย ผู้วิจัยใช้ไลบารี่ xgboost ด้วยคำสั่ง XGBRegressor ในการสร้างตัวแบบ Xgboost และมีการค้นหาค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดด้วย GridSearchCV โดยมีค่าในการปรับดังตารางที่ 3.5

ตารางที่ 3.5 ไฮเปอร์พารามิเตอร์และค่าที่ใช้ในการปรับของตัวแบบ Xgboost

ไฮเปอร์พารามิเตอร์	ค่าในการปรับ
n_estimators	100, 200 ,300 ,500
max_depth	3, 5, 7, 10
learning_rate	0.05, 0.1, 0.3

max_depth เป็นตัวกำหนดระดับสูงสุดของโหนดที่มีในต้นไม้ตัดสินใจแต่ละต้น หากไม่กำหนดค่าตัวแบบจะสร้างโหนดจนกว่าจะไม่สามารถแยกโหนดได้อีก ทำให้เกิดปัญหา โมเดลเรียนรู้ข้อมูลเฉพาะมากไป โดยในงานวิจัยนี้ผู้วิจัยทำการปรับค่า max_depth ด้วย GridSearchCV โดยมีค่าในการปรับ คือ 3, 5, 7, 10

n_estimators เป็นตัวกำหนดจำนวนต้นไม้ตัดสินใจที่ใช้ในตัวแบบว่าจะใช้ต้นไม้ตัดสินใจกี่ต้นโดยในงานวิจัยนี้ผู้วิจัยทำการปรับค่า n_estimators ด้วย GridSearchCV โดยมีค่าในการปรับ คือ 100, 200 ,300 ,500

learning_rate เป็นตัวกำหนดการเรียนรู้ของตัวแบบและการกำหนดยังช่วยป้องกันการ overfitting หรือโมเดลเรียนรู้ข้อมูลเฉพาะมากไป โดยควบคุมการปรับค่าน้ำหนักของตัวแบบในขั้นตอนการเรียนรู้ โดยค่าที่ใช้ในการปรับคือ 0.05, 0.1, 0.3

3.1.6 การวัดประสิทธิภาพของตัวแบบพยากรณ์

ผู้วิจัยใช้ค่ารากที่สองของค่าเฉลี่ยความผิดพลาดยกกำลังสอง (Root Mean Square Error: *RMSE*) ในการวัดประสิทธิภาพของตัวแบบที่ใช้ในการพยากรณ์โดยคัดเลือกตัวแบบจากตัวแบบที่มีค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความผิดพลาดยกกำลังสอง (Root Mean Square Error:

เอกสารนี้ ... ไม่ว่ากรณีใดๆ ...

RMSE) น้อยที่สุดในการทำนายข้อมูลทั้ง 89 เมนูและนำตัวแบบที่ได้มาใช้ในการพยากรณ์จำนวนยอดขายรายวัน 7 วันข้างหน้าของสินค้าทั้ง 89 เมนู

3.2 ตัวแปรที่ใช้ในงานวิจัย

ตัวแปรที่ใช้ในงานวิจัยนี้แบ่งเป็นตัวแปรอิสระจำนวน 30 ตัวแปร และตัวแปรตาม 1 ตัวแปร โดยรายละเอียดของตัวแปรที่ใช้แสดงดังตารางที่ 3.6 และ 3.7

3.2.1 ตัวแปรอิสระ

ตารางที่ 3.6 ตัวแปรอิสระที่ใช้ในงานวิจัย

ชื่อตัวแปร	คำอธิบาย	ชนิดของตัวแปร
day_of_week_Monday	วันของสัปดาห์ : วันจันทร์	ตัวแปรหุ่น
day_of_week_Tuesday	วันของสัปดาห์ : วันอังคาร	ตัวแปรหุ่น
day_of_week_Wednesday	วันของสัปดาห์ : วันพุธ	ตัวแปรหุ่น
day_of_week_Thursday	วันของสัปดาห์ : วันพฤหัสบดี	ตัวแปรหุ่น
day_of_week_Friday	วันของสัปดาห์ : วันศุกร์	ใตัวแปรหุ่น
day_of_week_Saturday	วันของสัปดาห์ : วันเสาร์	ตัวแปรหุ่น
day_of_week_Sunday	วันของสัปดาห์ : วันอาทิตย์	ตัวแปรหุ่น
month_January	เดือน : มกราคม	ตัวแปรหุ่น
month_February	เดือน : กุมภาพันธ์	ตัวแปรหุ่น
month_March	เดือน : มีนาคม	ตัวแปรหุ่น
month_April	เดือน : เมษายน	ตัวแปรหุ่น
month_May	เดือน : พฤษภาคม	ตัวแปรหุ่น
month_June	เดือน : มิถุนายน	ตัวแปรหุ่น
month_July	เดือน : กรกฎาคม	ตัวแปรหุ่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 ตัวแปรอิสระที่ใช้ในงานวิจัย (ต่อ)

ชื่อตัวแปร	คำอธิบาย	ชนิดของตัวแปร
month_August	เดือน : สิงหาคม	ตัวแปรหุ่น
month_September	เดือน : กันยายน	ตัวแปรหุ่น
month_October	เดือน : ตุลาคม	ตัวแปรหุ่น
month_November	เดือน : พฤศจิกายน	ตัวแปรหุ่น
month_December	เดือน : ธันวาคม	ตัวแปรหุ่น
year_2023	ปี : 2023	ตัวแปรหุ่น
year_2024	ปี : 2024	ตัวแปรหุ่น
day_type_festival	ประเภทของวัน : เทศกาลแต่ไม่ใช่วันหยุด	ตัวแปรหุ่น
day_type_holiday	ประเภทของวัน : วันหยุดราชกาล	ตัวแปรหุ่น
day_type_normal	ประเภทของวัน : วันธรรมดา	ตัวแปรหุ่น
day_of_month	จำนวนวันของเดือน	ตัวเลข
PromotionType_A	ประเภทของโปรโมชัน : A	ตัวแปรหุ่น
PromotionType_B	ประเภทของโปรโมชัน : B	ตัวแปรหุ่น
PromotionType_C	ประเภทของโปรโมชัน : C	ตัวแปรหุ่น
PromotionType_D	ประเภทของโปรโมชัน : D	ตัวแปรหุ่น
Prev_Day_QUANT	ยอดขายวันก่อนหน้า	ตัวเลข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

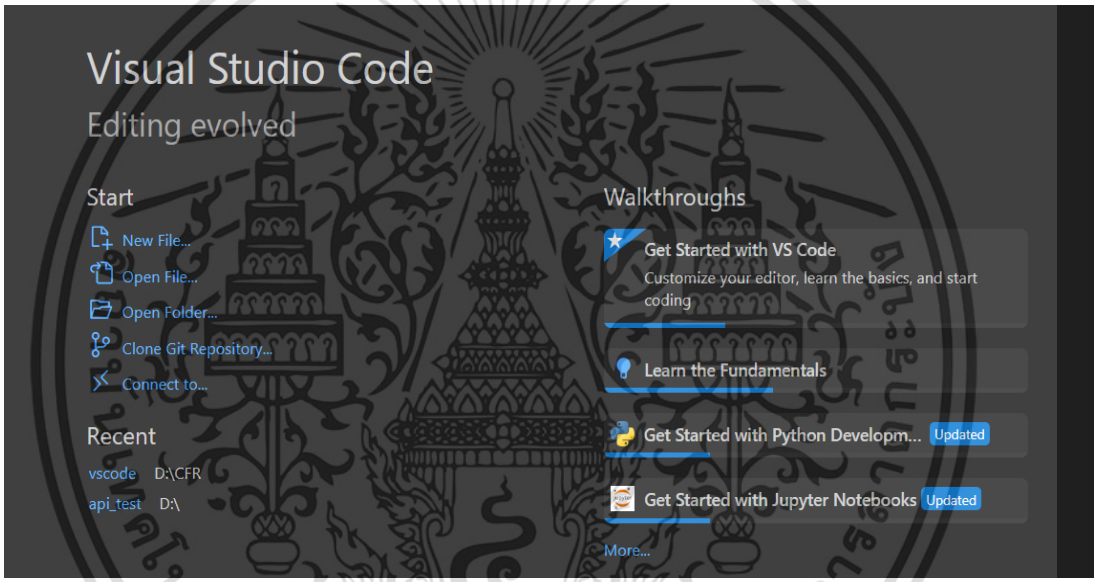
3.2.2 ตัวแปรตาม

ตารางที่ 3.7 อธิบายตัวแปรตามที่ศึกษา

ชื่อตัวแปร	คำอธิบาย	ชนิดของตัวแปร
QUANT	จำนวนสินค้าที่ขายรายวัน	ตัวเลข

3.3 เครื่องมือที่ใช้ในงานวิจัย

ในขั้นตอนการดำเนินงานวิจัยผู้วิจัยใช้ดำเนินการบนโปรแกรม Visual Studio Code ในการเขียนภาษา python โดยมีไลบารี่ที่ใช้ดังนี้



รูปที่ 3.5 หน้าแรกของ Visual Studio Code

- pandas ใช้ในการจัดการและวิเคราะห์ข้อมูลในรูปแบบตาราง
- numpy ใช้สำหรับการทำงานกับข้อมูลที่มีลักษณะ array หรือ matrix
- statsmodels ใช้สำหรับสร้างตัวแบบด้วย OLS
- sklearn ใช้สำหรับการแบ่งข้อมูล สร้างตัวแบบต้นไม้ตัดสินใจ (Decision Tree) ตัวแบบป่าสุ่ม (Random Forest) และใช้ในการวัดประสิทธิภาพของตัวแบบทั้ง 3 ตัวแบบ
- xgboost ใช้สำหรับการสร้างตัวแบบ XGBoost
- matplotlib ใช้สำหรับแสดงกราฟ
- plotly ใช้การสร้างและแสดงกราฟ

ในขั้นตอนการจัดทำรูปเล่มผู้วิจัยดำเนินการบนโปรแกรม Microsoft Excel ในการจัดเรียงข้อมูล และสร้างตารางแสดงผลต่างๆในงานวิจัยนี้

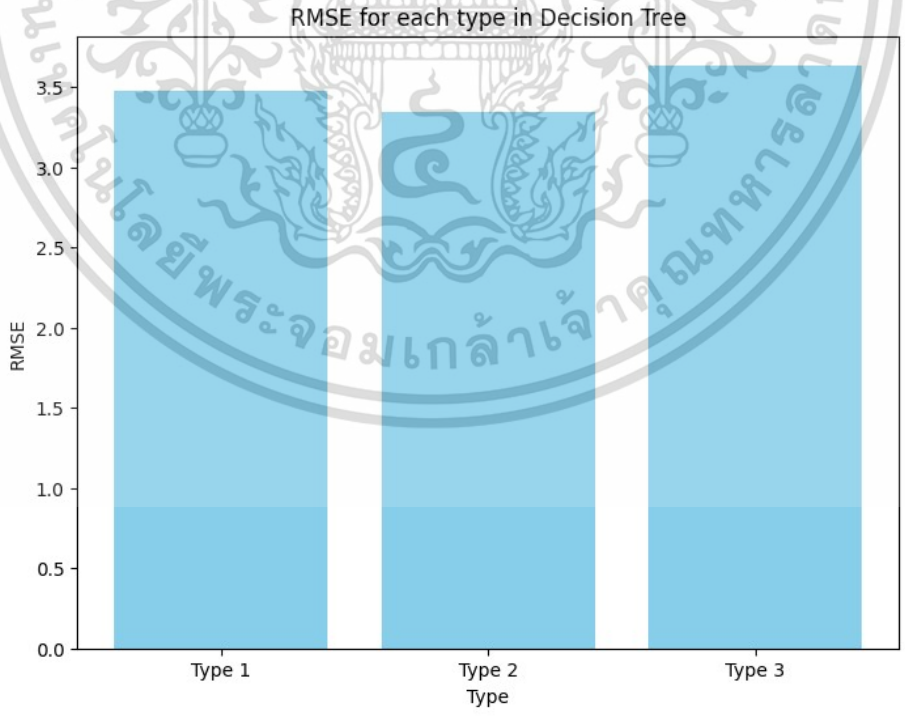
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัยและอภิปรายผล

ในบทนี้จะแสดงผลการวิจัยในการสร้างตัวแบบทั้ง 3 ตัวแบบนั้นคือ ตัวแบบต้นไม้ตัดสินใจ (Decision Tree) ตัวแบบป่าสุ่ม (Random Forest) และตัวแบบ XGboost จากข้อมูลทั้ง 3 รูปแบบ รูปแบบที่ 1 ใช้ตัวแปรอิสระทั้งหมดโดยใช้ข้อมูลยอดขายปกติ รูปแบบที่ 2 ใช้ตัวแปรอิสระที่ผ่านการคัดเลือกจากสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) โดยใช้ข้อมูลยอดขายปกติ รูปแบบที่ 3 ใช้ตัวแปรที่ผ่านการคัดเลือกจากสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) โดยใช้ข้อมูลลอการิทึมธรรมชาติของยอดขาย จากนั้นจึงวัดประสิทธิภาพของตัวแบบด้วยค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) และนำตัวแบบที่ได้มาพยากรณ์ยอดขาย 7 วันข้างหน้า

4.1 ผลการศึกษาตัวแบบต้นไม้ตัดสินใจ (Decision Tree)

ผู้วิจัยได้ทำการนำข้อมูลทั้ง 3 รูปแบบที่กล่าวไว้ข้างต้นเข้าตัวแบบและทำการปรับค่าไฮเปอร์พารามิเตอร์ max_depth ในการควบคุมจำนวนความลึกของต้นไม้ตัดสินใจ โดยใช้ GridSearchCV กำหนดค่าไว้เป็น 1,3,5,10 จากนั้นทำการคำนวณค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของแต่ละเมนูและนำค่าที่ได้มาหาค่าเฉลี่ยของข้อมูลทั้ง 3 รูปแบบแสดงดังรูปที่ 4.1



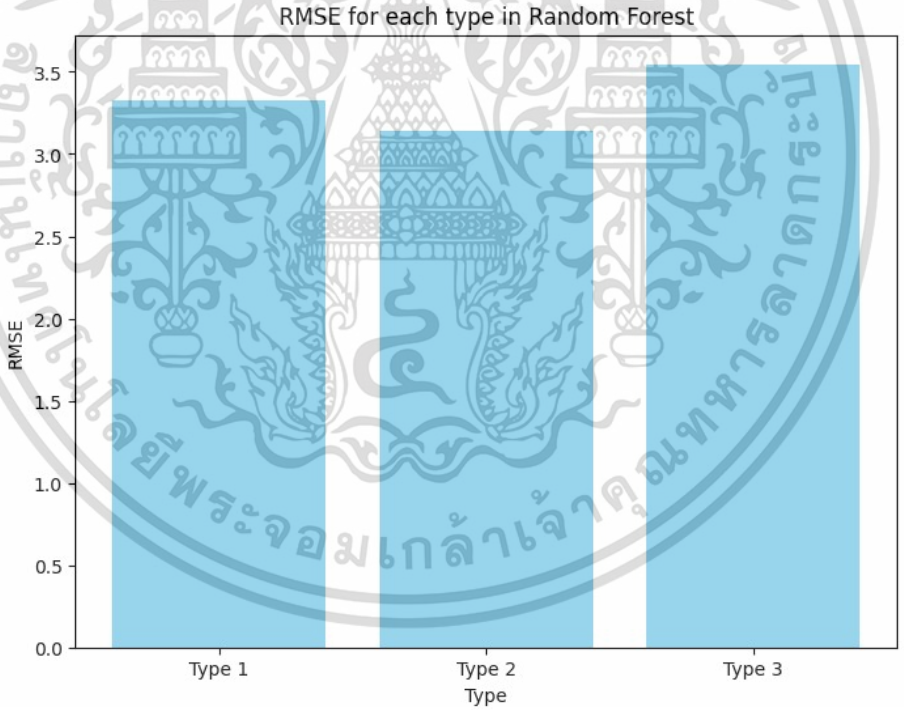
รูปที่ 4.1 ค่าเฉลี่ย RMSE ของทั้ง 3 รูปแบบการคัดเลือกตัวแปรที่ได้จากตัวแบบต้นไม้ตัดสินใจ

เอกสารนี้เป็นทรัพย์สินทางปัญญาของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ห้ามมิให้ทำซ้ำ ดัดแปลง หรือนำไปใช้เพื่อการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.1 แกน Y แสดงค่าเฉลี่ยที่คิดจากค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของเมนูทั้ง 89 เมนูที่ได้จากตัวแบบต้นไม้ตัดสินใจ โดยแบ่งตามรูปแบบการคัดเลือกตัวแปรโดย Type 1 Type 2และType 3 เป็นการคัดเลือกตัวแปรรูปแบบที่ 1 รูปแบบที่ 2 และรูปแบบที่ 3 ตามลำดับ และผลที่ได้การคัดเลือกตัวแปรรูปแบบที่ 2 ให้ผลค่าเฉลี่ย RMSE น้อยที่สุดรองลงมาคือรูปแบบที่ 1 และรูปแบบที่ 3 โดยค่าไฮเปอร์พารามิเตอร์และ RMSE ของแต่ละเมนูทั้ง 3 รูปแบบจะแสดงในตารางภาคผนวก ก-1, ก-2และก-3

4.2 ผลการศึกษาตัวแบบป่าสุ่ม (Random Forest)

ผู้วิจัยได้ทำการข้อมูลทั้ง 3 รูปแบบเข้าตัวแบบและทำการปรับค่าไฮเปอร์พารามิเตอร์ด้วยค่า max_depth ในการควบคุมความลึกของต้นไม้ตัดสินใจแต่ละต้นในตัวแบบป่าสุ่มโดยใช้ค่า 3, 5, 7, 10 และค่า n_estimators ในการควบคุมจำนวนต้นไม้ตัดสินใจที่สร้างในตัวแบบป่าสุ่มโดยใช้ค่า 100, 200, 300, 500 จากนั้นทำการคำนวณค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของแต่ละเมนูและนำค่าที่ได้มาหาค่าเฉลี่ยของข้อมูลทั้ง 3 รูปแบบ แสดงดังรูปที่ 4.2



รูปที่ 4.2 ค่าเฉลี่ย RMSE ของทั้ง 3 รูปแบบการคัดเลือกตัวแปรที่ได้จากตัวแบบป่าสุ่ม

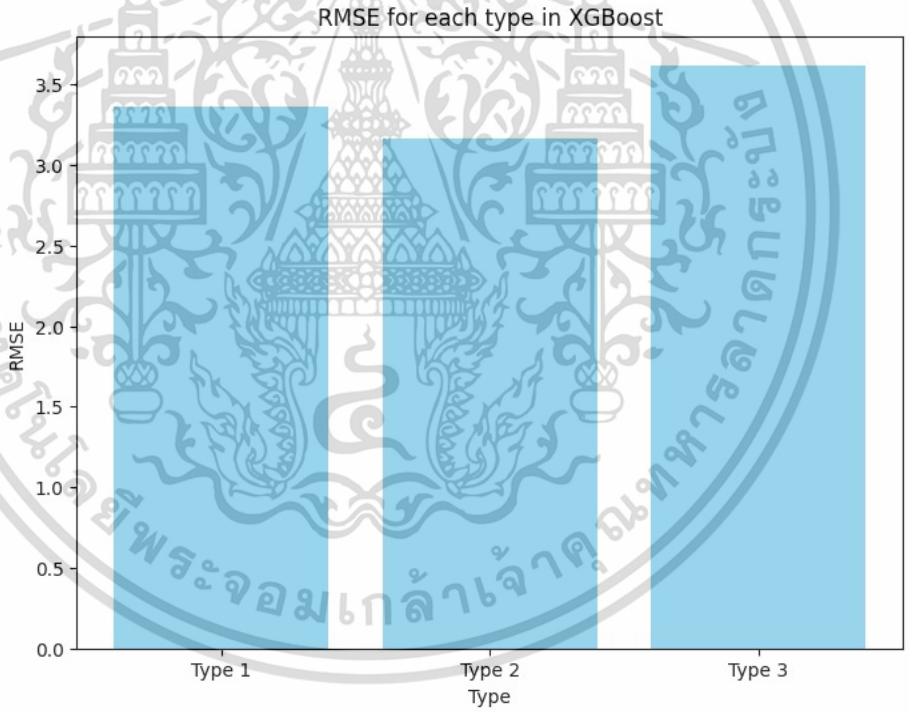
จากรูปที่ 4.2 แกน Y แสดงค่าเฉลี่ยที่คิดจากค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของเมนูทั้ง 89 เมนูที่ได้จากตัวแบบป่าสุ่ม โดยแบ่งตามรูปแบบการคัดเลือกตัวแปรโดย Type 1 Type 2และType 3 เป็นการคัดเลือกตัวแปรรูปแบบที่ 1 รูปแบบที่ 2 และรูปแบบที่ 3 ตามลำดับ และผลที่ได้การคัดเลือกตัวแบบรูปแบบที่ 2 ให้ค่าเฉลี่ยค่ารากที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า การแสวงหาผลประโยชน์ ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) น้อยที่สุดรองลงมา คือรูปแบบที่ 1 และรูปแบบที่ 3 โดยค่าไฮเปอร์พารามิเตอร์ของแต่ละเมนูและค่า RMSE ของแต่ละเมนูที่ได้จากตัวแบบป่าสุ่ม (Random Forest) ที่ปรับค่าไฮเปอร์พารามิเตอร์โดยใช้ GridSearchCV จะแสดงในตารางภาคผนวก ก-4 ก-5และก-6

4.3 ผลการศึกษาตัวแบบ eXtreme Gradient Boosting (XGboost)

ผู้วิจัยได้ทำการนำเข้าข้อมูลทั้ง 3 รูปแบบและทำการปรับค่าไฮเปอร์พารามิเตอร์ด้วยค่า max_depth ในการควบคุมความลึกของต้นไม้ตัดสินใจแต่ละต้นในตัวแบบ XGBoost โดยใช้ค่า 3, 5, 7, 10 และค่า n_estimators ในการควบคุมจำนวนต้นไม้ตัดสินใจที่สร้างในตัวแบบ XGBoost โดยใช้ค่า 100, 200, 300, 500 และค่า learning_rate เพื่อควบคุมการเรียนรู้ของตัวแบบไม่ให้เรียนรู้ข้อมูลเฉพาะมากเกินไปโดยใช้ค่า 0.05, 0.1, 0.3 จากนั้นทำการคำนวณค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของแต่ละเมนูและนำค่าที่ได้มาเฉลี่ยเป็นค่า RMSE เฉลี่ยของแต่ละรูปแบบโดยแสดงค่า RMSE เฉลี่ยของแต่ละรูปแบบดังรูปที่ 4.3



รูปที่ 4.3 ค่าเฉลี่ย RMSE ของทั้ง 3 รูปแบบการคัดเลือกตัวแปรที่ได้จากตัวแบบ XGBoost

จากรูปที่ 4.3 แกน Y แสดงค่าเฉลี่ยที่คิดจากค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของเมนูทั้ง 89 เมนูที่ได้จากตัวแบบ XGBoost โดยแบ่งตามรูปแบบการคัดเลือกตัวแปรโดย Type 1 Type 2และType 3 เป็นการคัดเลือกตัวแปรรูปแบบที่ 1 รูปแบบที่ 2 และรูปแบบที่ 3 ตามลำดับ และผลที่ได้การคัดเลือกตัวแบบรูปแบบที่ 2 ให้ค่าเฉลี่ยค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE)น้อยที่สุดรองลง

เอกสารนี้เป็นลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย ใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มาคือรูปแบบที่ 1 และรูปแบบที่ 3 โดยค่าไฮเปอร์พารามิเตอร์และ RMSE ของแต่ละเมนูที่ได้จากตัวแบบ eXtreme Gradient Bosting (XGboost) ที่ทำการปรับค่าไฮเปอร์พารามิเตอร์ด้วย GridSearchCV จะแสดงดังตารางในภาคผนวก ก-7 ก-8และ ก-9

4.4 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบ

หลังจากผู้วิจัยได้ทำการคัดเลือกตัวแปรอิสระจากข้อมูลทั้ง 3 รูปแบบคือ รูปแบบที่ 1 ใช้ตัวแปรอิสระทุกตัวในการทำนายยอดขาย รูปแบบที่ 2 ใช้ตัวแปรอิสระที่ผ่านการคัดเลือกจากสมการถดถอยเชิงเส้นพหุคูณ (Multiple linear Regression) รูปแบบที่ 3 ใช้ตัวแปรอิสระที่ผ่านการคัดเลือกจากสมการถดถอยเชิงเส้นพหุคูณ (Multiple linear Regression) ที่แปลงยอดขายด้วยลอกการิทึมธรรมชาติ จากนั้นนำตัวแปรอิสระที่ได้จากทั้ง 3 รูปแบบนำเข้าตัวแบบการเรียนรู้ของเครื่อง 3 ตัวแบบคือ ตัวแบบต้นไม้ตัดสินใจ (Decision Tree) ตัวแบบป่าสุ่ม (Random Forest) และตัวแบบ eXtreme Gradient Boosting (Xgboost) และปรับค่าไฮเปอร์พารามิเตอร์ด้วย GridSearchCV ได้ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของแต่ละวิธีการคัดเลือกตัวแปรแต่ละตัวแบบดังแสดงในตารางที่ 4.1

ตารางที่ 4.1 ค่าเฉลี่ย RMSE ทั้ง 3 รูปแบบการคัดเลือกตัวแปรที่ได้จากตัวแบบทั้ง 3 ตัวแบบ

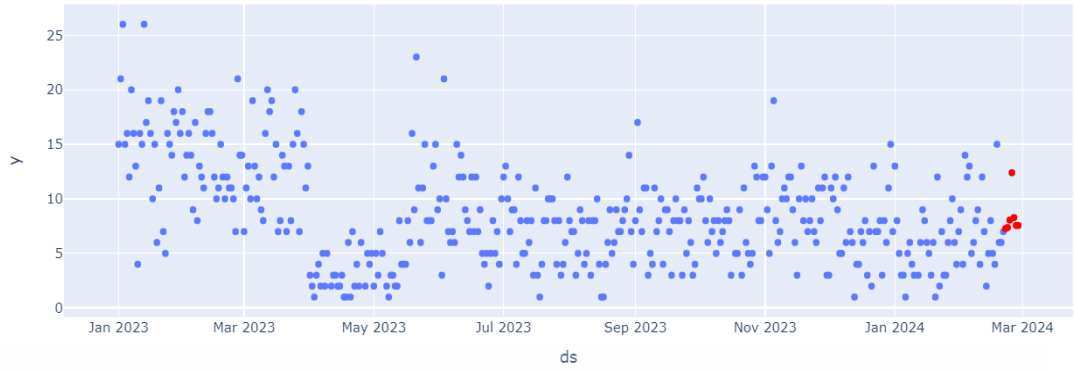
ตัวแบบ	รูปแบบที่ 1	รูปแบบที่ 2	รูปแบบที่ 3
ต้นไม้ตัดสินใจ (Decision Tree)	3.47	3.35	3.64
ตัวแบบป่าสุ่ม (Random Forest)	3.33	3.14	3.55
ตัวแบบ eXtreme Gradient Boosting (Xgboost)	3.36	3.17	3.62

เมื่อพิจารณาค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของแต่ละตัวแบบสำหรับแต่ละรูปแบบการคัดเลือกตัวแปรแล้วพบว่าตัวแบบป่าสุ่ม (Random Forest) โดยใช้การคัดเลือกตัวแปรรูปแบบที่ 2 ได้ค่า RMSE น้อยที่สุดที่ 3.14 ชิ้น

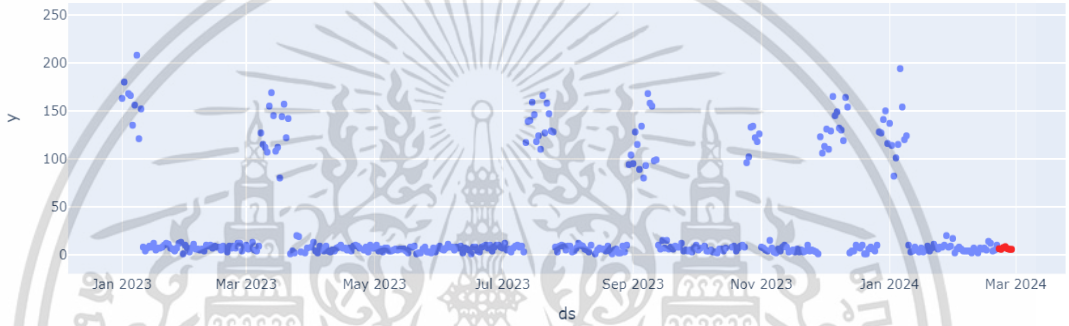
4.5 ผลการพยากรณ์ยอดขาย 7 วันข้างหน้าด้วยตัวแบบที่ดีที่สุด

จากการเปรียบเทียบตัวแบบและรูปแบบการคัดเลือกตัวแปรให้ผลว่าตัวแบบป่าสุ่ม (Random Forest) ที่ใช้การคัดเลือกตัวแปรรูปแบบที่ 2 เป็นตัวแบบที่ดีที่สุดจึงนำตัวแปรที่ได้จากรูปแบบที่ 2 แล้วค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุดจาก GridSearchCV มาใช้ในการทำนายด้วยตัวแบบป่าสุ่ม (Random Forest) โดยทำนายยอดขาย 7 วันข้างหน้า โดยตัวอย่างผลการทำนายแสดงดังรูป 4.4 และ 4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.4 ตัวอย่างการทำนายยอดขายสินค้า 7 วันข้างหน้าเมนู A



รูปที่ 4.5 ตัวอย่างการทำนายยอดขายสินค้า 7 วันข้างหน้าเมนู B

หมายเหตุ

จุดสีน้ำเงิน แทนค่า ข้อมูลยอดขายรายวันของสินค้า
จุดสีแดง แทนค่า ผลการทำนายยอดขายรายวันของสินค้า

จากรูปที่ 4.4 และ 4.5 เป็นตัวอย่างการทำนายยอดขายสินค้า 7 วันข้างหน้าของสินค้า 2 เมนู เพื่อประกอบการตัดสินใจผลิตสินค้าเพื่อให้เหมาะสมกับความต้องการของลูกค้าเพื่อลดการผลิตสินค้าเกินความต้องการของลูกค้าทำให้เกิดการทิ้งและสูญเสียกำไรในส่วนที่ทิ้งสินค้าไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์ในการวิเคราะห์หาตัวแบบในการทำนายยอดขายของสินค้าเบเกอรี่และทำนายยอดขายสินค้า 7 วันข้างหน้าเพื่อประกอบการตัดสินใจผลิตสินค้าของร้านค้าที่ศึกษา เพื่อลดการทิ้งสินค้าที่ผลิตมาเกินความต้องการของลูกค้าที่เข้ามาซื้อสินค้า โดยผู้วิจัยทำการแบ่งข้อมูลเป็น 3 รูปแบบคือ รูปแบบที่ 1 ใช้ตัวแปรอิสระทุกตัวในการทำนายยอดขายปกติ รูปแบบที่ 2 ใช้ตัวแปรอิสระที่ผ่านการคัดเลือกด้วยสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) รูปแบบที่ 3 ใช้ตัวแปรอิสระที่ผ่านการคัดเลือกด้วยสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ที่แปลงข้อมูลยอดขายด้วยลอการิทึมธรรมชาติ จากนั้นนำตัวแปรที่ได้จากทั้ง 3 รูปแบบนำเข้าตัวแบบ 3 ตัวแบบคือ ตัวแบบต้นไม้ตัดสินใจ (Decision Tree) ตัวแบบป่าสุ่ม (Random Forest) และตัวแบบ eXtreme Gradient Boosting (Xgboost) โดยใช้การปรับไฮเปอร์พารามิเตอร์ด้วยการค้นหาแบบกริด (Grid Search) จากนั้นคำนวณค่าเฉลี่ยของค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของแต่ละเมนูที่ได้จากตัวแบบทั้ง 3 ตัวแบบที่ใช้ข้อมูลตัวแปรอิสระทั้ง 3 รูปแบบที่กล่าวไว้ข้างต้น เพื่อวิเคราะห์หาตัวแบบและวิธีการคัดเลือกตัวแปรที่ดีที่สุดนำมาพยากรณ์ยอดขาย 7 วันข้างหน้าของสินค้าแต่ละเมนู

5.1 สรุปผลการวิจัย

จากผลการวิจัยวิธีการคัดเลือกตัวแบบทั้ง 3 รูปแบบและตัวแบบการเรียนรู้ของเครื่องทั้ง 3 ตัวแบบ นำมาสู่การสรุปผลการวิจัยดังนี้

1) การคัดเลือกตัวแปรทั้ง 3 รูปแบบ

จากผลการวิจัยทั้ง 3 รูปแบบคือ รูปแบบที่ 1 ใช้ตัวแปรอิสระทุกตัวในการทำนายยอดขายปกติ รูปแบบที่ 2 ใช้ตัวแปรอิสระที่ผ่านการคัดเลือกด้วยสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) รูปแบบที่ 3 ใช้ตัวแปรอิสระที่ผ่านการคัดเลือกด้วยสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ที่แปลงข้อมูลยอดขายด้วยลอการิทึมธรรมชาติ พบว่ารูปแบบที่ 2 ให้ผลการทำนายดีที่สุดทั้ง 3 ตัวแบบการเรียนรู้ของเครื่อง เนื่องมาจากรูปแบบที่ 1 ซึ่งเป็นการนำตัวแปรอิสระทุกตัวเข้าตัวแบบการเรียนรู้ของเครื่อง ทำให้ตัวแบบการเรียนรู้ของเครื่องมีความซับซ้อนของข้อมูลมากเกินไปทำให้ตัวแบบการเรียนรู้ของเครื่องเรียนรู้ข้อมูลของชุดฝึกมากเกินไปและส่งผลให้ประสิทธิภาพในการทำนายชุดข้อมูลใหม่ต่ำลง และรูปแบบที่ 3 เป็นการคัดเลือกตัวแปรด้วยสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ที่แปลงข้อมูลยอดขายด้วยลอการิทึมธรรมชาติ อาจเปลี่ยนแปลงลักษณะของข้อมูลให้มีความซับซ้อนมากขึ้นทำให้สมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) ไม่สามารถจับความสัมพันธ์ที่แท้จริงของตัวแปรอิสระและตัวแปรตามได้เมื่อนำตัวแปรที่ได้จากสมการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) มาใช้กับตัวแบบการเรียนรู้ของเครื่องทั้ง 3 ตัวแบบทำให้ประสิทธิภาพการทำนายลดลง

2) ตัวแบบต้นไม้ตัดสินใจ (Decision Tree)

การวิเคราะห์หาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมที่สุดสำหรับการทำนายยอดขายเบเกอรี่โดยใช้ตัวแบบต้นไม้ตัดสินใจใช้การค้นหาแบบกริด (Grid Search) เพื่อให้โมเดลมี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ทำซ้ำ ดัดแปลง เผยแพร่ หรือนำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น

ประสิทธิภาพในการทำนายและเหมาะสมกับข้อมูลมากที่สุดโดยค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมกับแต่ละเมนูแสดงในตารางภาคผนวก ก โดยค่าเฉลี่ย RMSE ของแต่ละรูปแบบการคัดเลือกตัวแปร 3 รูปแบบมีค่าเป็น 3.47, 3.35, 3.64 ชิ้น ตามลำดับ

โดยตัวแบบต้นไม้ตัดสินใจ (Decision Tree) เป็นตัวแบบที่มีความหลากหลายในการใช้งานสามารถนำไปใช้งานในด้านอื่นนอกเหนือจากการทำ Regression เช่น การจำแนกประเภท (Classification) ใช้ในการตรวจจับการทุจริต ใช้ในการจำแนกคำพูดหรือข้อความ เป็นต้น โดยมีหลักการทำงานคือแบ่งโหนด (node) โดยใช้ตัวแปรอิสระในการแบ่งจนกว่าจะไม่สามารถแบ่งโหนดได้อีก แล้วนำค่าที่ได้ในใบ (Leaf) มาเป็นผลลัพธ์ของตัวแบบ โดยตัวแบบต้นไม้ตัดสินใจเป็นตัวแบบที่มีค่าเฉลี่ย RMSE มากที่สุดเมื่อเทียบกับตัวแบบป่าสุ่ม (Random Forest) และตัวแบบ eXtreme Gradient Boosting (Xgboost)

3) ตัวแบบป่าสุ่ม (Random Forest)

การวิเคราะห์หาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมที่สุดสำหรับการทำนายยอดขายเบเกอรี่โดยใช้ตัวแบบป่าสุ่ม (Random Forest) ใช้การค้นหาแบบกริด (Grid Search) โดยปรับค่า n_estimators, max_depth เพื่อให้โมเดลมีประสิทธิภาพในการทำนายและเหมาะสมกับข้อมูลมากที่สุดโดยค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมกับแต่ละเมนูแสดงในตารางภาคผนวก ก โดยค่าเฉลี่ย RMSE ของแต่ละรูปแบบการคัดเลือกตัวแปร 3 รูปแบบมีค่าเป็น 3.33, 3.14, 3.55 ชิ้น ตามลำดับ

ตัวแบบป่าสุ่ม (Random Forest) เป็นตัวแบบที่มีหลักการทำงานโดยใช้การสร้างต้นไม้ตัดสินใจจำนวนมาก โดยในแต่ละต้นไม้ตัดสินใจจะมีข้อมูลนำเข้าไม่เหมือนกันและผลลัพธ์ที่ได้มาจากผลของต้นไม้ตัดสินใจแต่ละต้น โดยงานวิจัยนี้ตัวแบบป่าสุ่มเป็นตัวแบบที่มีค่า RMSE น้อยที่สุดเมื่อเทียบกับตัวแบบต้นไม้ตัดสินใจและตัวแบบeXtreme Gradient Boosting (Xgboost)

4) ตัวแบบ eXtreme Gradient Boosting (Xgboost)

การวิเคราะห์หาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมที่สุดสำหรับการทำนายยอดขายเบเกอรี่โดยใช้ตัวแบบ eXtreme Gradient Boosting (Xgboost) ใช้การค้นหาแบบกริด (Grid Search) โดยปรับค่า n_estimators, max_depth, learning_rate เพื่อให้โมเดลมีประสิทธิภาพในการทำนายและเหมาะสมกับข้อมูลมากที่สุดโดยค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมกับแต่ละเมนูแสดงในตารางภาคผนวก ก โดยค่าเฉลี่ย RMSE ของแต่ละรูปแบบการคัดเลือกตัวแปร 3 รูปแบบมีค่าเป็น 3.36, 3.17, 3.62 ชิ้น ตามลำดับ

ตัวแบบ eXtreme Gradient Boosting (Xgboost) เป็นตัวแบบการเรียนรู้ของเครื่องที่มีหลักการทำงานคือสร้างต้นไม้ตัดสินใจหลายๆต้นต่อกันและแก้ไขความผิดพลาดของต้นไม้ตัดสินใจก่อนหน้าจนกว่าจะถึงค่าที่กำหนดด้วยไฮเปอร์พารามิเตอร์ โดยในงานวิจัยนี้ตัวแบบeXtreme Gradient Boosting (Xgboost) เป็นตัวแบบที่มีค่าเฉลี่ย RMSE มากกว่าตัวแบบป่าสุ่ม (Random Forest) แต่น้อยกว่าตัวแบบต้นไม้ตัดสินใจ (Decision Tree)

เมื่อนำค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Square Error: RMSE) ของตัวแบบการเรียนรู้ของเครื่องทั้ง 3 ตัวแบบมาเปรียบเทียบกันเพื่อหาตัวแบบที่ดีที่สุดเพื่อนำไปทำนายยอดขายรายวันของสินค้าเบเกอรี่ 7 วันข้างหน้า พบว่าตัวแบบป่าสุ่ม (Random Forest) ที่ใช้วิธีคัดเลือกตัวแบบรูปแบบที่ 2 มีค่า RMSE น้อยที่สุด 3.14 ชิ้น จากตัวแบบที่ได้ไปใช้

เอกสารนี้ ... ไม่ว่ากรณีใดๆ ...

ทำนายยอดขายสินค้าเบเกอรี่ 7 วันข้างหน้าโดยนำไปประกอบการตัดสินใจผลิตสินค้าเพื่อให้ผลิตสินค้าให้เหมาะสมกับความต้องการของลูกค้าช่วยในการลดการทิ้งสินค้าที่ขายไม่หมดทำให้กำไรของร้านเบเกอรี่ที่ศึกษาดีขึ้น

5.2 ข้อเสนอแนะ

การปรับค่าไฮเปอร์พารามิเตอร์ของตัวแบบแต่ละตัวแบบมีค่าไฮเปอร์พารามิเตอร์ในการปรับตัวอื่นและช่วงในการปรับที่มากขึ้นอาจทำให้ตัวแบบทำนายข้อมูลได้ดีมากขึ้น แต่ควรระวังเรื่องการใช้ทรัพยากรในการทำงานของตัวแบบเช่น ทรัพยากรเรื่องเวลาในการทำงานของตัวแบบ ทรัพยากรคอมพิวเตอร์ที่ใช้สร้างตัวแบบ เป็นต้น

5.3 ข้อจำกัดในงานวิจัย

จากการที่ผู้วิจัยได้ไปสำรวจร้านค้าที่ศึกษาพบว่ามีความผิดพลาดที่ควบคุมไม่ได้ในการเก็บข้อมูลคือ การยิงบาร์โค๊ดสินค้าที่มีราคาเดียวกันแต่คนละชนิดกันอาจทำให้ข้อมูลเกิดความผิดพลาดได้ทำให้การทำนายยอดขายรายวันของสินค้านั้นๆผิดพลาดได้ โดยในงานวิจัยนี้ผู้วิจัยทำการวิจัยโดยสมมุติว่าข้อมูลทั้งหมดเก็บมาอย่างถูกต้องและไม่มีความผิดพลาดในการเก็บ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

ชัยพฤกษ์ นิละนนท์. (2563). **การพยากรณ์ปริมาณความต้องการเหล็กรีดร้อนภายในประเทศด้วยวิธีเรียนรู้ของเครื่อง.** วิทยานิพนธ์. หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมอุตสาหการ จุฬาลงกรณ์มหาวิทยาลัย.

ชยากรณ์ อุปกรณ์. (2564). **การพยากรณ์ยอดขายสินค้าที่มีราคาหลากหลายในธุรกิจค้าปลีกโดยใช้วิธีเรียนรู้ของเครื่อง.** วิทยานิพนธ์.หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมอุตสาหการ. จุฬาลงกรณ์มหาวิทยาลัย

พรรณทิพา วาณิชย์จิรัฐติกาล. (2564). **การวิเคราะห์การถดถอยและสหสัมพันธ์ : เอกสารประกอบการสอน.** สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. กรุงเทพมหานคร.

พรทิวา วิศิษฏ์สรรอรรถ. (2564). **ระบบวิเคราะห์ข้อมูลอนุกรมเวลาด้วยเทคนิดทางสถิติและการเรียนรู้ของเครื่อง.** สารนิพนธ์. หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมข้อมูลขนาดใหญ่ มหาวิทยาลัยธุรกิจบัณฑิตย์.

พิภพ ลลิตาภรณ์. (2553). **ระบบการวางแผนและควบคุมการผลิต (ฉบับปรับปรุง).** พิมพ์ครั้งที่ 15. กรุงเทพมหานคร. สมาคมส่งเสริมเทคโนโลยี (ไทย-ญี่ปุ่น).

วิโรจน์ อัศวรังสี. (2566). **Hands-on Machine Learning with scikit-learn, Keras & TensorFlow.** พิมพ์ครั้งที่ 1. กรุงเทพมหานคร. คอร์ฟังกชัน.

ศูนย์อัจฉริยะเพื่ออุตสาหกรรมอาหาร สถาบันอาหาร. (2563). **รายงานตลาดอาหารในประเทศไทย.** [Online]. การเข้าถึง : ตลาดขนมอบ_มี.ค.63.pdf (nfi.or.th)

สมศรี บัณฑิตวิไล. (2559). **การวิเคราะห์อนุกรมเวลา: เอกสารประกอบการสอน.** สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. กรุงเทพมหานคร.

Breiman et al. . (2527). **Classification And Regression Trees.** Wadsworth & Brooks/Cole.

Sasiwut Chaiyadecha. (2563). **Random forest regression + DD Balance หาราคาResidual value.** [Online]. การเข้าถึง : https://lengyi.medium.com/random-forest-regression-double-declining-balance-residual-value-model-6af8fd0c8709

Witchapong Daroontham. (2561). **รู้จัก Decision Tree, Random Forest, และ XGBoost!!! PART 1.** [Online]. การเข้าถึง: https://medium.com/@witchapongdaroontham/รู้จัก-decision-tree- random-forrest- และ-xgboost-part-1-cb49c4ac1315

Yang, T. (2023). **Sales Prediction of Walmart Sales Based on OLS, Random Forest, and XGBoost Models.** *Highlights in Science, Engineering and Technology AMMSAC*,49, 1-4.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

ตารางภาคผนวก ก ที่ 1 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนูที่ได้จากตัวแบบต้นไม้ตัดสินใจ โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 1

menu	Best Parameters	RMSE
0	{'max_depth': 5}	8.041
1	{'max_depth': 1}	0.957
2	{'max_depth': 1}	1.893
3	{'max_depth': 3}	4.130
4	{'max_depth': 1}	1.754
5	{'max_depth': 1}	3.025
6	{'max_depth': 1}	4.869
7	{'max_depth': 1}	0.694
8	{'max_depth': 1}	1.204
9	{'max_depth': 1}	1.806
10	{'max_depth': 1}	2.560
11	{'max_depth': 1}	1.628
12	{'max_depth': 1}	1.071
13	{'max_depth': 1}	2.840
14	{'max_depth': 1}	1.900
15	{'max_depth': 3}	31.647
16	{'max_depth': 1}	3.914
17	{'max_depth': 1}	0.798
18	{'max_depth': 1}	9.076
19	{'max_depth': 1}	5.821

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

20	{'max_depth': 1}	5.022
21	{'max_depth': 1}	1.079
22	{'max_depth': 1}	1.019
23	{'max_depth': 3}	6.179
24	{'max_depth': 3}	3.829
25	{'max_depth': 1}	0.499
26	{'max_depth': 1}	2.876
27	{'max_depth': 1}	0.691
28	{'max_depth': 1}	1.234
29	{'max_depth': 1}	1.230
30	{'max_depth': 1}	0.845
31	{'max_depth': 1}	0.548
32	{'max_depth': 1}	0.383
33	{'max_depth': 1}	0.503
34	{'max_depth': 1}	0.595
35	{'max_depth': 1}	0.728
36	{'max_depth': 1}	1.089
37	{'max_depth': 1}	2.490
38	{'max_depth': 1}	3.639
39	{'max_depth': 1}	14.142
40	{'max_depth': 3}	18.321
41	{'max_depth': 1}	2.140
42	{'max_depth': 1}	3.069
43	{'max_depth': 1}	2.681
44	{'max_depth': 1}	4.348

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

45	{'max_depth': 1}	3.879
46	{'max_depth': 1}	1.857
47	{'max_depth': 1}	2.446
48	{'max_depth': 1}	2.437
49	{'max_depth': 1}	1.976
50	{'max_depth': 1}	2.251
51	{'max_depth': 1}	1.817
52	{'max_depth': 1}	2.452
53	{'max_depth': 1}	1.398
54	{'max_depth': 1}	1.532
55	{'max_depth': 3}	1.369
56	{'max_depth': 3}	1.415
57	{'max_depth': 1}	1.391
58	{'max_depth': 1}	2.569
59	{'max_depth': 1}	2.843
60	{'max_depth': 1}	1.330
61	{'max_depth': 3}	1.124
62	{'max_depth': 1}	0.837
63	{'max_depth': 1}	0.456
64	{'max_depth': 3}	5.893
65	{'max_depth': 1}	0.744
66	{'max_depth': 1}	0.613
67	{'max_depth': 1}	1.061
68	{'max_depth': 1}	1.807
69	{'max_depth': 1}	4.501

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

70	{'max_depth': 1}	1.011
71	{'max_depth': 1}	6.137
72	{'max_depth': 1}	1.763
73	{'max_depth': 1}	1.685
74	{'max_depth': 1}	1.610
75	{'max_depth': 1}	2.847
76	{'max_depth': 1}	2.106
77	{'max_depth': 1}	1.004
78	{'max_depth': 1}	1.186
79	{'max_depth': 1}	0.401
80	{'max_depth': 1}	0.668
81	{'max_depth': 1}	0.803
82	{'max_depth': 1}	0.950
83	{'max_depth': 1}	2.199
84	{'max_depth': 10}	11.754
85	{'max_depth': 3}	16.885
86	{'max_depth': 1}	17.572
87	{'max_depth': 3}	10.274
88	{'max_depth': 1}	9.551

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางภาคผนวก ก ที่ 2 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนู ที่ได้จากตัวแบบต้นไม้ตัดสินใจ โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 2

menu	Best Parameters	RMSE
0	{'max_depth': 3}	8.453
1	{'max_depth': 3}	0.936
2	{'max_depth': 1}	1.893
3	{'max_depth': 3}	4.190
4	{'max_depth': 5}	2.053
5	{'max_depth': 1}	3.025
6	{'max_depth': 10}	5.044
7	{'max_depth': 1}	0.707
8	{'max_depth': 1}	1.204
9	{'max_depth': 1}	1.806
10	{'max_depth': 1}	2.560
11	{'max_depth': 3}	1.696
12	{'max_depth': 3}	1.063
13	{'max_depth': 1}	2.840
14	{'max_depth': 3}	1.895
15	{'max_depth': 3}	23.513
16	{'max_depth': 5}	3.947
17	{'max_depth': 3}	0.778
18	{'max_depth': 1}	9.076
19	{'max_depth': 1}	5.821
20	{'max_depth': 3}	4.483
21	{'max_depth': 5}	1.141

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	{'max_depth': 1}	1.019
23	{'max_depth': 5}	6.220
24	{'max_depth': 3}	3.912
25	{'max_depth': 1}	0.499
26	{'max_depth': 3}	2.862
27	{'max_depth': 3}	0.648
28	{'max_depth': 1}	1.234
29	{'max_depth': 3}	1.227
30	{'max_depth': 3}	0.809
31	{'max_depth': 1}	0.548
32	{'max_depth': 1}	0.383
33	{'max_depth': 1}	0.503
34	{'max_depth': 1}	0.595
35	{'max_depth': 5}	0.751
36	{'max_depth': 3}	1.074
37	{'max_depth': 1}	2.629
38	{'max_depth': 3}	3.675
39	{'max_depth': 1}	14.142
40	{'max_depth': 5}	13.520
41	{'max_depth': 3}	2.161
42	{'max_depth': 3}	2.901
43	{'max_depth': 10}	2.735
44	{'max_depth': 1}	4.348
45	{'max_depth': 10}	3.591
46	{'max_depth': 1}	1.857

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

47	{'max_depth': 5}	2.411
48	{'max_depth': 5}	2.461
49	{'max_depth': 3}	2.046
50	{'max_depth': 5}	2.265
51	{'max_depth': 3}	1.859
52	{'max_depth': 3}	2.240
53	{'max_depth': 1}	1.398
54	{'max_depth': 1}	1.532
55	{'max_depth': 3}	1.269
56	{'max_depth': 3}	1.440
57	{'max_depth': 1}	1.428
58	{'max_depth': 3}	2.664
59	{'max_depth': 1}	2.843
60	{'max_depth': 1}	1.302
61	{'max_depth': 3}	0.957
62	{'max_depth': 1}	0.837
63	{'max_depth': 1}	0.456
64	{'max_depth': 1}	5.149
65	{'max_depth': 5}	0.863
66	{'max_depth': 1}	0.627
67	{'max_depth': 3}	1.097
68	{'max_depth': 5}	1.733
69	{'max_depth': 3}	3.475
70	{'max_depth': 3}	0.942
71	{'max_depth': 3}	4.352

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

72	{'max_depth': 1}	1.763
73	{'max_depth': 1}	1.685
74	{'max_depth': 10}	1.563
75	{'max_depth': 1}	2.373
76	{'max_depth': 3}	2.185
77	{'max_depth': 3}	1.209
78	{'max_depth': 3}	1.045
79	{'max_depth': 1}	0.401
80	{'max_depth': 1}	0.637
81	{'max_depth': 3}	0.784
82	{'max_depth': 5}	0.921
83	{'max_depth': 5}	2.213
84	{'max_depth': 5}	10.511
85	{'max_depth': 3}	22.514
86	{'max_depth': 1}	17.572
87	{'max_depth': 3}	11.424
88	{'max_depth': 1}	9.551

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางภาคผนวก ก ที่ 3 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนูที่ได้จากตัวแบบต้นไม้ตัดสินใจ โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 3

menu	Best Parameters	RMSE
0	{'max_depth': 5}	7.072
1	{'max_depth': 3}	0.975
2	{'max_depth': 1}	1.997
3	{'max_depth': 3}	4.368
4	{'max_depth': 1}	1.878
5	{'max_depth': 5}	3.192
6	{'max_depth': 10}	5.229
7	{'max_depth': 3}	0.714
8	{'max_depth': 1}	1.235
9	{'max_depth': 1}	1.889
10	{'max_depth': 3}	2.509
11	{'max_depth': 5}	1.749
12	{'max_depth': 3}	1.128
13	{'max_depth': 1}	3.086
14	{'max_depth': 5}	2.082
15	{'max_depth': 3}	24.480
16	{'max_depth': 5}	4.397
17	{'max_depth': 1}	0.848
18	{'max_depth': 1}	8.967
19	{'max_depth': 1}	5.750
20	{'max_depth': 1}	4.871
21	{'max_depth': 5}	1.129

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	{'max_depth': 1}	1.088
23	{'max_depth': 1}	4.959
24	{'max_depth': 3}	4.010
25	{'max_depth': 1}	0.498
26	{'max_depth': 3}	2.815
27	{'max_depth': 3}	0.709
28	{'max_depth': 1}	1.312
29	{'max_depth': 3}	1.296
30	{'max_depth': 3}	0.849
31	{'max_depth': 1}	0.526
32	{'max_depth': 1}	0.378
33	{'max_depth': 1}	0.488
34	{'max_depth': 1}	0.587
35	{'max_depth': 5}	0.675
36	{'max_depth': 1}	1.153
37	{'max_depth': 1}	2.738
38	{'max_depth': 5}	3.878
39	{'max_depth': 5}	16.270
40	{'max_depth': 5}	17.874
41	{'max_depth': 10}	2.245
42	{'max_depth': 1}	2.926
43	{'max_depth': 10}	2.665
44	{'max_depth': 1}	4.777
45	{'max_depth': 1}	3.921
46	{'max_depth': 1}	1.895

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

47	{'max_depth': 5}	2.527
48	{'max_depth': 5}	2.670
49	{'max_depth': 3}	2.077
50	{'max_depth': 10}	2.361
51	{'max_depth': 5}	2.090
52	{'max_depth': 10}	2.475
53	{'max_depth': 1}	1.486
54	{'max_depth': 1}	1.580
55	{'max_depth': 10}	1.299
56	{'max_depth': 3}	1.397
57	{'max_depth': 1}	1.428
58	{'max_depth': 3}	2.767
59	{'max_depth': 1}	3.053
60	{'max_depth': 3}	1.446
61	{'max_depth': 1}	0.980
62	{'max_depth': 1}	0.859
63	{'max_depth': 1}	0.428
64	{'max_depth': 1}	6.250
65	{'max_depth': 3}	0.875
66	{'max_depth': 1}	0.643
67	{'max_depth': 1}	1.093
68	{'max_depth': 5}	1.862
69	{'max_depth': 5}	3.865
70	{'max_depth': 3}	0.932
71	{'max_depth': 1}	6.466

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

72	{'max_depth': 3}	1.803
73	{'max_depth': 5}	1.677
74	{'max_depth': 10}	1.658
75	{'max_depth': 1}	3.016
76	{'max_depth': 3}	2.145
77	{'max_depth': 3}	1.013
78	{'max_depth': 3}	1.081
79	{'max_depth': 1}	0.402
80	{'max_depth': 1}	0.630
81	{'max_depth': 1}	0.796
82	{'max_depth': 5}	0.874
83	{'max_depth': 5}	2.312
84	{'max_depth': 3}	11.466
85	{'max_depth': 3}	28.581
86	{'max_depth': 3}	16.679
87	{'max_depth': 3}	17.129
88	{'max_depth': 5}	9.430

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางภาคผนวก ก ที่ 4 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนูที่ได้จากตัวแบบป่าสุ่ม โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 1

menu	Best Parameters	RMSE
0	{'max_depth': 10, 'n_estimators': 200}	8.631
1	{'max_depth': 7, 'n_estimators': 300}	0.949
2	{'max_depth': 3, 'n_estimators': 300}	1.870
3	{'max_depth': 10, 'n_estimators': 100}	3.815
4	{'max_depth': 3, 'n_estimators': 300}	1.733
5	{'max_depth': 7, 'n_estimators': 500}	3.035
6	{'max_depth': 3, 'n_estimators': 200}	4.630
7	{'max_depth': 3, 'n_estimators': 500}	0.717
8	{'max_depth': 3, 'n_estimators': 100}	1.187
9	{'max_depth': 3, 'n_estimators': 500}	1.786
10	{'max_depth': 3, 'n_estimators': 200}	2.596
11	{'max_depth': 3, 'n_estimators': 500}	1.590
12	{'max_depth': 3, 'n_estimators': 100}	1.055
13	{'max_depth': 3, 'n_estimators': 500}	2.817
14	{'max_depth': 3, 'n_estimators': 200}	1.843
15	{'max_depth': 7, 'n_estimators': 100}	26.462
16	{'max_depth': 3, 'n_estimators': 300}	3.996
17	{'max_depth': 3, 'n_estimators': 100}	0.746
18	{'max_depth': 10, 'n_estimators': 100}	11.362
19	{'max_depth': 7, 'n_estimators': 200}	6.654
20	{'max_depth': 3, 'n_estimators': 100}	4.493
21	{'max_depth': 3, 'n_estimators': 300}	1.082

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	{'max_depth': 3, 'n_estimators': 500}	0.979
23	{'max_depth': 3, 'n_estimators': 100}	5.372
24	{'max_depth': 5, 'n_estimators': 100}	4.067
25	{'max_depth': 3, 'n_estimators': 500}	0.490
26	{'max_depth': 3, 'n_estimators': 200}	2.963
27	{'max_depth': 3, 'n_estimators': 500}	0.675
28	{'max_depth': 3, 'n_estimators': 300}	1.289
29	{'max_depth': 3, 'n_estimators': 100}	1.200
30	{'max_depth': 3, 'n_estimators': 500}	0.839
31	{'max_depth': 3, 'n_estimators': 300}	0.571
32	{'max_depth': 3, 'n_estimators': 300}	0.376
33	{'max_depth': 3, 'n_estimators': 200}	0.451
34	{'max_depth': 3, 'n_estimators': 200}	0.602
35	{'max_depth': 3, 'n_estimators': 500}	0.687
36	{'max_depth': 3, 'n_estimators': 100}	1.034
37	{'max_depth': 5, 'n_estimators': 100}	2.427
38	{'max_depth': 5, 'n_estimators': 100}	3.431
39	{'max_depth': 3, 'n_estimators': 100}	14.042
40	{'max_depth': 10, 'n_estimators': 200}	16.902
41	{'max_depth': 7, 'n_estimators': 500}	2.155
42	{'max_depth': 3, 'n_estimators': 500}	2.956
43	{'max_depth': 3, 'n_estimators': 100}	2.619
44	{'max_depth': 10, 'n_estimators': 500}	3.771
45	{'max_depth': 3, 'n_estimators': 100}	3.590
46	{'max_depth': 3, 'n_estimators': 300}	1.798

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

47	{'max_depth': 3, 'n_estimators': 100}	2.406
48	{'max_depth': 3, 'n_estimators': 100}	2.432
49	{'max_depth': 3, 'n_estimators': 300}	2.030
50	{'max_depth': 3, 'n_estimators': 500}	2.256
51	{'max_depth': 3, 'n_estimators': 300}	1.847
52	{'max_depth': 3, 'n_estimators': 300}	2.397
53	{'max_depth': 3, 'n_estimators': 500}	1.400
54	{'max_depth': 3, 'n_estimators': 300}	1.424
55	{'max_depth': 3, 'n_estimators': 100}	1.235
56	{'max_depth': 3, 'n_estimators': 100}	1.338
57	{'max_depth': 3, 'n_estimators': 500}	1.365
58	{'max_depth': 3, 'n_estimators': 500}	2.542
59	{'max_depth': 3, 'n_estimators': 300}	2.907
60	{'max_depth': 3, 'n_estimators': 500}	1.345
61	{'max_depth': 5, 'n_estimators': 300}	1.045
62	{'max_depth': 3, 'n_estimators': 500}	0.791
63	{'max_depth': 3, 'n_estimators': 200}	0.641
64	{'max_depth': 3, 'n_estimators': 500}	5.246
65	{'max_depth': 3, 'n_estimators': 300}	0.789
66	{'max_depth': 3, 'n_estimators': 100}	0.644
67	{'max_depth': 3, 'n_estimators': 500}	1.136
68	{'max_depth': 3, 'n_estimators': 500}	1.702
69	{'max_depth': 3, 'n_estimators': 200}	3.599
70	{'max_depth': 3, 'n_estimators': 500}	0.995
71	{'max_depth': 3, 'n_estimators': 200}	5.652

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

72	{'max_depth': 3, 'n_estimators': 200}	1.730
73	{'max_depth': 3, 'n_estimators': 200}	1.619
74	{'max_depth': 3, 'n_estimators': 300}	1.571
75	{'max_depth': 3, 'n_estimators': 100}	2.772
76	{'max_depth': 3, 'n_estimators': 500}	2.141
77	{'max_depth': 3, 'n_estimators': 100}	1.075
78	{'max_depth': 3, 'n_estimators': 500}	1.182
79	{'max_depth': 3, 'n_estimators': 500}	0.515
80	{'max_depth': 3, 'n_estimators': 200}	0.653
81	{'max_depth': 3, 'n_estimators': 500}	0.814
82	{'max_depth': 7, 'n_estimators': 200}	0.926
83	{'max_depth': 7, 'n_estimators': 500}	1.998
84	{'max_depth': 5, 'n_estimators': 500}	6.823
85	{'max_depth': 7, 'n_estimators': 100}	17.837
86	{'max_depth': 3, 'n_estimators': 500}	16.976
87	{'max_depth': 5, 'n_estimators': 500}	10.679
88	{'max_depth': 5, 'n_estimators': 200}	9.534

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางภาคผนวก ก ที่ 5 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนูที่ได้จากตัวแบบป่าสุ่ม โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 2

menu	Best Parameters	RMSE
0	{'max_depth': 3, 'n_estimators': 200}	9.337
1	{'max_depth': 3, 'n_estimators': 200}	0.936
2	{'max_depth': 3, 'n_estimators': 500}	1.940
3	{'max_depth': 5, 'n_estimators': 300}	3.921
4	{'max_depth': 3, 'n_estimators': 200}	1.850
5	{'max_depth': 3, 'n_estimators': 300}	2.974
6	{'max_depth': 7, 'n_estimators': 300}	4.787
7	{'max_depth': 3, 'n_estimators': 100}	0.743
8	{'max_depth': 3, 'n_estimators': 100}	1.210
9	{'max_depth': 3, 'n_estimators': 300}	1.760
10	{'max_depth': 3, 'n_estimators': 100}	2.540
11	{'max_depth': 3, 'n_estimators': 200}	1.606
12	{'max_depth': 3, 'n_estimators': 100}	1.080
13	{'max_depth': 3, 'n_estimators': 500}	2.815
14	{'max_depth': 5, 'n_estimators': 100}	1.910
15	{'max_depth': 5, 'n_estimators': 100}	19.385
16	{'max_depth': 3, 'n_estimators': 200}	3.928
17	{'max_depth': 3, 'n_estimators': 100}	0.777
18	{'max_depth': 3, 'n_estimators': 300}	9.029
19	{'max_depth': 3, 'n_estimators': 200}	6.650
20	{'max_depth': 3, 'n_estimators': 300}	4.332
21	{'max_depth': 7, 'n_estimators': 300}	1.126

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	{'max_depth': 3, 'n_estimators': 300}	1.061
23	{'max_depth': 3, 'n_estimators': 300}	5.228
24	{'max_depth': 5, 'n_estimators': 100}	4.194
25	{'max_depth': 7, 'n_estimators': 100}	0.528
26	{'max_depth': 3, 'n_estimators': 500}	2.888
27	{'max_depth': 3, 'n_estimators': 100}	0.660
28	{'max_depth': 3, 'n_estimators': 300}	1.279
29	{'max_depth': 3, 'n_estimators': 300}	1.193
30	{'max_depth': 3, 'n_estimators': 100}	0.800
31	{'max_depth': 3, 'n_estimators': 200}	0.545
32	{'max_depth': 3, 'n_estimators': 300}	0.413
33	{'max_depth': 3, 'n_estimators': 300}	0.434
34	{'max_depth': 3, 'n_estimators': 200}	0.578
35	{'max_depth': 5, 'n_estimators': 100}	0.745
36	{'max_depth': 3, 'n_estimators': 300}	1.072
37	{'max_depth': 3, 'n_estimators': 300}	2.399
38	{'max_depth': 5, 'n_estimators': 500}	3.377
39	{'max_depth': 3, 'n_estimators': 200}	10.674
40	{'max_depth': 10, 'n_estimators': 100}	10.527
41	{'max_depth': 3, 'n_estimators': 200}	2.113
42	{'max_depth': 3, 'n_estimators': 100}	2.889
43	{'max_depth': 5, 'n_estimators': 500}	2.701
44	{'max_depth': 7, 'n_estimators': 500}	3.679
45	{'max_depth': 10, 'n_estimators': 100}	3.573
46	{'max_depth': 3, 'n_estimators': 100}	1.821

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

47	{'max_depth': 3, 'n_estimators': 200}	2.424
48	{'max_depth': 3, 'n_estimators': 300}	2.452
49	{'max_depth': 3, 'n_estimators': 300}	2.050
50	{'max_depth': 3, 'n_estimators': 100}	2.251
51	{'max_depth': 3, 'n_estimators': 100}	1.850
52	{'max_depth': 3, 'n_estimators': 100}	2.223
53	{'max_depth': 3, 'n_estimators': 500}	1.373
54	{'max_depth': 3, 'n_estimators': 100}	1.448
55	{'max_depth': 3, 'n_estimators': 100}	1.236
56	{'max_depth': 3, 'n_estimators': 100}	1.377
57	{'max_depth': 3, 'n_estimators': 500}	1.430
58	{'max_depth': 3, 'n_estimators': 200}	2.605
59	{'max_depth': 3, 'n_estimators': 500}	2.853
60	{'max_depth': 3, 'n_estimators': 200}	1.438
61	{'max_depth': 3, 'n_estimators': 100}	0.957
62	{'max_depth': 3, 'n_estimators': 500}	0.853
63	{'max_depth': 3, 'n_estimators': 200}	0.456
64	{'max_depth': 3, 'n_estimators': 200}	4.365
65	{'max_depth': 3, 'n_estimators': 300}	0.846
66	{'max_depth': 3, 'n_estimators': 100}	0.616
67	{'max_depth': 3, 'n_estimators': 300}	1.096
68	{'max_depth': 3, 'n_estimators': 300}	1.705
69	{'max_depth': 3, 'n_estimators': 200}	3.910
70	{'max_depth': 3, 'n_estimators': 100}	0.941
71	{'max_depth': 3, 'n_estimators': 500}	5.643

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

72	{'max_depth': 5, 'n_estimators': 100}	1.828
73	{'max_depth': 3, 'n_estimators': 200}	1.628
74	{'max_depth': 7, 'n_estimators': 500}	1.556
75	{'max_depth': 3, 'n_estimators': 100}	2.403
76	{'max_depth': 3, 'n_estimators': 500}	2.180
77	{'max_depth': 3, 'n_estimators': 200}	1.174
78	{'max_depth': 3, 'n_estimators': 500}	1.043
79	{'max_depth': 3, 'n_estimators': 500}	0.384
80	{'max_depth': 3, 'n_estimators': 300}	0.640
81	{'max_depth': 3, 'n_estimators': 500}	0.818
82	{'max_depth': 5, 'n_estimators': 300}	0.918
83	{'max_depth': 3, 'n_estimators': 200}	2.196
84	{'max_depth': 5, 'n_estimators': 500}	7.353
85	{'max_depth': 3, 'n_estimators': 100}	20.592
86	{'max_depth': 3, 'n_estimators': 200}	16.793
87	{'max_depth': 3, 'n_estimators': 100}	10.945
88	{'max_depth': 3, 'n_estimators': 100}	9.227

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางภาคผนวก ก ที่ 6 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนูที่ได้จากตัวแบบป่าสุ่ม โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 3

menu	Best Parameters	RMSE
0	{'max_depth': 5, 'n_estimators': 200}	6.729
1	{'max_depth': 3, 'n_estimators': 100}	0.983
2	{'max_depth': 3, 'n_estimators': 200}	2.005
3	{'max_depth': 3, 'n_estimators': 200}	4.366
4	{'max_depth': 3, 'n_estimators': 100}	1.874
5	{'max_depth': 5, 'n_estimators': 100}	3.173
6	{'max_depth': 7, 'n_estimators': 300}	4.990
7	{'max_depth': 3, 'n_estimators': 100}	0.712
8	{'max_depth': 3, 'n_estimators': 100}	1.252
9	{'max_depth': 3, 'n_estimators': 300}	1.869
10	{'max_depth': 3, 'n_estimators': 100}	2.462
11	{'max_depth': 5, 'n_estimators': 300}	1.697
12	{'max_depth': 3, 'n_estimators': 100}	1.119
13	{'max_depth': 3, 'n_estimators': 200}	3.003
14	{'max_depth': 5, 'n_estimators': 100}	1.994
15	{'max_depth': 3, 'n_estimators': 300}	23.764
16	{'max_depth': 5, 'n_estimators': 300}	4.411
17	{'max_depth': 3, 'n_estimators': 100}	0.830
18	{'max_depth': 5, 'n_estimators': 500}	8.893
19	{'max_depth': 3, 'n_estimators': 300}	6.944
20	{'max_depth': 5, 'n_estimators': 100}	4.459
21	{'max_depth': 3, 'n_estimators': 300}	1.103

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	{'max_depth': 3, 'n_estimators': 300}	1.108
23	{'max_depth': 3, 'n_estimators': 300}	4.870
24	{'max_depth': 3, 'n_estimators': 100}	4.452
25	{'max_depth': 3, 'n_estimators': 100}	0.496
26	{'max_depth': 3, 'n_estimators': 500}	2.849
27	{'max_depth': 3, 'n_estimators': 200}	0.711
28	{'max_depth': 3, 'n_estimators': 200}	1.333
29	{'max_depth': 3, 'n_estimators': 500}	1.274
30	{'max_depth': 3, 'n_estimators': 100}	0.839
31	{'max_depth': 3, 'n_estimators': 200}	0.526
32	{'max_depth': 3, 'n_estimators': 300}	0.378
33	{'max_depth': 3, 'n_estimators': 300}	0.437
34	{'max_depth': 3, 'n_estimators': 200}	0.580
35	{'max_depth': 3, 'n_estimators': 100}	0.680
36	{'max_depth': 3, 'n_estimators': 500}	1.117
37	{'max_depth': 3, 'n_estimators': 200}	2.679
38	{'max_depth': 7, 'n_estimators': 500}	3.124
39	{'max_depth': 3, 'n_estimators': 200}	16.634
40	{'max_depth': 10, 'n_estimators': 100}	15.920
41	{'max_depth': 3, 'n_estimators': 300}	2.184
42	{'max_depth': 3, 'n_estimators': 500}	2.872
43	{'max_depth': 3, 'n_estimators': 500}	2.643
44	{'max_depth': 7, 'n_estimators': 200}	3.953
45	{'max_depth': 5, 'n_estimators': 100}	3.609
46	{'max_depth': 3, 'n_estimators': 100}	1.781

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

47	{'max_depth': 3, 'n_estimators': 200}	2.542
48	{'max_depth': 3, 'n_estimators': 300}	2.671
49	{'max_depth': 3, 'n_estimators': 300}	2.067
50	{'max_depth': 3, 'n_estimators': 100}	2.359
51	{'max_depth': 3, 'n_estimators': 100}	2.005
52	{'max_depth': 7, 'n_estimators': 300}	2.421
53	{'max_depth': 3, 'n_estimators': 500}	1.495
54	{'max_depth': 3, 'n_estimators': 100}	1.477
55	{'max_depth': 3, 'n_estimators': 100}	1.239
56	{'max_depth': 3, 'n_estimators': 500}	1.316
57	{'max_depth': 3, 'n_estimators': 300}	1.427
58	{'max_depth': 3, 'n_estimators': 500}	2.755
59	{'max_depth': 5, 'n_estimators': 100}	3.112
60	{'max_depth': 3, 'n_estimators': 100}	1.445
61	{'max_depth': 3, 'n_estimators': 100}	0.983
62	{'max_depth': 3, 'n_estimators': 500}	0.870
63	{'max_depth': 3, 'n_estimators': 200}	0.428
64	{'max_depth': 3, 'n_estimators': 100}	6.102
65	{'max_depth': 3, 'n_estimators': 300}	0.853
66	{'max_depth': 3, 'n_estimators': 100}	0.641
67	{'max_depth': 3, 'n_estimators': 300}	1.094
68	{'max_depth': 3, 'n_estimators': 200}	1.853
69	{'max_depth': 5, 'n_estimators': 200}	3.765
70	{'max_depth': 3, 'n_estimators': 100}	0.931
71	{'max_depth': 3, 'n_estimators': 200}	6.349

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

72	{'max_depth': 3, 'n_estimators': 200}	1.807
73	{'max_depth': 3, 'n_estimators': 200}	1.702
74	{'max_depth': 7, 'n_estimators': 500}	1.655
75	{'max_depth': 3, 'n_estimators': 100}	3.069
76	{'max_depth': 3, 'n_estimators': 100}	2.141
77	{'max_depth': 3, 'n_estimators': 300}	1.013
78	{'max_depth': 3, 'n_estimators': 500}	1.081
79	{'max_depth': 3, 'n_estimators': 500}	0.372
80	{'max_depth': 3, 'n_estimators': 300}	0.630
81	{'max_depth': 3, 'n_estimators': 300}	0.819
82	{'max_depth': 5, 'n_estimators': 500}	0.874
83	{'max_depth': 5, 'n_estimators': 200}	2.302
84	{'max_depth': 3, 'n_estimators': 500}	7.997
85	{'max_depth': 7, 'n_estimators': 300}	27.752
86	{'max_depth': 7, 'n_estimators': 500}	18.874
87	{'max_depth': 3, 'n_estimators': 200}	16.574
88	{'max_depth': 5, 'n_estimators': 100}	9.058

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางภาคผนวก ก ที่ 7 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนู ที่ได้จากตัวแบบXGBoost โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 1

menu	Best Parameters	RMSE
0	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	10.216
1	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.936
2	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.869
3	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.754
4	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.845
5	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.997
6	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.316
7	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.751
8	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.227
9	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.795
10	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.681
11	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.585
12	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.069
13	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.796
14	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.871
15	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	23.802
16	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.037
17	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.769
18	{'learning_rate': 0.1, 'max_depth': 7, 'n_estimators': 100}	11.045
19	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	7.560
20	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	4.539
21	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.146

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.991
23	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 300}	5.203
24	{'learning_rate': 0.3, 'max_depth': 3, 'n_estimators': 100}	3.492
25	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.506
26	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.855
27	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.702
28	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.393
29	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.184
30	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.865
31	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.614
32	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.378
33	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.448
34	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.660
35	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.685
36	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.052
37	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.606
38	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.423
39	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	13.630
40	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	15.861
41	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.137
42	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.977
43	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.731
44	{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 200}	3.758
45	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.436
46	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.849

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

47	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.418
48	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.485
49	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.069
50	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.344
51	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.892
52	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.361
53	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.493
54	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.475
55	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.269
56	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.393
57	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.583
58	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.516
59	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.011
60	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.399
61	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.139
62	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.802
63	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	0.875
64	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	5.998
65	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.890
66	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.741
67	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.295
68	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.661
69	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.753
70	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.987
71	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	5.856

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

72	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.908
73	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.632
74	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.574
75	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.169
76	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.384
77	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.167
78	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.154
79	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.542
80	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.687
81	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.857
82	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.868
83	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.994
84	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	7.598
85	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	17.795
86	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	16.382
87	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	10.893
88	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	9.681

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางภาคผนวก ก ที่ 8 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนูที่ได้จากตัวแบบXGBoost โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 2

menu	Best Parameters	RMSE
0	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	10.931
1	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.936
2	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.939
3	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.782
4	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	2.023
5	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.037
6	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.746
7	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.789
8	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	1.217
9	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.748
10	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.541
11	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.618
12	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.083
13	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.864
14	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.866
15	{'learning_rate': 0.05, 'max_depth': 10, 'n_estimators': 100}	18.457
16	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.945
17	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.777
18	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	6.187
19	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	7.881
20	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.425
21	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.119

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.062
23	{'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 100}	5.249
24	{'learning_rate': 0.3, 'max_depth': 3, 'n_estimators': 100}	3.712
25	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.524
26	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.867
27	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.647
28	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.298
29	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.195
30	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.804
31	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.563
32	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.434
33	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.421
34	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.577
35	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.726
36	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.074
37	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.342
38	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.453
39	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 200}	14.142
40	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 200}	14.980
41	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.112
42	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.936
43	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.691
44	{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100}	3.638
45	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.516
46	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.966

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

47	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.411
48	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.457
49	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.049
50	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.259
51	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.886
52	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.203
53	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.370
54	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.531
55	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.264
56	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.380
57	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.427
58	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.627
59	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.859
60	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.430
61	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.958
62	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.882
63	{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100}	0.456
64	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.779
65	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.855
66	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.620
67	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.096
68	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.713
69	{'learning_rate': 0.05, 'max_depth': 7, 'n_estimators': 100}	4.466
70	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.942
71	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	5.705

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

72	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.846
73	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.636
74	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.552
75	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.396
76	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.181
77	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.283
78	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.046
79	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.387
80	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.640
81	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.817
82	{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 200}	0.921
83	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.216
84	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 300}	8.367
85	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	15.550
86	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	16.115
87	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	10.802
88	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	9.862

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางภาคผนวก ก ที่ 9 ตารางค่าไฮเปอร์พารามิเตอร์ที่ดีที่สุด และค่า RMSE ของแต่ละเมนู ที่ได้จากตัวแบบXGBoost โดยใช้การคัดเลือกตัวแปรอิสระรูปแบบที่ 3

menu	Best Parameters	RMSE
0	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	8.785
1	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.975
2	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.022
3	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.145
4	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.965
5	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.152
6	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 200}	5.003
7	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.714
8	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.252
9	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.867
10	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.470
11	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.689
12	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.123
13	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.029
14	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.962
15	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	23.561
16	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.400
17	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.832
18	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	8.370
19	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	6.540
20	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.524
21	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.123

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.108
23	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.853
24	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.513
25	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.496
26	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.815
27	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.709
28	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.340
29	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.292
30	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.844
31	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.554
32	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.378
33	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.419
34	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 500}	0.579
35	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.674
36	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.125
37	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.692
38	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.082
39	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	25.248
40	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	16.262
41	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.202
42	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.884
43	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.655
44	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.892
45	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.615
46	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.894

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

47	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.526
48	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.672
49	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.098
50	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.367
51	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.061
52	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.411
53	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.547
54	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.533
55	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.270
56	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.362
57	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.428
58	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.753
59	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.114
60	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.445
61	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.981
62	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.870
63	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.428
64	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.770
65	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.863
66	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.642
67	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.092
68	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.857
69	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	4.353
70	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.932
71	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	6.022

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

72	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.851
73	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.685
74	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.653
75	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	3.070
76	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	2.141
77	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	1.012
78	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	1.082
79	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.368
80	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.630
81	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	0.871
82	{'learning_rate': 0.3, 'max_depth': 3, 'n_estimators': 200}	0.874
83	{'learning_rate': 0.05, 'max_depth': 5, 'n_estimators': 100}	2.311
84	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	8.761
85	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 200}	27.341
86	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	16.316
87	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	17.042
88	{'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100}	9.015

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

โค๊ดภาษา python ที่ใช้ในงานวิจัย

```
#เรียกใช้แพ็คเกจ
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OneHotEncoder
import statsmodels.api as sm
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
import xgboost as xgb
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objs as go
#นำเข้าข้อมูล
df_vis = pd.read_csv("C:/xxx/xxxxxx/xxxxxx/pivot_pilot_latest.csv")
pormotion_C = pd.read_excel("C:/xxx/xxxxxx/xxxxxx/Promotion_bakery_2021_24.xlsx")
#สำรวจข้อมูลสูญหาย
df_vis.isna().sum()
#เตรียมข้อมูล
#สร้างฟังชั้นเพื่อแปลงตัวแปรอิสระที่ใช้ด้วย one-hot encoder
def get_data_varidate_combine(df):
    data2 = pd.pivot_table(df, values='QUANT',index=['date','day_of_week','month',
                                                     'year','day_type','PromotionType',
                                                     'day_of_month'], aggfunc="sum")
    data2['Prev_Day_QUANT'] = data2['QUANT'].shift(1)
    data2 = data2.reset_index()
    encoder = OneHotEncoder()
    feature_groups = ['day_of_week', 'month', 'year', 'day_type','PromotionType']
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    encoded_vars = encoder.fit_transform(data2[feature_groups]).toarray()
    encoded_df = pd.DataFrame(encoded_vars,columns=encoder.get_feature_names_
    out(feature_groups), index=data2.index)
    data2=pd.concat([data2[['date']],encoded_df,data2[['day_of_month']],
                                data2[['Prev_Day_QUANT']],data2[['QUANT']]], axis = 1)
    return data2
#ฟังก์ชันในการคัดเลือกตัวแปรรูปแบบที่2และ3
def select_features_with_linear_regression(X_train, y_train):
    X_train_with_const = sm.add_constant(X_train)
    model = sm.OLS(y_train, X_train_with_const)
    results = model.fit()
    p_values = results.pvalues
    significant_features = p_values[p_values < 0.05].index.tolist()
    return significant_features
#ตัวแบบต้นไม้ตัดสินใจ
def find_best_model_DT (df_, random_state_):
    df_tmp = df_.copy()
    results = []
    for pr_code_use in df_tmp['PR_CODE'].unique(): #ทำการวนลูปทีละเมนู
        df_menu = df_tmp[df_tmp['PR_CODE'] == pr_code_use]
        menu_ = df_menu['ITEM_LONG_ENG'].unique()
        pivot_data_train = get_data_varidate_combine(df_menu) #ใช้ฟังก์ชันแปลงข้อมูล
        pivot_data_train = pivot_data_train.rename(columns={'date': 'ds', 'QUANT': 'y'})
        if len(pivot_data_train) > 60: #ตั้งเงื่อนไขว่าใช้แค่เมนูที่มีการขายมากกว่า 60 วัน
            time_frame = len(pivot_data_train)
            X_ = pivot_data_train.drop(columns=['ds', 'y'])
            y_ = pivot_data_train[['y']]
            #y_['y'] = np.log(y_['y']) #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่ 3 ลอการิทึมธรรมชาติ
            X_train, X_test, y_train, y_test = train_test_split(X_, y_, test_size=0.2,
          random_state=42)#แบ่งข้อมูลชุดฝึกและชุดทดสอบ
           #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่2และ3
            #significant_features = select_features_with_linear_regression(X_train,
            y_train)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        #X_train_selected = X_train[significant_features]
        #X_test_selected = X_test[significant_features]
        #ค้นหาพารามิเตอร์ที่ดีที่สุดด้วย GridSearch
        param_grid = { 'max_depth': [1, 3, 5, 10]  }
        grid_search=GridSearchCV(estimator=DecisionTreeRegressor(random_state=
                                 random_state_), param_grid=param_grid,
                                 cv=3,scoring='neg_root_mean_squared_error')
        grid_search.fit(X_train_selected, y_train)
        best_params = grid_search.best_params_
        #ใช้ค่าไฮเปอร์พารามิเตอร์ที่ได้กับตัวแบบต้นไม้ตัดสินใจ
        DT_model = DecisionTreeRegressor(random_state=random_state_,
                                         **best_params)
        DT_model.fit(X_train_selected, y_train)
        predictions = DT_model.predict(X_test_selected)
        #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่ 3 แปลงข้อมูลยอดขายรายวันกลับเป็นค่าเดิม
        #y_test_exp = np.exp(y_test)
        #predictions_exp = np.exp(predictions)
        #คิดค่า RMSE ของแต่ละเมนู
        rmse = np.sqrt(mean_squared_error(y_test_exp, predictions_exp))
        #เก็บข้อมูลลงใน results
        results.append({
          "PR_CODE": pr_code_use,
          "Menu": menu_,
          "RMSE": rmse,
          "Best Parameters": best_params,
         "Significant Features": significant_features #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่
2 และ 3 })
    results_df = pd.DataFrame(results)
    results_df['RMSE'] = results_df['RMSE'].astype(float)
    return results_df #แสดงผลข้อมูลในรูปแบบตารางโดยใช้ข้อมูลจาก results
#ตัวแบบป่าสุ่ม
def find_best_model_RF (df_, random_state_):
    df_tmp = df_.copy()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
results = []
for pr_code_use in df_tmp['PR_CODE'].unique(): #ทำการวนลูปทีละเมนู
    df_menu = df_tmp[df_tmp['PR_CODE'] == pr_code_use]
    menu_ = df_menu['ITEM_LONG_ENG'].unique()
    pivot_data_train = get_data_varidate_combine(df_menu) #ใช้ฟังก์ชันแปลงข้อมูล
    pivot_data_train = pivot_data_train.rename(columns={'date': 'ds', 'QUANT': 'y'})
    if len(pivot_data_train) > 60: #ตั้งเงื่อนไขว่าใช้แค่เมนูที่มีการขายมากกว่า 60 วัน
        time_frame = len(pivot_data_train)
        X_ = pivot_data_train.drop(columns=['ds', 'y'])
        y_ = pivot_data_train[['y']]
        #y_['y'] = np.log(y_['y']) #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่ 3 ลอการิทึมธรรมชาติ
        X_train, X_test, y_train, y_test = train_test_split(X_, y_, test_size=0.2,
       random_state=42)#แบ่งข้อมูลชุดฝึกและชุดทดสอบ
        #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่2และ3
        #significant_features = select_features_with_linear_regression(X_train,
        y_train)
        #X_train_selected = X_train[significant_features]
        #X_test_selected = X_test[significant_features]
        #ค้นหาพารามิเตอร์ที่ดีที่สุดด้วย GridSearch
        param_grid = { 'n_estimators': [100, 200, 300, 500],
                       'max_depth': [3, 5, 7, 10] }
       grid_search=GridSearchCV(estimator=RandomForestRegressor(random_state=
                                random_state_), param_grid=param_grid,
                                cv=3,scoring='neg_root_mean_squared_error')
        grid_search.fit(X_train_selected, y_train)
        best_params = grid_search.best_params_
        #ใช้ค่าไฮเปอร์พารามิเตอร์ที่ได้กับตัวแบบต้นไม้ตัดสินใจ
        RF_model = RandomForestRegressor(random_state=random_state_,
                                         **best_params)
        RF_model.fit(X_train_selected, y_train)
        predictions = RF_model.predict(X_test_selected)
        #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่ 3 แปลงข้อมูลยอดขายรายวันกลับเป็นค่าเดิม
        #y_test_exp = np.exp(y_test)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            #predictions_exp = np.exp(predictions)
            #คิดค่า RMSE ของแต่ละเมนู
            rmse = np.sqrt(mean_squared_error(y_test_exp, predictions_exp))
            #เก็บข้อมูลลงใน results
            results.append({
                "PR_CODE": pr_code_use,
                "Menu": menu_,
                "RMSE": rmse,
                "Best Parameters": best_params,
               "Significant Features": significant_features #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่
2 และ 3 })
    results_df = pd.DataFrame(results)
    results_df['RMSE'] = results_df['RMSE'].astype(float)
    return results_df #แสดงผลข้อมูลในรูปแบบตารางโดยใช้ข้อมูลจาก results
#ตัวแบบ XGBoost
def find_best_model_XGB (df_, random_state_):
    df_tmp = df_.copy()
    results = []
    for pr_code_use in df_tmp['PR_CODE'].unique(): #ทำการวนลูปทีละเมนู
        df_menu = df_tmp[df_tmp['PR_CODE'] == pr_code_use]
        menu_ = df_menu['ITEM_LONG_ENG'].unique()
        pivot_data_train = get_data_varidate_combine(df_menu) #ใช้ฟังก์ชันแปลงข้อมูล
        pivot_data_train = pivot_data_train.rename(columns={'date': 'ds', 'QUANT': 'y'})
        if len(pivot_data_train) > 60: #ตั้งเงื่อนไขว่าใช้แค่เมนูที่มีการขายมากกว่า 60 วัน
            time_frame = len(pivot_data_train)
            X_ = pivot_data_train.drop(columns=['ds', 'y'])
            y_ = pivot_data_train[['y']]
            #y_['y'] = np.log(y_['y']) #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่ 3 ลอการิทึมธรรมชาติ
            X_train, X_test, y_train, y_test = train_test_split(X_, y_, test_size=0.2,
           random_state=42) #แบ่งข้อมูลชุดฝึกและชุดทดสอบ
            #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่2และ3
            #significant_features = select_features_with_linear_regression(X_train,
            y_train)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
#X_train_selected = X_train[significant_features]
#X_test_selected = X_test[significant_features]
#ค้นหาพารามิเตอร์ที่ดีที่สุดด้วย GridSearch
param_grid = {
    'n_estimators': [100, 200, 300, 500],
    'max_depth': [3, 5, 7, 10],
    'learning_rate' : [0.05 , 0.1 , 0.3]
}
grid_search=GridSearchCV(estimator=xgb.XGBRegressor(objective ='reg:square
                                derror',eval_metric ='rmse',param_grid=param_grid,
                                cv=3,scoring='neg_root_mean_squared_error')
grid_search.fit(X_train_selected, y_train)
best_params = grid_search.best_params_
#ใช้ค่าไฮเปอร์พารามิเตอร์ที่ได้กับตัวแบบต้นไม้ตัดสินใจ
XGB_model = xgb.XGBRegressor(objective ='reg:squarederror',
                                eval_metric ='rmse' , **best_params)
XGB_model.fit(X_train_selected, y_train)
predictions = XGB_model.predict(X_test_selected)
#กรณีใช้การคัดเลือกตัวแปรรูปแบบที่ 3 แปลงข้อมูลยอดขายรายวันกลับเป็นค่าเดิม
#y_test_exp = np.exp(y_test)
#predictions_exp = np.exp(predictions)
#คิดค่า RMSE ของแต่ละเมนู
rmse = np.sqrt(mean_squared_error(y_test_exp, predictions_exp))
#เก็บข้อมูลลงใน results
results.append({
    "PR_CODE": pr_code_use,
    "Menu": menu_,
    "RMSE": rmse,
    "Best Parameters": best_params,
   "Significant Features": significant_features #กรณีใช้การคัดเลือกตัวแปรรูปแบบที่
2และ 3    })
results_df = pd.DataFrame(results)
results_df['RMSE'] = results_df['RMSE'].astype(float)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    return results_df #แสดงผลข้อมูลในรูปแบบตารางโดยใช้ข้อมูลจาก results
#ฟังก์ชั่นในการทำนายยอดขายรายวัน7วันข้างหน้า
def predict_pilot_3(df_full,start_date,end_date,promotion_df,res_rfr_cht4):
  format_ = df_full['Format'].unique()[0]
  prediction_df = pd.DataFrame()
   #ทำการวนลูปทุกเมนู
  for pr_code_use in df_full['PR_CODE'].unique():
    df_train = df_full[df_full['PR_CODE']==pr_code_use]
    menu_ = df_train['ITEM_LONG_ENG'].unique()[0]
    pivot_data_train = get_data_train(df_train)
    pivot_data_train = pivot_data_train.rename(columns={'date': 'ds', 'QUANT': 'y'})
    #ตั้งเงื่อนไขเลือกเฉพาะเมนูที่มีการขายมากกว่า 60 วัน
    if len(pivot_data_train)>60:
      X = pivot_data_train.drop(columns=['ds', 'y'])
      y = pivot_data_train[['y']]
      X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
                                                          random_state=42)
      #ใช้ตัวแปรอิสระและไฮเปอร์พารามิเตอร์ที่ดีที่สุดจากตัวแบบป่าสุ่มที่ใช้การคัดเลือกตัวแปรแบบ
ที่2
      best_params = res_rfr_cht4[res_rfr_cht4['PR_CODE'] == pr_code_use]
                                 ['Best Parameters'].iloc[0]
      best_n_estimators = best_params['n_estimators']
      best_max_depth = best_params['max_depth']
      # best_learning_rate = best_params['learning_rate']
      selected_features = res_rfr_cht4[res_rfr_cht4['PR_CODE'] == pr_code_use]
['Significant Features'].iloc[0]
      X_train_selected = X_train[selected_features]
      #ฝึกสอนตัวแบบRandomForestRegressor
      model = RandomForestRegressor(n_estimators=best_n_estimators,
                                    max_depth=best_max_depth, random_state=123)
      model.fit(X_train_selected, y_train)
      df_predict = create_date_df(start_date, end_date, public_holidays, festival
                                  ,promotion_df, pr_code_use, format_)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    df_predict = get_data_predict(df_predict)
    df_predict.at[0, 'Prev_Day_QUANT'] = pivot_data_train.iloc[-1]['y']
    df_predict = df_predict.rename(columns={'date': 'ds', 'QUANT': 'y'})
    df_predict['Menu'] = menu_
    df_predict['PR_CODE'] = pr_code_use
    X_ = df_predict.drop(columns=['ds', 'y', 'Menu', 'PR_CODE'])
    X_ = X_[selected_features]
    X_pre = X_[:1]
    #ทำนายยอดขายรายวัน
    predictions = model.predict(X_pre)
    df_predict.at[0, 'y'] = predictions
    for i in range(1, len(df_predict)):
      if 'Prev_Day_QUANT' in selected_features:
        df_predict.at[i, 'Prev_Day_QUANT'] = df_predict.iloc[i-1]['y']
        X_.at[i, 'Prev_Day_QUANT'] = df_predict.iloc[i-1]['y']
      X_pre = X_[i:i+1]
      predictions = model.predict(X_pre)
      df_predict.at[i, 'y'] = predictions
    prediction_df = pd.concat([prediction_df, df_predict])
  else: pass
prediction_df['ds'] = prediction_df['ds'].dt.date.astype(str)
pivot_table = prediction_df.pivot_table(index=['PR_CODE', 'Menu'], columns='ds',
                                        values='y', aggfunc='sum').reset_index()
return pivot_table
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

งานทะเบียนคณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
คำรับรองเล่มโครงงานพิเศษ/ปัญหาพิเศษ/สหกิจศึกษา

วันที่ 11 เดือน พฤษภาคม พ.ศ 2567

ข้าพเจ้า นาย ธนกฤต เปาวสันต์ รหัสประจำตัว 63050625
นักศึกษาหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชา สถิติประยุกต์ ภาควิชา สถิติ ขอรับรองว่าโครงงานสหกิจศึกษา เรื่อง

ชื่อภาษาไทย	การพยากรณ์จำนวนการขายสินค้าของร้านขายเบเกอรี่ กรณีศึกษาร้านเบเกอรี่แห่งหนึ่ง
ชื่อภาษาอังกฤษ	Forecasting Bakery Product Sales: A Case Study of a Bakery Shop
ปีการศึกษา	2566

เป็นผลงานวิจัยที่มิได้คัดลอกหรือละเมิดลิขสิทธิ์ของผู้อื่นและได้ผ่านการตรวจสอบความซ้ำซ้อนเรียบร้อยแล้วและได้แนบเอกสารการตรวจสอบการลอกเลียนงานวรรณกรรมที่ตรวจสอบจากเล่มโครงงานสหกิจศึกษาฉบับสมบูรณ์แล้ว
โปรแกรมอักขราวิสุทธิ์ 1.08 %

ลงชื่อ.....
(นายธนกฤต เปาวสันต์)
นักศึกษา

ข้าพเจ้า รศ.ดร วลัยลักษณ์ อัตธีรวงศ์ อาจารย์ที่ปรึกษาโครงงานสหกิจศึกษา ได้ตรวจสอบโครงงานสหกิจศึกษาของนักศึกษาข้างต้น แล้ว ขอรับรองว่าเป็นผลงานวิจัยของนักศึกษาจริงและมีเนื้อหาสมบูรณ์ จึงลงชื่อไว้เป็นหลักฐาน

ลงชื่อ.....
อาจารย์ที่ปรึกษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Submission Information

ID	SUBMISSION DATE	SUBMITTED BY	ORGANIZATION	FILENAME	STATUS	SIMILARITY INDEX
3734314	May 11, 2024 at 05:52 AM	63050625@kmitl.ac.th	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง	ธนกฤต_เปาวสันต์_63050625.pdf	Completed	1.08 %

Match Overview

Show 10 entries Search:

NO.	TITLE	AUTHOR(S)	SOURCE	SIMILARITY INDEX
1	ระบบการพยากรณ์ราคาข้าวเปลือกเจ้านาปี 5 เปอร์เซ็นต์ โดยใช้โครงข่ายใยประสาทเทียมผ่านเครือข่ายอินเทอร์เน็ต	สุวรรณ บุญวิจิตร	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ	0.26 %
2	การกำหนดเป้าหมายยอดขายที่เหมาะสมด้วยเทคนิคการวิเคราะห์อนุกรมเวลา : กรณีศึกษา ยาของบริษัทยาแห่งหนึ่ง	สุพิชญา พันธุมจินดา	มหาวิทยาลัยธุรกิจบัณฑิตย์	0.24 %
3	PERCEPTION SOCIAL CARRYING CAPACITY: SMALL BUSINESS FOR REAL ESTATE DEVELOP CASE STUDY: HAT CHA-AM AREA	ศิริจานุสรณ์, อรุณ	วารสารศรีปทุมปริทัศน์ ฉบับมนุษยศาสตร์และสังคมศาสตร์	0.10 %
4	THE ENTERING OF LIFELONG LEARNING OF THE ELDERLY PEOPLE IN THE NORTHEASTERN THAILAND	ธีระฐิติ, นภเรณู สัจจรักษ์	วารสารศรีปทุมปริทัศน์ ฉบับมนุษยศาสตร์และสังคมศาสตร์	0.09 %
5	http://regist.pn.psu.ac.th/documents/parinya53/name/pedu4.pdf	regist.pn.psu.ac.th	regist.pn.psu.ac.th_nutch	0.08 %
6	https://mathematics.sc.mahidol.ac.th/wp-content/uploads/2018/12/ผังห้อง-L-01.pdf	mathematics.sc.mahidol.ac.th	mathematics.sc.mahidol.ac.th_nutch	0.07 %
7	Results of health products knowledge on consumption behavior Ban Mai Patthana community, Nam Kien sub district, Phupieng district, Nan province	อิศรางกูร ณ อยุธยา, เจริญลักษณ์	วารสารการคุ้มครองผู้บริโภคด้านสุขภาพ	0.07 %

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้