Journal of Information Science and Technology

JIST Journal of Information Science and Technology
Volume 13, NO 2 | JUL – DEC 2023 | 1-12
ISSN: 2651-1053 (Online)
ISSN: 1906-9553 (Print)

การศึกษาการเรียนรู้ของเครื่องสำหรับการทำนายราคาน้ำมันปาล์มดิบ

A Study of Machine Learning for Crude Palm Oil Price Prediction

*กัญญา ศิริภิรมย์ และ ดัชกรณ์ ตันเจริญ**

*Kanya Siripirom and Datchakorn Tancharoen**

คณะวิศวกรรมศาสตร์และเทคโนโลยี สถาบันการจัดการปัญญาภิวัฒน์

Faculty of Engineering and Technology, Panyapiwat Institute of Management

Received: June 10, 2023; Revised: December 14, 2023; Accepted: December 14, 2023; Published: December 29, 2023

ABSTRACT – This paper focuses on Thailand's Crude Palm Oil price prediction. By searching the best model from a comparative study between the Baseline Model, RNN-LSTM, and the Classic models. The Classic models in this paper are Exponential Smoothing, Linear Regression, Random Forest, LightGBM Model, and ARIMA. The prediction of crude palm oil prices uses time series. And for the evaluation uses the Mean Absolute Percentage Error (MAPE) metric to measure the percentage deviation between the projected price and actual prices in order to provides the insights performance of the model. This research found that the Linear Regression model has the lowest MAPE of 5.71. Therefore, the Linear Regression is the best accuracy for forecasting crude palm oil prices in Thailand.

KEYWORDS: Baseline Model, LSTM, Classic Model, MAPE, Price Prediction

บทคัดย่อ -- บทความนี้มุ่งเน้นไปที่การคาดการณ์ราคาน้ำมันปาล์มดิบในประเทศไทยเพื่อหาโมเดลที่ดีที่สุดในการทำนายราคา ในการศึกษานี้เป็นการเปรียบเทียบแบบจำลอง 3 แบบ ได้แก่ โมเดลพื้นฐาน (Baseline Model) โมเดล RNN-LSTM และโมเดลคลาสสิก (Classic Model) เพื่อทำนายราคาน้ำมันปาล์มดิบในประเทศไทย โดยโมเดลคลาสสิก ได้แก่ Exponential Smoothing, Linear Regression, Random Forest, LightGBM Model และ ARIMA การทำนายราคาของน้ำมันปาล์มดิบใช้อนุกรมเวลาในการทดสอบ ในประเมินประสิทธิภาพนั้นจะใช้ Mean Absolute Percentage Error (MAPE) เพื่อวัดเปอร์เซ็นต์ความเบี่ยงเบนระหว่างราคาที่คาดการณ์และราคาที่เป็นจริงเพื่อทราบประสิทธิภาพของแบบจำลองเชิงลึก ผลการศึกษาแสดงให้เห็นว่าโมเดล Linear Regression มีค่า MAPE ต่ำที่สุดอยู่ที่ 5.71 ดังนั้นแบบจำลอง Linear Regression จึงมีประสิทธิภาพดีที่สุดในการพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทยในงานวิจัยนี้

คำสำคัญ: โมเดลพื้นฐาน, LSTM, โมเดลคลาสสิก, MAPE, การพยากรณ์ราคา

**Corresponding Author: datchakorntan@pim.ac.th*

1. บทนำ

น้ำมันปาล์มดิบเป็นสินค้าสำคัญที่มีการใช้งานหลากหลายในอุตสาหกรรมต่าง ๆ รวมถึงอาหาร เครื่องสำอาง และเชื้อเพลิงชีวภาพ อุตสาหกรรมน้ำมันปาล์มมีบทบาทสำคัญในเศรษฐกิจโลก โดยเฉพาะอย่างยิ่งในประเทศไทย ซึ่งการผลิตและการส่งออกน้ำมันปาล์มมีส่วนสำคัญต่อเศรษฐกิจของประเทศ ด้วยธรรมชาติที่ผันผวนของตลาดสินค้าโภคภัณฑ์ การคาดการณ์ราคาน้ำมันปาล์มดิบที่แม่นยำ จึงกลายเป็นสิ่งจำเป็นสำหรับผู้มีส่วนได้ส่วนเสียในอุตสาหกรรมน้ำมันปาล์มในการตัดสินใจในการจัดการความเสี่ยง และเพิ่มประสิทธิภาพการดำเนินงาน มีงานวิจัยหลากหลายที่ทำการคาดการณ์ราคาแบบดั้งเดิม เช่น แบบจำลองอนุกรมเวลาทางสถิติ [1] จะจับลักษณะที่ซับซ้อนและเปลี่ยนแปลงตลอดเวลาของราคาน้ำมันปาล์มดิบ เนื่องจากมีปัจจัยที่มีอิทธิพลหลายอย่าง ส่งผลให้มีความสนใจในการใช้ประโยชน์จากเทคนิคการเรียนรู้ของเครื่องเพิ่มขึ้นเพื่อปรับปรุงความแม่นยำของการคาดการณ์ราคาในอุตสาหกรรมน้ำมันปาล์ม มีงานวิจัยที่ทำการเปรียบเทียบผลการคาดการณ์ ได้แก่ Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN) และ Hybrid ARIMA-ANN [2] เป็นต้น การเรียนรู้ของเครื่องนำเสนอศักยภาพในการแสดงรูปแบบและความสัมพันธ์ของข้อมูล ทำให้สามารถคาดการณ์ได้แม่นยำและมีประสิทธิภาพมากขึ้นในการพยากรณ์ราคา รวมไปถึงการพยากรณ์ราคาอื่น ๆ เช่น ราคาน้ำมันดิบโลก โดยใช้ Long Shot Term Memory (LSTM) การใช้การเรียนรู้เชิงลึก (Deep Learning) เพื่อกำหนดรูปแบบการคาดการณ์ราคาน้ำมันดิบ แล้วแสดงผลออกมาเป็นกราฟเพื่อแสดงให้เห็นถึงความซับซ้อนในการวิเคราะห์ความผันผวนของราคาน้ำมันดิบ [3] นอกจากนี้มีการทดลองโดยเพิ่ม Layer ของ LSTM การเพิ่ม Layer ปรับเปลี่ยน Loopback และการปรับเปลี่ยนจำนวน Epoch เพื่อเพิ่มความแม่นยำในการทำนายราคาน้ำมันดิบ [4] มีงานวิจัยอื่น ๆ ที่ใช้แบบจำลอง Deep Learning ควบคู่กับโมเดล Deep Belief Network (DBN), Long Shot Term Memory (LSTM), Random Walk (RW), Autoregressive Moving Average (ARMA) และ Hybrid Model เช่น RW-DBN, ARMA-DBN, RW-LSTM, ARMA-LSTM ด้วย [5] นอกจากนั้นมีโมเดลที่ชื่อว่า VMD-RE-LSTM ที่แสดงให้เห็นว่าเป็นโมเดลที่สามารถทำนายราคาน้ำมันได้ดีกว่า LSTM [6] อย่างไรก็ตาม LSTM เป็นที่นิยมในการทำนายราคาน้ำมันดิบเพื่อเพิ่มความแม่นยำ และเสถียรภาพระยะยาวจึงนำ LSTM และ Facebook's Prophet มาใช้ในการประมาณข้อมูล Time Series ที่พัฒนาโดย Facebook ซึ่งมีความสามารถในการทำนายข้อมูลราคาน้ำมันเบรนท์ ซึ่ง ผลการทดลองชี้ให้เห็นว่ามีความแม่นยำสูง [7] การคาดการณ์ราคาน้ำมันดิบมีวิธีใหม่เพิ่มขึ้นอย่างต่อเนื่อง เช่น Support Vector Machine (SVM) เป็นขั้นตอนการพัฒนาโมเดลเครื่องเวกเตอร์พอร์ตสนับสนุนสำหรับการคาดการณ์อนุกรมเวลา เกี่ยวข้องกับการสุ่มตัวอย่างข้อมูล โดย เปรียบเทียบกับ Autoregressive Integrated Moving Average (ARIMA) และ Back Propagation Neural Network (BPNN) ซึ่งผลการทดลองนั้นแสดงให้เห็นว่า SVM สามารถเป็นตัวเลือกที่ดีสำหรับการพยากรณ์ราคาน้ำมันเมื่อแปรผันตามเวลา (Time Series) [8] นอกจากนี้มีการพยากรณ์ราคาน้ำมันปาล์มในประเทศอื่นเช่นมาเลเซีย และใช้โมเดลอื่นๆ ได้แก่ Autoregressive Distributed Lag (ARDL), Autoregressive Integrated Moving Average (ARIMA) และ ARIMAX [9] และโมเดลอื่นโดยใช้อนุกรมเวลา เช่น Exponential Smoothing เพื่อหาค่า Mean Absolute Percentage Error (MAPE) เปรียบเทียบกับ Regression เพื่อหาค่า Error เปรียบเทียบความแม่นยำในการพยากรณ์การผลิตน้ำมันปาล์มอีกด้วย [10] ไปใช้ในการทำนายราคาและทำนายเทรนของหุ้น โดยใช้ WEKA Software และสามารถ ปรับปรุงเทคนิค Regression เพื่อปรับปรุงการทำนายราคาหุ้นให้แม่นยำขึ้นโดยใช้เพียงชุดข้อมูลต้นฉบับ [11] นอกจากนี้การพยากรณ์ราคาบ้านและนำมาเปรียบเทียบในแต่ละโมเดลหรือหลายวิธีการนั้น โดยมีการใช้ Neural Network, Random Forests, Inverse Distance Weighting, และ Kriging โดยพบว่า Neural Network และ Random Forests สามารถทำนายราคาได้แม่นยำ [12] และมีการนำ LSTM กับ LightGBM มาผสมผสานแล้วบันทึกผลการฝึกโมเดล ราคาเปิด ราคาปิดทั้งสูงสุดและต่ำสุด ปริมาณการซื้อขาย แล้วมาทำนายราคาผันผวนของหุ้น [12] การพยากรณ์ราคาหุ้นเป็นงานที่ท้าทาย แสดงถึงคุณสมบัติที่แบบจำลองเดียวทำไม่ได้ มีนักวิจัยหลายคนนำแบบจำลองต่าง ๆ มาผสมผสานกันระหว่างโดยนำวิธี Empirical Mode Decomposition (EMD) และวิธี Theta โดยพิจารณาเพื่อเพิ่มศักยภาพในการพยากรณ์ที่ดี [13] โดยในการทดสอบแบบจำลองเพื่อทำนายราคาที่เกี่ยวกับอนุกรมเวลา สามารถใช้ฟังก์ชันของ Dart เพื่อสร้างโมเดลคลาสสิกได้อย่างหลากหลายเพื่อใช้ในทดสอบข้อมูลอนุกรมเวลา [14] ในการ

พยากรณ์ราคาด้วยโมเดลที่หลากหลายนั้น มีนักวิจัยที่นำโมเดลที่เป็น Traditional Statistic เช่น โมเดลพื้นฐานที่มีตัวแปรเกี่ยวกับเวลามาเปรียบเทียบกับ Machine Learning เพื่อทำนายราคาหุ้นในอนาคต โดยนำโมเดลต่าง ๆ เช่น Simple Moving Average, Weighted Moving Average, Exponential Smoothing, Naive Model และ Machine Learning เช่น Linear Regression, Lasso, Ridge, K-Nearest Neighbors, Support Vector Machine, Random Forest, Single Layer Perceptron, Multi-Layer Perceptron, Long-Short Term Memory มาใช้ในการพยากรณ์ราคาและพบว่า Long-Short Term Memory ให้ผลการทำนายที่ดีที่สุด [15] อย่างไรก็ตามก็มีวรรณกรรมที่นำแบบจำลองการเรียนรู้ของเครื่องมาใช้ในการทำนายราคาในอุตสาหกรรมเหมืองแร่ โดยนำ RNNL-STM และ Traditional Model เช่น ARIMA มาทำการทำนายราคาของหินปูน (Limestone) ซึ่งเป็นวัตถุดิบที่สำคัญตัวหนึ่งที่ใช้สำหรับอุตสาหกรรมเซรามิก แล้วมาเปรียบเทียบความแม่นยำของแบบจำลอง ผลปรากฏว่าใน ARIMA ก็มีความแม่นยำที่ดีกว่า LSTM และใช้เวลาในการเทรนข้อมูลที่น้อยกว่า LSTM อีกด้วย [16] ยังมีวรรณกรรมอื่นๆอีกมากมายที่นำ LSTM เปรียบเทียบประสิทธิภาพกับ Linear Regression เช่นในการทำนายดัชนีราคาของหุ้น ผลปรากฏว่า Linear Regression ก็ให้ค่าความแม่นยำถึง 98% เมื่อเทียบกับ LSTM ที่ให้ค่าแม่นยำเพียง72.3% [17] เป็นที่น่าสังเกตได้ว่าในบางครั้งการใช้ Deep Learning RNN-LSTM ก็อาจจะไม่ได้ ผลลัพท์ที่ดีกว่า Traditional Model แม้ว่าจะมีเทรนข้อมูลโดยใช้เวลานานและมีจำนวน Architecture หรือโมเดลที่ซับซ้อน แต่ก็มีวรรณกรรมไม่น้อยที่ทำการทดสอบวัดความแม่นยำเทียบกับ LSTM และ LSTM ให้ผลดีกว่าแบบจำลองอื่น ดังนั้นผลการทดลองจากวรรณกรรมต่างๆ ผลลัพท์อาจจะขึ้นอยู่กับปัจจัย ชุดของมูล หรือแบบจำลองของแต่ละแบบทดสอบในวรรณกรรมนั้นๆได้อีกด้วย ซึ่งในบทความนี้มุ่งเน้นไปที่งานคาดการณ์ราคาน้ำมันปาล์มดิบในประเทศไทยโดยใช้โมเดลการเรียนรู้ของเครื่องโดยเปรียบเทียบประสิทธิภาพของโมเดลพื้นฐาน โมเดล RNN-LSTM และโมเดลคลาสสิก 5 แบบ

ส่วนถัดไปของบทความคือ ส่วนที่ 2 แบบจำลองและทฤษฎีที่เกี่ยวข้อง ส่วนที่ 3 วิธีการดำเนินการวิจัย ส่วนนี้อธิบายถึงการรวบรวมข้อมูล การประมวลผล และการนำโมเดลการเรียนรู้ของเครื่องไปใช้ ส่วนที่ 4 แสดงผลการทดลองและการวิเคราะห์เปรียบเทียบประสิทธิภาพของแบบจำลอง ส่วนสุดท้ายเป็นบทสรุป การอภิปราย และทิศทางการวิจัยในอนาคต

2. แบบจำลองและทฤษฎีที่เกี่ยวข้อง

ปัญญาประดิษฐ์ (Artificial Intelligence) คือ เทคโนโลยีที่ออกแบบให้มีระบบทำงานเหมือนกับการทำงานของสมองมนุษย์ ปัญญาประดิษฐ์ถูกพัฒนาให้ฉลาดมากขึ้น ทำงานได้หลากหลายมากขึ้น และมีความสามารถใกล้เคียงกับมนุษย์มากยิ่งขึ้น เช่น Machine Learning (ML) การป้อนข้อมูลเพื่อให้ระบบคอมพิวเตอร์ทำการเรียนรู้กับชุดข้อมูลเพื่อแสดงผลตามที่ต้องการ และมีการเรียนรู้เชิงลึก (Deep Learning) โดย Deep Learning คือศาสตร์แขนงหนึ่งของ Machine Learning ที่เลียนแบบระบบเซลล์ประสาทในสมองของมนุษย์ (Neural Network) ชนิดของโครงข่าย Deep Learning จะมีโครงข่ายประสาทแบบป้อนไปหน้า (Feed-Forward Neural Networks) และโครงข่ายแบบวนซ้ำ Recurrent Neural Networks (RNN)

2.1 แบบจำลองพื้นฐาน

ในการทำนายราคานั้นการที่จะกำหนดการทำนายราคาพื้นฐานที่ใช้ในการทำนายราคาโดยแบบจำลองพื้นฐานเริ่มต้นที่กำหนดเป็น Baseline เมื่อเริ่มต้นงานหรือโครงการต่างๆ ก็เริ่มจากการสร้างโมเดลที่ง่ายที่สุดก่อนเพื่อแก้ปัญหานั้นๆ อาจจะกำหนดจากบรรทัดฐานหรือวิธีการปฏิบัติดั้งเดิมที่ดีอยู่แล้ว ยกตัวอย่างเช่นตั้งต้นจากสิ่งที่ง่ายที่สุดที่ใช้ในการทำราคา เช่น นำราคาเดือนที่แล้วมาเพื่อนำไปสู่ผลลัพท์ในการทำนายราคาที่มีประสิทธิภาพของการทำงานด้าน Data science มากขึ้น สามารถปรับแต่งหรือเพิ่มตัวแปรอื่นๆหรือนำโมเดลอื่นๆที่มีความซับซ้อนมากกว่า แล้วหลังจากนั้นจึงพิจารณาต่อว่าจะมีโมเดลอื่นๆมาทำนายราคามีประสิทธิภาพดีกว่าหรือไม่

2.1.1 Last Month Prediction คือโมเดลที่ง่ายที่สุดที่กำหนดเป็นโมเดลพื้นฐานในการทำนายราคาในเดือนถัดไปและในโมเดลถัดไปที่จะกำหนดมาเป็นโมเดลพื้นฐานอื่นๆก็จะมีการปรับด้วยเปอร์เซ็นการเปลี่ยนแปลงราคาของเดือนก่อนหน้าหรือเพิ่มความซับซ้อนอื่นๆเพิ่มเติมจากโมเดลพื้นฐานนี้

2.1.2 Simple Moving Average (SMA) คือ โมเดลที่ทำการคำนวนค่าเฉลี่ยแบบเคลื่อนที่บนคาบเวลาที่กำหนด โดยใช้ข้อมูลย้อนหลังมาคำนวณค่าเฉลี่ยโดยทุกค่าข้อมูลจะถูกให้ความน้ำ หนักทุกตัวเท่ากัน เช่น

กำหนด T=3

โดย Raw = [1,2,3,4,5,6,7,8,9,10]

จะได้ SMA = [NA, NA, 2, 3, 4,5,6,7,8,9]

ดังนั้นโมเดลพื้นฐานที่จะมาใช้การทำนายอาจจะการนำ Last month รวมกับ Simple Moving Average และมีการปรับด้วยเปอร์เซ็นการเปลี่ยนแปลงราคาของเดือนก่อนหน้าได้หรือหลายเดือนก่อนหน้าก็ได้ หรือกับ Simple Moving Average โดยกำหนด T หรือจำนวนเดือนก่อนหน้าได้เพื่อกำหนดเป็นโมเดลพื้นฐานถัดไป

2.1.3 Exponential Moving Average คือ โมเดลการคำนวณค่าเฉลี่ยแบบเคลื่อนที่บนคาบเวลาที่กำหนดแต่มีการให้ค่าน้ำหนักของค่าล่าสุดมากกว่าค่าอื่นๆ

$$EMA = Pricetoday\left(\frac{2}{1+Days}\right) + EMAPrice\left(1 - \frac{2}{1+Days}\right) \quad (1)$$

โดยมี $\frac{2}{1+Days}$ เป็น Smoothing Factor จะดูว่าจำนวนเวลาที่ต้องการจะดูย้อนหลังไปเช่น 2 วัน

ในส่วน $EMAPrice$ เป็นข้อมูลของวันก่อนหน้า

ถ้ามี Raw = [1,2,3,4,5,6,7,8,9,10]

เช่นต้องดูราคาย้อนหลังไป 3 วัน Smoothing Factor = 0.5

EMA1 = 1

EMA2 = (2 × 0.5) + (1 × (1 - 0.5)) = 1.5

EMA3 = (3 × 0.5) + (1.5 × (1 - 0.5)) = 2.25

EMA4 = (4 × 0.5) + (2.25 × (1 - 0.5)) = 3.125 เป็นต้น

ดังนั้นโมเดลพื้นฐานต่อมาอาจจะเป็น Exponential Moving Average ของ 3 เดือนก่อนหน้าได้

2.2 แบบจำลอง RNN-LSTM

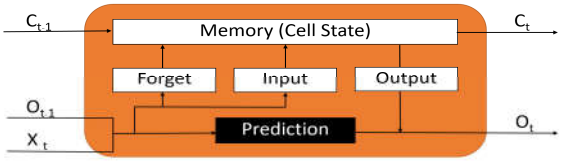
Long Short-Term Memory (LSTM) จัดว่าเป็นโครงข่ายประเภท Recurrent Neural Network (RNN) ซึ่งสามารถจำลองลำดับของข้อมูลได้ เช่น การทำนายคำต่อไปในประโยค การทำนายอากาศในวันถัดไป ทำนายสีต่อไปเพื่อไล่ลำดับสี โดยใช้ชุดข้อมูลตามอนุกรมเวลา (Time Series) เพื่อทำ Data Training และนำผลที่ได้ (Output) ของ Period นี้นำไปเป็น Input ของ Period หน้า ซึ่ง RNN ไม่สามารถเรียนรู้ความสัมพันธ์ระยะยาวได้ เนื่องจากไม่มีการเก็บข้อมูลใน Memory ดังนั้น LSTM เปรียบเสมือนโครงข่ายประสาทเทียมแบบหนึ่งที่ถูกออกแบบมาสำหรับการประมวลผลลำดับโดยเป็นเครือข่ายประสาทที่เกิดซ้ำและได้รับการแก้ไข ซึ่งช่วยให้จดจำข้อมูลที่ผ่านมาในหน่วยความจำได้ง่ายขึ้น กล่าวคือ LSTM คือ RNN ที่มีหน่วยความจำเพิ่มเข้ามาและเรียกว่า Cell State เพื่อเก็บข้อมูลทุกๆช่วงเวลา และมีกระบวนการเพิ่มการตัดสินใจว่าข้อมูลประเภทใดที่ควรจะเป็น Output กระบวนการที่เพิ่มขึ้นมาใน LSTM ดังนี้

1) ฟังก์ชั่นที่ตัดสินใจว่าข้อมูลใดควรจะลบออกจากหน่วยความจำ (Forgetting Gate)

2) ฟังก์ชั่นที่ตัดสินใจว่าข้อมูลใหม่ข้อมูลใดที่ควรเพิ่มในหน่วยความจำ (Input Gate)

3) ฟังก์ชั่นที่ตัดสินใจว่าข้อมูลใดในหน่วยความจำมาส่งผลกระทบกับข้อมูลที่จะทำนาย (Output Gate)

ซึ่งกระบวนการของ LSTM คือจะมีข้อมูลที่เป็น O_{t-1} และ X_t เป็นข้อมูล Input โดยที่ O_{t-1} เป็น Output จาก Period ที่แล้ว และ X_t คือ Input ใน Period นี้ และมี C_{t-1} เป็น Input สำหรับหน่วยความจำ ดังรูปที่ 1



รูปที่ 1. กระบวนการทำงานของ LSTM

จากทฤษฎีของ LSTM ได้มีการพัฒนาให้มีหน่วยความจำเพิ่มขึ้นมาซึ่งเป็นข้อดีและเป็นไปได้ว่าเป็นโมเดลที่มีประสิทธิภาพในการทำนายราคามากกว่าโมเดลอื่น ๆ

2.3 แบบจำลองคลาสสิก (Classic Model)

คือแบบจำลอง Traditional หรือแบบจำลองคลาสสิกอื่นๆที่นำมาใช้ในงานวิจัยนี้และเป็นแบบจำลองในการทดสอบเพื่อศึกษาเป็นแนวทางว่าแบบจำลอง Traditional อื่นๆ อาจจะทำงานได้มีประสิทธิภาพทำนายราคาได้ดี เช่นเดียวกับการใช้แบบจำลองพื้นฐานและ RNN-LSTM ก็เป็นได้ โดยแบบจำลองคลาสสิกที่เลือกใช้ในงานวิจัยนี้ได้แก่

2.3.1 Exponential Smoothing เป็นวิธีที่ใช้หลักการเดียวกันกับการพยากรณ์แบบค่าเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก คือ ให้ความสำคัญกับข้อมูลชุดใหม่ ที่สุดมากที่สุด (ค่าถ่วงน้ำหนักสูงสุด) และค่อยๆ ลดค่า ถ่วงน้ำหนักลง วิธีเอกซ์โปเนนเชียลปรับเรียบจะทำการพยากรณ์โดยนำค่าพยากรณ์ ของช่วงเวลาที่ผ่านมาบวกเข้ากับอัตราส่วนความแตกต่างระหว่างข้อมูลจริงกับค่า พยากรณ์ ณ ช่วงเวลานั้นๆ ดังแสดงตามสมการที่ (2)

$$F(t+1) = \alpha * Y(t) + (1-\alpha) * F(t) \quad (2)$$

โดย F(t+1) คือ การคาดการณ์สำหรับช่วงเวลาถัดไป

Y(t) คือ การสังเกตจริง ณ เวลา t

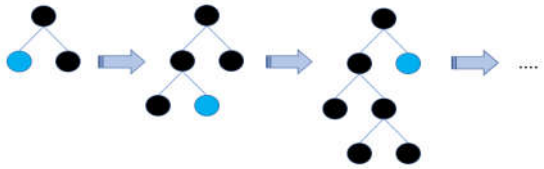
F(t) คือ การคาดการณ์สำหรับช่วงเวลาปัจจุบัน

α คือ ปัจจัยการปรับให้เรียบ ซึ่งจะกำหนดน้ำหนักให้กับการสังเกตปัจจุบันเทียบกับการคาดการณ์ก่อนหน้า

2.3.2 Linear Regression Linear Regression ที่มาจากคำว่า Linear หมายความว่าเส้นตรง และคำว่า Regression ที่หมายความว่าถดถอย ซึ่งก็คือการนำเอาข้อมูลหรือตัวแปรมาหาความสัมพันธ์กัน แล้วโดยใช้ความสัมพันธ์หรือการทำนายข้อมูลจะออกมาใกล้เคียงกับเส้นตรง โดย Linear Regression ถือว่าเป็น Machine Learning การเรียนรู้ที่มีผู้สอนประเภท (Supervised Learning) ชนิดแบบ Statistical Regression ที่จำเป็นต้องป้อนชุดข้อมูลเข้าไปให้เรียนรู้ก่อน หลังจากเรียนรู้แล้วจะนำตัวแปรต้นและตัวแปรตามไปคำนวณด้วยสถิติทางคณิตศาสตร์ แล้วจะได้ข้อมูล Output ออกมาเป็นตัวเลข เป็นการวิเคราะห์การถดถอยเชิงเส้น โดยสามารถนำ Linear Regression มาทำการทำนายราคาสินค้าได้ โดยกำหนดชุดข้อมูลตัวอย่างเช่นราคาเป็นข้อมูลระยะเวลาเท่าไร จากนั้นก็ทำการเทรนโมเดลแล้วหาค่าความสัมพันธ์ของ Output ว่าแบบจำลองนั้นได้ค่าดีหรือไม่ดี

2.3.3 Random Forest เป็นกลุ่มของโมเดลที่มีหลักการในการเทรนโมเดลที่เหมือนกันหลายๆครั้งบนชุดข้อมูลเดียวกัน โดยแต่ละครั้งในการเทรนจะเลือกส่วนของข้อมูลที่เทรนแตกต่างกัน และตัดสินใจโดยการโหวดว่าตัวไหนถูกเลือกมากที่สุด อัลกอริธึมนี้เป็นการเรียนรู้ของเครื่องยอดนิยมที่ใช้สำหรับทั้งการจัดหมวดหมู่และการถดถอย เป็นวิธีการทั้งมวลที่รวมการคาดคะเนของแผนผังการตัดสินใจหลาย ๆ แบบเพื่อให้การคาดคะเนแม่นยำและมีประสิทธิภาพมากขึ้น สร้างโมเดล จาก Decision Tree หลายๆโมเดลย่อยๆ โดยแต่ละ โมเดลจะได้รับ Data set ไม่เหมือนกัน ซึ่งเป็น Subset ของ Data set ทั้งหมด และทำหน้าที่ Prediction ให้แต่ละ Decision Tree ของตนเอง และคำนวณผล Prediction ด้วยการ Vote output ที่ ถูกเลือกโดย Decision Tree มากที่สุด (กรณี Classification) หรือ หาค่า Mean จาก Output ของแต่ละ Decision Tree (กรณี regression) โดย Random Forest ใช้ได้ทั้งกับปัญหา Classification และ Regression

2.3.4 โมเดล Light Gradient Boosting Machine (LGBM) เป็นการสร้างโมเดลที่มีประสิทธิภาพดีขึ้น แนวคิดหลักที่อยู่เบื้องหลังการเร่งความเร็ว คือการฝึกแต่ละโมเดลที่ตามมาเพื่อแก้ไขข้อผิดพลาดที่เกิดจากโมเดลก่อนหน้า ซึ่งจะเป็นการปรับปรุงความแม่นยำในการทำนายโดยรวม โมเดลทางคณิตศาสตร์ที่มีโครงสร้างเป็นแบบต้นไม้หลาย ๆ ต้น (Trees) โดยต้นไม้เหล่านี้จะถูกสร้างขึ้นจากข้อมูลที่ใช้สอนโมเดล โดยโมเดลจะใช้ข้อมูลที่ทางทีมส่งให้โมเดลเรียนรู้ในการสร้างต้นไม้แต่ละครั้ง ดังรูปที่ 2



รูปที่ 2. Light Gradient Boosting Machine (LGBM)

LGBM ได้รับความนิยมเนื่องจากมีความสามารถในการจัดการชุดข้อมูลขนาดใหญ่ได้อย่างมีประสิทธิภาพและให้การคาดการณ์ที่แม่นยำ โดยทั่วไปจะใช้สำหรับงานต่างๆ เช่น การถดถอย การจำแนกประเภท และการจัดอันดับ นอกจากนี้ยังรองรับคุณสมบัติขั้นสูงต่างๆ เช่น การหยุดก่อนเวลา การฝึกอบรมแบบคู่ขนาน และฟังก์ชันการสูญเสียที่กำหนดเอง ซึ่งทำให้เป็นเครื่องมืออเนกประสงค์และทรงพลังในด้านการเรียนรู้ของเครื่อง

2.3.5 ARIMA Model เป็นแบบจำลองการคาดการณ์อนุกรมเวลาที่นิยมใช้ในการวิเคราะห์และทำนายมูลค่าในอนาคตตามรูปแบบในอดีตในข้อมูล รวมสามองค์ประกอบ ได้แก่ การถดถอยอัตโนมัติ (AR) ส่วนต่าง (I) และค่าเฉลี่ยเคลื่อนที่ (MA) โดย ARIMA ย่อมาจาก Auto Regressive Integrated Moving Average หลักการณ์ทำงานจะประกอบไปด้วยอนุกรมเวลา โดยหลักการคือแบบจำลองนี้จะทำการขจัด Noise ออกจากข้อมูลเพื่อลด Error ให้มากที่สุด เพื่อให้ข้อมูลนั้นน่าเชื่อถือส่งผลให้

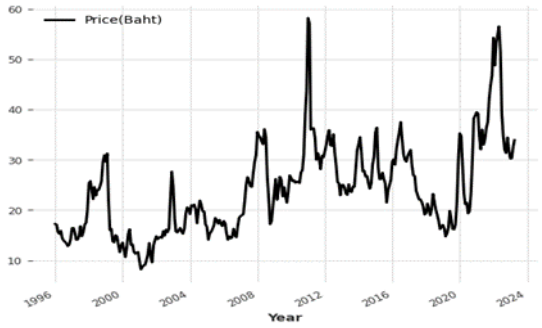
การทำนายราคาในขั้นตอนต่อไปและมีประสิทธิภาพเพิ่มขึ้น โดยอธิบายการทำงานอย่างง่ายดังนี้ โดยเริ่มจากทำการตรวจสอบข้อมูลที่เป็น Time series ว่ามี Noise ในชุดข้อมูลมากน้อยเพียงใดและสามารถนำ Noise ออกได้ข้อมูลได้หรือไม่ ในส่วนต่อมาคือดูว่าข้อมูลเป็น Stationary และลบแนวโน้มในข้อมูล ปรับข้อมูลจนแน่ใจว่าข้อมูลไม่มี Trend หรือ Seasonal และในขั้นตอนสุดท้ายคือเริ่มใช้ค่าหรือข้อมูลตัวเองในการทำนายข้อมูลในอนาคตโดยมีพารามิเตอร์ที่เป็นจำนวนเวลาหรือว่า Lag ในอนุกรมเวลา เพื่อเป็นตัวกำหนดว่าข้อมูลในอดีตมีความสัมพันธ์กับตัวข้อมูลในอนาคต

3. วิธีการดำเนินการวิจัย

ขั้นตอนการดำเนินงานในการพยากรณ์ราคาน้ำมันปาล์มดิบเกรด A ในประเทศไทย จะใช้วิธีการทำนายราคาโดยการวัดผลจากโมเดลพื้นฐานและวัดผลโมเดลที่เป็นแบบแบบจำลองการเรียนรู้เชิงลึก (Deep Learning) ของ LSTM, Exponential Smoothing, Linear Regression, Random Forest, LightGBM Model และ ARIMA โดยใช้ชุดข้อมูลเดียวกันในการทดสอบและนำผลค่าคลาดเคลื่อนมาเปรียบเทียบกันเพื่อหาโมเดลที่มีประสิทธิภาพที่สุด โดยในงานวิจัยนี้มีขอบเขตและข้อจำกัดในการดำเนินการวิจัยคือใช้ปัจจัยคืออนุกรมเวลาเป็นปัจจัยเดียวในการทดสอบ เนื่องจากหากนำปัจจัยอื่นมาเข้าร่วมทดสอบด้วยจะใช้ระยะเวลานานในการรวบรวมข้อมูลและพัฒนาโมเดล รวมถึงการทดสอบในแต่ละโมเดลและปัจจัยที่เพิ่มขึ้น

3.1 การเตรียมข้อมูล (Data Preparation)

การวิจัยนี้ได้นำข้อมูลราคาน้ำมันปาล์มดิบในประเทศไทยจากเว็บไซต์กรมการค้าภายในกระทรวงพาณิชย์ เป็นข้อมูลราคาปาล์มน้ำมันดิบเกรด A เฉลี่ยรายเดือน โดยเป็นข้อมูลตั้งแต่เดือนมกราคม พ.ศ.2539 – เดือนเมษายน พ.ศ. 2566 มาชุดข้อมูลในการเรียนรู้ข้อมูลตัวอย่าง ดังรูปที่ 3



รูปที่ 3. กราฟแสดงข้อมูลราคาน้ำมันปาล์มดิบในประเทศไทย

3.2 การกำหนดแบบจำลองที่ใช้ในการวิจัย

กำหนดแบบจำลองพื้นฐาน แบบจำลองMaการเรียนรู้เชิงลึก (Deep Learning) และ แบบจำลองคลาสสิก ในการพยากรณ์ราคาน้ำมันปาล์มดิบ ดังนี้

3.2.1 แบบจำลอง Last month คือ เรียนรู้ข้อมูลโดยใช้ราคาน้ำมันปาล์มดิบ 1 เดือนก่อนหน้าเพื่อทำนายราคาน้ำมันปาล์มดิบเดือนถัดไป

3.2.2 แบบจำลอง Last Month and Percent Change คือ เรียนรู้ข้อมูลโดยใช้ราคาน้ำมันปาล์มดิบเดือนก่อนหน้าและปรับด้วยเปอร์เซ็นการเปลี่ยนแปลงราคาของ 1 เดือนก่อนหน้าเพื่อทำนายราคาน้ำมันปาล์มดิบเดือนถัดไป

3.2.3 แบบจำลอง Last Month and 3 Months Moving Average Percent Change คือเรียนรู้ข้อมูลโดยใช้ราคาน้ำมันปาล์มดิบ 1 เดือนก่อนหน้าและปรับด้วยเปอร์เซ็นค่าเฉลี่ยเคลื่อนที่ของราคาน้ำมันปาล์มดิบ 3 เดือนก่อนหน้าเพื่อทำนายราคาน้ำมันปาล์มดิบเดือนถัดไป

3.2.4 แบบจำลอง Last Month and 5 Months Moving Average Percent Change คือเรียนรู้ข้อมูลโดยใช้ราคาน้ำมันปาล์มดิบ 1 เดือนก่อนหน้าและปรับด้วยเปอร์เซ็นค่าเฉลี่ยเคลื่อนที่ของราคาน้ำมันปาล์มดิบ 5 เดือนก่อนหน้าเพื่อทำนายราคาน้ำมันปาล์มดิบเดือนถัดไป

3.2.5 แบบจำลอง 3 Months Moving Average คือเรียนรู้ข้อมูลโดยใช้ค่าเฉลี่ยเคลื่อนที่ของราคาน้ำมันปาล์มดิบ 3 เดือน ก่อนหน้าเพื่อทำนายราคาน้ำมันปาล์มดิบเดือนถัดไป

3.2.6 แบบจำลอง 3 Months Exponential Weighted Moving Average คือเรียนรู้ข้อมูลโดยใช้ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักของราคาน้ำมันปาล์มดิบ 3 เดือนก่อนหน้าเพื่อทำนายราคาน้ำมันปาล์มดิบเดือนถัดไป

3.2.7 แบบจำลอง RNN-LSTM คือแบบจำลองการเรียนรู้เชิงลึกโดยมีการกำหนด LSTM Architecture เริ่มต้นที่ 64, 128, 192, 256 และ 320 Nodes และมีจำนวน Layer ที่ 2 layer , 3 layer, 4 layer และ 5 Layer ตามลำดับ รวมแบบจำลองทั้ง 20 Architecture เพื่อพยากรณ์ราคาน้ำมันปาล์มดิบ

3.2.8 แบบจำลองคลาสสิก ได้แก่ โมเดล Exponential Smoothing, Regression Model, Random Forest, LightGBM Model และ ARIMA โดยจะใช้ฟังก์ชั่นของ Darts เพื่อสร้างโมเดลหลังจากนั้นจะทำการทดสอบโดย Back-Testing สรุปแบบจำลองทั้งหมดที่ใช้ในการทดสอบราคาน้ำมันปาล์มดิบของประเทศไทย ดังตารางที่ 1

ตารางที่ 1. แสดงโมเดลพื้นฐานและ RNN-LSTM *ที่ใช้ในการทำนายราคาน้ำมันปาล์มดิบ*

Baseline	• Last month • last month and % change • Last Month and 3 months Moving Average % change • Last month and 5 months Moving Average %Change • 3 months Moving Average • 3 months Exponential Weighted Moving Average	
Deep Learning	RNN-LSTM	จำนวน 2, 3, 4 และ 5 Layer ที่ 64 Node จำนวน 2, 3, 4 และ 5 Layer ที่ 128 Node จำนวน 2, 3, 4 และ 5 Layer ที่ 192 Node จำนวน 2, 3, 4 และ 5 Layer ที่ 256 Node จำนวน 2, 3, 4 และ 5 Layer ที่ 320 Node
Classic Model	• Exponential Smoothing • Linear Regression • Random Forest • LightGBM • ARIMA	Back-Testing และปรับพารามิเตอร์ Lag ตั้งแต่ 12-24 เช่น Lag=1 ใช้ข้อมูลย้อนหลัง 1 Time step ทำนายเพื่อ 1 Time Step ในอนาคต

3.3 กำหนดเมตริกสำหรับวัดประสิทธิภาพของแบบจำลอง

เมตริก (Metric) หรือ Loss Function มีเป้าหมายเพื่อหาค่า Loss หรือค่า Error หลังจากมีการเทรนโมเดล โดยเมตริกที่ใช้ในงานวิจัยที่ใช้วัดประสิทธิภาพของแบบจำลองมีดังต่อไปนี้ ดังสมการที่ (3) - (7)

- Root Mean Square Error (RMSE)

$$RMSE = sqrt\left(\left(\frac{1}{n}\right) * \Sigma(yi - \hat{y}i)^2\right) \quad (3)$$

- R squared score

$$R^2 \quad (4)$$

- Mean Squared Error (MSE)

$$MSE = \left(\frac{1}{n}\right) * \Sigma(yi - \hat{y}i)^2 \quad (5)$$

- Mean Absolute Error (MAE)

$$MAE = \left(\frac{1}{n}\right) * \Sigma|yi - \hat{y}i| \quad (6)$$

- Mean Absolute Percentage Error (MAPE)

$$MAPE = \left(\frac{1}{n}\right) * \Sigma(|(yi - \hat{y}i) / yi| * 100 \quad (7)$$

โดยนำค่า MAPE มาเปรียบเทียบกันเป็นสำคัญ ซึ่งปรกติค่า RMSE, MAE, MSE จะแปรผันตามค่า MAPE และยังมีค่า R^2 ที่ช่วยในการตัดสินใจว่าโมเดลใดค่า R^2 เข้าใกล้ 1 คือโมเดลที่ดีที่สุด แต่ R^2 ก็สามารถมีค่าติดลบ ในกรณีที่การทำนายราคาได้ค่าที่แตกจากข้อมูลจริงอย่างมากซึ่งหมายความว่าแบบจำลองนั้นไม่ดีเลยหรือชุดข้อมูลนั้นไม่เหมาะสมกับแบบจำลองนั้น เหตุผลในการใช้ค่า MAPE เพื่อการพิจารณาเนื่องด้วยค่าความผิดพลาดที่ได้จะเป็นเปอร์เซ็นต์ซึ่งสามารถนำข้อมูลมาเปรียบเทียบและอธิบายได้ง่ายกว่าค่าอื่นๆ

3.4 การทำการเทรนโมเดลและการปรับแต่งโมเดล

การปรับแต่งโมเดลเพื่อเตรียมพร้อมสำหรับการนำข้อมูลเข้ามาเรียนรู้ โดยในการเรียนรู้นั้น ก็จะมีขั้นตอนโดยละเอียดมากมาย ยกตัวอย่างเช่น เพิ่มจำนวน Layer เพื่อหาจุดที่เหมาะสม หาค่า Learning ของโมเดลเพื่อให้โมเดลสามารถเรียนรู้ได้ดี การกำหนด Split Data ที่เหมาะสม เป็นต้น โดยในการทำนายราคาน้ำมันปาล์มดิบได้การแบ่งข้อมูลโดยแบ่ง Train Set และ Test Set โดยกำหนด Test Set มีขนาดเท่ากับ 12 เนื่องจากต้องการทำนายราคาน้ำมันปาล์มดิบเป็นระยะเวลา 1 ปี และเพื่อป้องกันการ Overfitting ของโมเดล Machine Learning ได้มีการทำ Cross Validation สำหรับ Train Set และ Test set โดยข้อมูลจะถูกแบ่งออกเป็นชุดย่อยๆ ทำงานตามจำนวนรอบและสับเปลี่ยนกันทำให้มีให้มีชุดข้อมูลมีการใช้งานทุกส่วน ในการทดสอบนี้ด้วยโดยกำหนด Validation Split เท่ากับ 0.1 เมื่อปรับแต่ง

โมเดลที่เหมาะสมแล้วหลังจากนั้นก็จะเข้าขั้นตอนทดสอบการพยากรณ์ราคาและประเมินประสิทธิภาพต่อไปจากข้อมูลราคาปาล์มมีทั้งที่เป็นข้อมูลราคาผลปาล์ม น้ำมันปาล์ม ปาล์มทั้งทะลาย แต่ผู้วิจัยนำข้อมูลราคาน้ำมันปาล์มดิบในประเทศไทยมาเป็นชุดตัวอย่างโดย Split Data โดยแบ่ง Train Set และ Test Set โดยกำหนด Test Set มีขนาดเท่ากับ 12 เนื่องจากต้องการทำนายราคาน้ำมันปาล์มดิบเป็นระยะเวลา 1 ปี และเพื่อป้องกันการ Overfitting ของโมเดล Machine Learning ได้มีการทำ Cross Validation สำหรับ Train Set และ Test set โดยข้อมูลจะถูกแบ่งออกเป็นชุดย่อยๆ ทำงานตามจำนวนรอบและสับเปลี่ยนกันทำให้มีให้มีชุดข้อมูลมีการใช้งานทุกส่วน ในการทดสอบนี้ด้วยโดยกำหนด Validation Split เท่ากับ 0.1 จากนั้นทำการทดสอบการพยากรณ์ราคาและทำการเก็บรวบรวมข้อมูลเพื่อทำการวิเคราะห์ข้อมูลต่อไป

3.5 ประเมินประสิทธิภาพของแบบจำลองและวิเคราะห์ผลการทดลอง

ทำการวิเคราะห์ข้อมูลโดยนำ ค่า Error ของ RMSE, R^2, MSE, MAE และ MAPE มาพิจารณาค่า โดยค่า MAPE แบบจำลองใดที่มีค่า MAPE น้อยที่สุดให้ถือว่าเป็นแบบจำลองที่ดีที่สุด จากนั้นทำการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งหมดในการพยากรณ์ราคาน้ำมันปาล์มดิบ

4. ผลการทดลอง

ผลการทดสอบจะได้ผลการหาค่า Error ของทุกโมเดลพื้นฐานของการพยากรณ์ราคาน้ำมันปาล์มดิบของไทบ โดยหาค่าของ RMSE, MSE, MAE, MAPE และ R^2 จากการทำนายราคาจากแบบจำลองต่างๆดังต่อไปนี้

4.1 ผลการทดสอบโมเดลพื้นฐานในการพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทย

จากการทดสอบการพยากรณ์ราคาน้ำมันปาล์มดิบที่มีแบบจำลองอยู่ 6 แบบ เมื่อทดสอบประสิทธิภาพได้ค่าความคาดเคลื่อน ดังตารางที่ 2

ตารางที่ 2. แสดงค่า Error และ R2 ของทุกแบบจำลองพื้นฐานในการทำนายราคาน้ำมันปาล์มดิบ

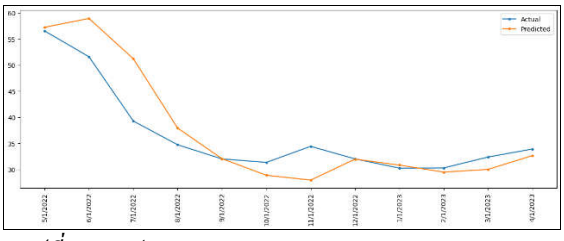
Model Description	RMSE	MSE	MAE	MAPE	R^2
Last Month	4.40	19.37	3.15	8.42	0.71
Last Month and Percent Change	3.97	15.78	3.10	8.46	0.77
Last Month and 3 Months Moving Average Percent Change	4.79	22.91	3.55	9.52	0.66
Last Month and 5 Months Moving Average Percent Change	4.68	21.91	3.10	8.13	0.68
3 Month Moving Average (MA)	7.06	49.86	5.15	14.29	0.26
3 Months Exponential Weighted Moving Average (EWM)	6.22	38.74	4.66	13.00	0.43

ผลการทดสอบประสิทธิภาพแบบจำลองพื้นฐานของการทำนายราคาน้ำมันปาล์มดิบในประเทศไทย ได้ผลลัพท์จากการเปรียบเทียบค่า MAPE ของโมเดลพื้นฐานทั้ง 6 แบบ โมเดลที่ให้ค่า MAPE ที่ดีที่สุดคือโมเดล Last Month and 5 Months Moving Average Percent Change โดยมีค่าความผิดพลาดอยู่ที่ 8.13 ดังตารางที่ 3

ตารางที่ 3. เปรียบเทียบค่า MAPE ของทุกแบบจำลองพื้นฐานในการทำนายราคาน้ำมันปาล์มดิบ

Model Description	MAPE
Last month	8.42
Last Month and Percent Change	8.46
Last Month and 3 Months Moving Average Percent Change	9.52
Last Month and 5 Months Moving Average Percent Change	**8.13**
3 Months Moving Average	14.29
3 Months Exponential Weighted Moving Average	13.00

โดยแบบจำลอง Last Month and 5 Months Moving Average Percent Change ที่มีค่า Error น้อยที่สุด แสดงการทำนายราคาน้ำมันปาล์มดิบในเดือนถัดไป ดังรูปที่4



รูปที่ 4. กราฟ Last Month and 5 Months Moving Average Percent Change

4.2 ผลการทดสอบโมเดล RNN-LSTM ในการพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทย

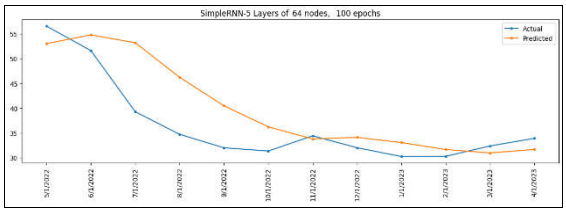
ในการทดลองการพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทย ผู้วิจัยได้ทำการใช้โมเดล RNN-LSTM ที่มี LSTM Architecture จำนวนเท่ากับ 2 layer, 3 layer, 4 layer และ 5 Layer โดยแต่ละ Layer มีจำนวน Node เท่ากับ 64, 128, 192, 256 และ 320 Node ตามลำดับ รวมจำนวนแบบจำลองทั้งสิ้นคือ 20 แบบ และผลลัพท์จากการเปรียบเทียบค่า Error ของ RNN-LSTM ทั้ง 20 แบบจำลอง มีค่าของ RMSE, R^2, MSE, MAE และ MAPE จากทุก Architecture แสดงดังตารางที่ 4

ตารางที่ 4. แสดงค่า Error ของ LSTM ในการทำนายราคาน้ำมันปาล์มดิบ

No.	Model Description	RMSE	MSE	MAE	MAPE	R^2
1	RNN-LSTM 2 Layer 64 Node	11.94	142.48	11.60	33.95	-1.11
2	RNN-LSTM 2 Layer 128 Node	34.04	1158.81	33.03	89.96	-16.14
3	RNN-LSTM 2 Layer 192 Node	30.98	959.53	29.87	80.98	-13.19
4	RNN-LSTM 2 Layer 256 Node	28.51	812.91	27.30	73.67	-11.02
5	RNN-LSTM 2 Layer 320 Node	11.01	121.29	7.33	16.87	-0.79
6	RNN-LSTM 3 Layer 64 Node	18.36	337.15	16.42	42.73	-3.99
7	RNN-LSTM 3 Layer 128 Node	31.31	980.27	30.21	81.95	-13.50
8	RNN-LSTM 3 Layer 192 Node	32.72	1070.48	31.67	86.09	-14.83
9	RNN-LSTM 3 Layer 256 Node	16.55	273.88	14.36	36.88	-3.05
10	RNN-LSTM 3 Layer 320 Node	21.44	459.50	19.80	52.33	-5.79
11	RNN-LSTM 4 Layer 64 Node	29.34	861.00	28.17	76.14	-11.73
12	RNN-LSTM 4 Layer 128 Node	31.42	987.31	30.33	82.28	-13.60
13	RNN-LSTM 4 Layer 192 Node	27.60	761.85	26.35	70.96	-10.27
14	RNN-LSTM 4 Layer 256 Node	32.25	1040.01	31.18	84.71	-14.38
15	RNN-LSTM 4 Layer 320 Node	30.05	903.22	28.91	78.24	-12.36
16	RNN-LSTM 5 Layer 64 Node	10.32	106.52	6.25	**13.83**	-0.58
17	RNN-LSTM 5 Layer 128 Node	33.93	1151.48	32.92	89.66	-16.03
18	RNN-LSTM 5 Layer 192 Node	27.83	774.36	26.58	71.64	-10.45
19	RNN-LSTM 5 Layer 256 Node	26.32	692.78	25.00	67.14	-9.24
20	RNN-LSTM 5 Layer 320 Node	28.34	803.18	27.12	73.16	-10.88

ผลการเปรียบเทียบประสิทธิภาพของแบบจำลอง Deep Learning RNN-LSTM ที่ดีที่สุดได้แก่ RNN-LSTM จำนวน 5 layer และจำนวน 64 Node ให้ค่า MAPE ที่ดีที่สุดอยู่ที่ 13.83 แต่อย่างไรแล้วจะเห็นว่าค่า R^2 มีค่าติดลบทุก Architecture ซึ่งปรกติค่า R^2 จะมีค่าอยู่ระหว่าง 0-1 แต่เมื่อไรก็ตามที่มีค่าติดลบนั้นหมายความว่า ค่าทำนายที่ได้จากโมเดลนั้นให้ผลลัพธ์ที่แย่กว่าการใช้ค่าเฉลี่ยหรือหมายความว่าโมเดลนั้นไม่ดี หรืออาจจะไม่เหมาะสมกับชุดข้อมูลนั้นๆ

โดยแบบจำลองนี้แสดงกราฟการทำนายราคาน้ำมันปาล์มดิบในเดือนถัดไป RNN-LSTM ที่ดีที่สุดได้แก่ RNN-LSTM จำนวน 5 layer และ 64 Node ดังรูปที่ 5



รูปที่ 5. กราฟทำนายราคาน้ำมันปาล์มดิบเดือนถัดไปของ LSTM- 5 layer จำนวน 64 Node

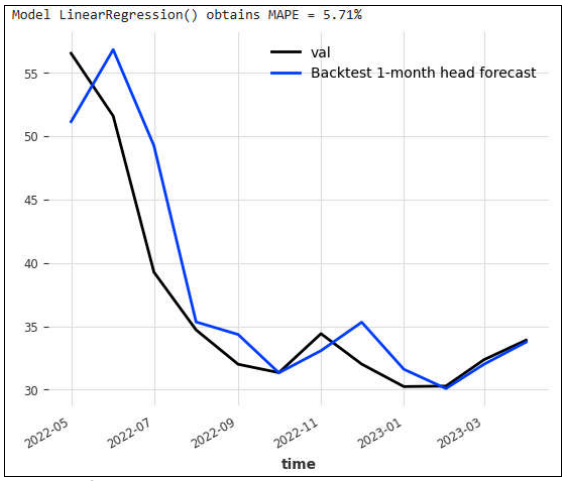
4.3 ผลการทดสอบโมเดลคลาสสิกในการพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทย

ผู้วิจัยทำการเพิ่มแบบจำลองในการทดสอบเพิ่มเติมคือแบบจำลองคลาสสิกมาเพื่อทดสอบการพยากรณ์ราคาน้ำมันปาล์มดิบ โดยแบบจำลองคลาสสิกได้แก่ Linear Regression Model, Exponential Smoothing, Random Forest, LightGBM และ ARIMA มาเพื่อทำการทดสอบ Back-Testing และทดสอบโมเดลคลาสสิก โดยเริ่มต้นการทดสอบโดยการปรับพารามิเตอร์คือค่า Lag โดยเริ่มจากค่าที่ 12 ไปจนถึงค่า 24 และทุกๆค่า Lag หมายถึงการนำข้อมูลย้อนหลัง ตามค่า Lag ในหน่วยเวลามาใช้เพื่อการทำนายข้อมูลในอนาคต เมื่อทำการกำหนดค่า Lag แล้วทำการป้อนข้อมูลให้กับแบบจำลอง หลังจากนั้นทำ Back-Testing กับแบบจำลองคลาสสิกและเก็บผลค่าความผิดพลาด ในทุกๆการปรับค่า Lag โดยผลลัพท์ของการทำ Back-Testing ของ Exponential Smoothing, Linear Regression Model, Random Forest, LightGBM และ ARIMA ได้ผลลัพท์ค่า MAPE ที่มีค่าน้อยที่สุดในแต่ละแบบจำลองและมีค่า Lag ดังตารางที่ 5

ตารางที่ 5 แสดงค่า MAPE ของโมเดลคลาสสิกในการทำนายราคาน้ำมันปาล์มดิบ

Model Description	Lag	MAPE
Exponential Smoothing	15	6.92
Linear Regression	16	**5.71**
Random Forest	16	6.96
LightGBM	13	10.34
ARIMA	12	5.97

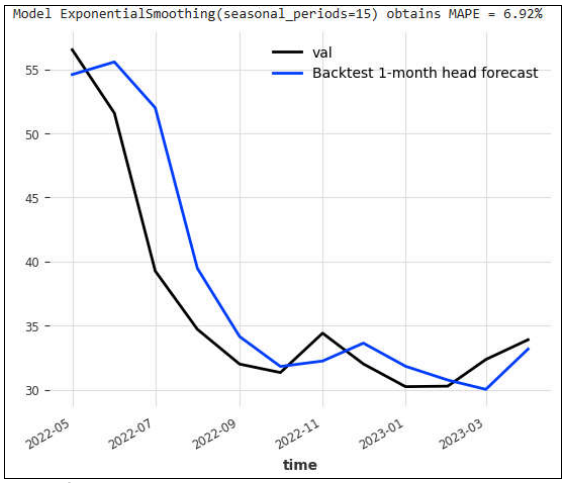
โดยผลการทำ Back-Testing ของโมเดลคลาสสิกแนวโน้มค่า Error เป็นไปในทางเดียวกัน คือเมื่อเริ่มต้นทดสอบที่กำหนดค่า Lag เท่ากับ 12 ไปค่าความคลาดเคลื่อนที่ได้จะเริ่มดีขึ้นเรื่อยๆ และดีที่สุดที่ Lag เท่ากับ 16 โดยค่า Loss ของแบบจำลอง Linear Regression จะดีที่สุดที่ Lag เท่ากับ 16 โดยมีค่า MAPE เท่ากับ 5.71 ดังรูปที่ 6



รูปที่ 6. กราฟ Back-Testing ของ Linear Regression

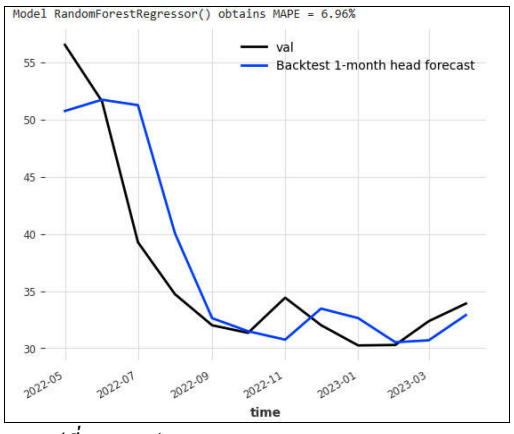
ในส่วนแบบจำลองคลาสสิกอื่นๆผลของ MAPE จะมีแนวโน้มตามค่า Lag ในทางเดียวกันแต่มีค่าไม่ดีเท่า Linear Regression โดยแต่ละแบบจำลองแสดงกราฟและข้อมูล ดังต่อไปนี้

4.3.1 ผลการทำ Back-Testing ที่ดีที่สุดของ Exponential Smoothing จะมีค่า MAPE ดีที่สุดเท่ากับ 6.92 อยู่ที่ Lag =15 ดังภาพที่ 7



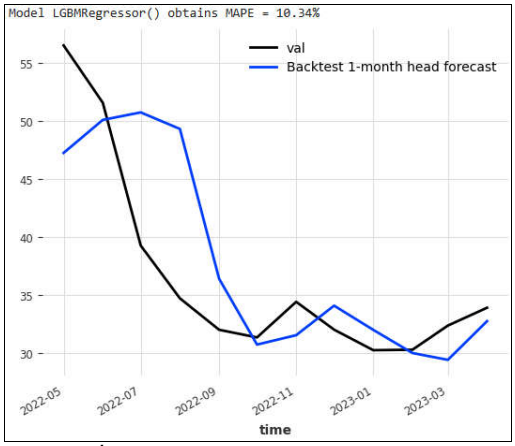
รูปที่ 7. กราฟ Back-Testing ของ Exponential Smoothing

4.3.2 ผลการทำ Back-Testing ที่ดีที่สุดของ Random Forest จะมีค่า MAPE ดีที่สุดเท่ากับ 6.96 อยู่ที่ Lag =16 ดังรูปที่ 8



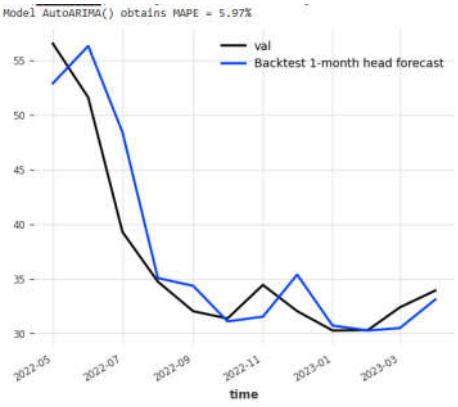
รูปที่ 8. กราฟ Back-Testing ของ Random Forest

4.3.3 ผลการทำ Back-Testing ที่ดีที่สุดของ LightGBM จะมีค่า MAPE ดีที่สุดเท่ากับ 10.34 อยู่ที่ Lag =13 ดังรูปที่ 9



รูปที่ 9. กราฟ Back-Testing ของ LightGBM

4.3.4 ผลการทำ Back-Testing ที่ดีที่สุดของ ARIMA จะมีค่า MAPE ดีที่สุดเท่ากับ 5.97 อยู่ที่ Lag =12 ดังรูปที่ 10



รูปที่ 10. กราฟ Back-Testing ของ ARIMA

4.4 ผลการเปรียบเทียบประสิทธิภาพทุกแบบจำลองในการพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทย

จากการนำค่า MAPE ของทุกโมเดลพื้นฐาน 6 โมเดล, โมเดล Deep Learning RNN-LSTM ทั้งหมด 20 โมเดล และโมเดลคลาสสิก 5 โมเดล มาเปรียบเทียบกันนั้นเพื่อหาแบบจำลองที่ทำการพยากรณ์ราคาเชิงปริมาณแบบอนุกรมเวลาและตรวจสอบค่าความคลาดเคลื่อน(ค่า Loss) เพื่อหาค่าพยากรณ์ที่ดีที่สุดสำหรับการพยากรณ์น้ำมันปาล์มดิบในเดือนถัดไปนั้น แสดงการเปรียบเทียบค่า MAPE ของทุกแบบจำลองได้ ดังตารางที่ 6

ตารางที่ 6 เปรียบเทียบค่า MAPE ของโมเดลแต่ละกลุ่มในการทำนายราคาน้ำมันปาล์มดิบ

Model	MAPE
Last Month and 5 Months Moving Average Percent Change	8.13
RNN- LSTM 5-layer 64Node	13.83
Exponential Smoothing	6.92
Linear Regression	**5.71**
Random Forest	6.96
LightGBM	10.34
ARIMA	5.97

การเปรียบเทียบค่า MAPE ในการพยากรณ์ราคาในอนาคตของน้ำมันปาล์มดิบในประเทศไทย ของแบบจำลองพบว่า Linear Regression มีค่า MAPE น้อยที่สุดอยู่ที่ 5.71

5. บทสรุปและการอภิปราย

บทความนี้แสดงผลการพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทย โดยการศึกษาและเปรียบเทียบประสิทธิภาพของแบบจำลองต่าง ๆ ในการพยากรณ์ราคา โดยพิจารณาจากค่า MAPE หรือ ค่าความคลาดเคลื่อนที่มีค่าที่ดีที่สุดหรือค่าน้อย โดยจากการศึกษานี้สามารถสรุปผลการวิจัยอภิปรายและมีดังต่อไปนี้ ในการทดลองแรกของการพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทย โดยแบบจำลองพื้นฐาน กำหนด Architecture ของ RNN-LSTM และนำแบบจำลองคลาสสิกมาทดสอบเพิ่มเติม ได้ผลลัพท์ว่าแบบจำลองคลาสสิก Linear Regression สามารถพยากรณ์ราคาน้ำมันปาล์มดิบในประเทศไทยได้มีประสิทธิภาพดีที่สุดในงานวิจัยนี้และในการทดลองนี้ยังมีแบบจำลอง ARIMA ที่ให้ผลลัพท์เป็นที่น่าพอใจแม้ว่าใช้ข้อมูลในอดีตไม่มากนักก็สามารถพยากรณ์ราคาออกมาได้ผลลัพท์ที่ดี ซึ่งมีวรรณกรรมมากมายที่ Traditional Model อย่างเช่น Linear Regression หรือ ARIMA ให้ผลความแม่นยำในการทำนายราคาได้ดีกว่า LSTM แม้ว่าในกรณีที่ LSTM จะมีการใช้เวลาการเรียนรู้ชุดข้อมูลที่มากกว่าและมีแบบจำลองที่ซับซ้อนมากกว่า เช่นมีหลาย Layer แต่ ARIMA ก็ สามารถพยากรณ์ราคาได้แม่นยำกว่าและใช้เวลาในการเทรนข้อมูลน้อยกว่าอีกด้วย จึงสรุปได้กว่าโมเดลที่ซับซ้อนกว่าอาจจะไม่ได้ให้ผลลัพท์ที่ดีกว่า ข้อเสนอแนะเพิ่มเติมสำหรับงานวิจัยนี้ในการทำนายราคาโดยการใช้ Machine learning ควรปรับแต่งจากโมเดลที่ไม่ซับซ้อนมากก่อน (Simple Model) ไปจนถึงโมเดลที่มีความซับซ้อนมากขึ้นตามลำดับ(Complex Model) แต่โมเดลที่มีความซับซ้อนก็ไม่ได้หมายความว่าจะได้ผลลัพท์ที่ดีหรือมีความแม่นยำเสมอไป และการทดลองสามารถปรับปรุงแบบจำลองให้ดีขึ้นได้โดยวิธีการปรับแต่ง(Tuning) โมเดล โดยเปลี่ยนตัวแปรด้าน เวลา การปรับระบบโครงข่ายประสาท (Neuron Network) โดยเปลี่ยน Architecture หรือระบุจุดอ่อนในแบบจำลองโดยปรับเปลี่ยนวิธีการตามผลการทดสอบย้อนหลัง(Back-Testing) เพื่อปรับปรุงประสิทธิภาพการพยากรณ์ราคาในอนาคตและการวิจัยนี้เป็นการทดลองโดยใช้ข้อมูลราคาตามอนุกรมเวลาปัจจัยเดียว เพื่อให้งานวิจัยดีขึ้นสามารถเพิ่มปัจจัยอื่นๆอีกได้เช่น ปัจจัยที่เกี่ยวกับ อุปสงค์ อุปทาน ตลาดการเงิน GDP ราคาหุ้น โรคระบาด COVID-19 สงคราม เป็นต้น เพื่อการทำนายราคาน้ำมันปาล์มดิบในอนาคตให้ดีขึ้นกว่าเดิมได้ นอกจากนี้ยังสามารถนำแบบจำลองไปใช้ในการทำนายราคาสินค้าอื่นๆได้อีกด้วย

เอกสารอ้างอิง

[1] Kanchymalay, Kasturi, et al. " Multivariate time series forecasting of crude palm oil price using machine learning techniques." IOP Conference Series: Materials Science and Engineering. Vol. 226. No. 1. IOP Publishing, 2017.

[2] Rahim, Nur Fazliana, Mahmod Othman, and Rajalingam Sokkalingam. "A comparative review on various method of forecasting crude palm oil prices." Journal of Physics: Conference Series. Vol. 1123. No. 1. IOP Publishing, 2018.

[3] Bristone, Makumbonori, Rajesh Prasad, and Adamu Ali Abubakar. "CPPCNDL: Crude oil price prediction using complex network and deep learning algorithms." Petroleum 6.4 (2020): 353-361.

[4] Gupta, Varun, and Ankit Pandey. "Crude oil price prediction using LSTM networks." International Journal of Computer and Information Engineering 12.3 (2018): 226-230.

[5] Zhao, Yang, Jianping Li, and Lean Yu. "A deep learning ensemble approach for crude oil price forecasting." Energy Economics 66 (2017): 9-16.

[6] Li, Zhanke, et al. "Oil Price Forecasting Based on Variational Mode Decomposition, Relative Entropy and LSTM Neural Network." IOP Conference Series: Materials Science and Engineering. Vol. 750. No. 1. IOP Publishing, 2020.

[7] Güleryüz, Didem, and Erdemalp Özden. "The prediction of Brent crude oil trend using LSTM and Facebook Prophet." Avrupa Bilim ve Teknoloji Dergisi 20 (2020): 1-9.

[8] Xie, Wen, et al. "A new method for crude oil price forecasting based on support vector machines." Computational Science–ICCS 2006: 6th International Conference, Reading, UK, May 28-31, 2006, Proceedings, Part IV 6. Springer Berlin Heidelberg, 2006.

[9] Khalid, Norlin, et al. "Crude palm oil price forecasting in Malaysia: An econometric approach." Jurnal Ekonomi Malaysia 52.3 (2018): 263-278.

[10] Siew, Han Lock, and Md Jan Nordin. "Regression techniques for the prediction of stock price trend." 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE). IEEE, 2012.

[11] Kim, Jeonghyeon, et al. "A comparative study of machine learning and spatial interpolation methods for predicting house prices." Sustainability 14.15 (2022): 9056.

[12] Guo, Yuankai, Yangyang Li, and Yuan Xu. "Study on the application of LSTM-LightGBM Model in stock rise and fall prediction." MATEC Web of Conferences. Vol. 336. EDP Sciences, 2021.

[13] Hossain, Mohammad Raquibul, and Mohd Tahir Ismail. "Empirical mode decomposition based on theta method for forecasting daily stock price." Journal of Information and Communication Technology 19. 4 (2020): 533-558.

[14] Herzen, Julien, et al. "Darts: User-friendly modern machine learning for time series." The Journal of Machine Learning Research 23.1 (2022): 5442-5447.

[15] Bhattacharjee, Indronil, and Pryonti Bhattacharja. "Stock price prediction: a comparative study between traditional statistical approach and machine learning approach." 2019 4th international conference on electrical information and communication technology (EICT). IEEE, 2019.

[16] Mbah, Tawum Juvert, et al. "Using LSTM and ARIMA to simulate and predict limestone Price variations." Mining, Metallurgy & Exploration 38 (2021): 913-926

[17] Ebenesh, C., and K. Anitha. "A Novel Approach to Minimize the Mean Square Error in Predicting Stock Price Index using Linear Regression in Comparison with LSTM Model." 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). IEEE, 2022.