การพยากรณ์ราคาทองแดงด้วยการเรียนรู้ของเครื่อง

COPPER PRICE FORECASTING USING MACHINE LEARNING

วรัญชัย ไชยปัญญา
WARUNCHAI CHAIPANYA

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2568
KMITL-2025-SC-M-017-042

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPPER PRICE FORECASTING USING MACHINE LEARNING

WARUNCHAI CHAIPANYA

AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS

KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2025

KMITL-2025-SC-M-017-042

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2025
SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ การพยากรณ์ราคาทองแดงด้วยการเรียนรู้ของเครื่อง
ชื่อนักศึกษา วรัญชัย ไชยปัญญา
รหัสประจำตัว 66056073
ปริญญา วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
พ.ศ. 2568
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ รองศาสตราจารย์ ดร.ณชญาดา กมลมิธิชม

บทคัดย่อ

ในยุคที่โลกกำลังเปลี่ยนผ่านสู่การใช้พลังงานสะอาด ความต้องการใช้ทองแดงในอุตสาหกรรมเพิ่มสูงขึ้น โดยเฉพาะอย่างยิ่งในอุตสาหกรรมผลิตสายไฟฟ้า ซึ่งถือเป็นโครงสร้างพื้นฐานสำคัญของระบบไฟฟ้า ทองแดงมีคุณสมบัติเด่นด้านการนำไฟฟ้าและความทนทาน จึงถูกนำมาใช้เป็นวัสดุตัวนำกระแสไฟฟ้าในระบบส่งจ่ายพลังงาน ความผันผวนของราคาทองแดงในตลาดโลกที่เกิดจากปัจจัยเศรษฐกิจ การเมือง และอุปสงค์-อุปทาน จึงส่งผลโดยตรงต่อต้นทุนการผลิตสายไฟและต้นทุนโครงการโครงสร้างพื้นฐานในหลายภาคส่วน

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาและเปรียบเทียบประสิทธิภาพของแบบจำลองทางสถิติและแบบจำลองการเรียนรู้ของเครื่องในการพยากรณ์ราคาทองแดง โดยใช้ข้อมูลอนุกรมเวลาราคาปิดทองแดงรายวันจากตลาด London Metal Exchange (LME) ในช่วงปี พ.ศ. 2553 ถึง พ.ศ. 2567 ร่วมกับตัวแปรต้นอื่น ๆ ที่เกี่ยวข้อง เช่น ราคาทองคำ GDP ของประเทศผู้ผลิตทองแดง และอัตราแลกเปลี่ยน

ในการศึกษาครั้งนี้ ได้พัฒนาแบบจำลองทั้งหมด 5 รูปแบบ ได้แก่แบบจำลอง ARIMA, แบบจำลองต้นไม้ตัดสินใจ, แบบจำลอง XGBoost, แบบจำลองป่าสุ่ม และ แบบจำลองการเรียนรู้เชิงลึก LSTM โดยประเมินผลลัพธ์ด้วยค่าความคลาดเคลื่อน ได้แก่ MAE, RMSE และ MAPE ผลการวิจัยพบว่าแบบจำลอง XGBoost มีประสิทธิภาพสูงสุด รองลงมาคือแบบจำลองป่าสุ่ม, แบบจำลองการเรียนรู้เชิงลึก LSTM, และแบบจำลองต้นไม้ตัดสินใจ ตามลำดับ ขณะที่ ARIMA ให้ผลลัพธ์แม่นยำน้อยที่สุด นอกจากนี้ตัวแปรตามที่นำมาพิจารณาเพิ่มเติมล้วนช่วยเพิ่มความแม่นยำของแบบจำลองในการพยากรณ์ราคาทองแดง ได้อย่างมีนัยสำคัญ

คำสำคัญ : การพยากรณ์ราคา, ARIMA, Decision Tree, LSTM, Random Forest, XGBoost

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Independent Study Title	Copper Price Forecasting Using Machine Learning
Student Name	Warunchai Chaipanya
Student ID	66056073
Degree	Master of Science (Data Science and Analytics)
	KMITL Digital Analytics and Intelligence Center
Year	2025
Independent Study Advisor	Assoc. Prof. Dr. Nachayadar Kamolmitisom

Abstract

As the world transitions toward clean energy, the demand for copper in various industries has significantly increased especially in the electric wire manufacturing sector, which forms a crucial part of electrical infrastructure. Due to its excellent conductivity and durability, copper is widely used as a conductor in power transmission systems. However, copper prices in the global market are highly volatile, influenced by economic, political, and supply-demand factors. Such fluctuations directly affect the production costs of electric wires and infrastructure projects across multiple sectors.

This study aims to develop and compare the performance of statistical and machine learning models for forecasting copper prices. The research utilizes daily closing price time series data of copper from the London Metal Exchange (LME) spanning from 2010 to 2024, along with other relevant exogenous variables such as gold prices, GDP of copper-producing countries, and exchange rates.

Five models were developed in this study: ARIMA model, Decision Tree model, XGBoost model, Random Forest model, and LSTM deep learning model. Model performance was evaluated using error metrics including MAE, RMSE, and MAPE. The research findings reveal that the XGBoost model achieved the highest performance, followed by Random Forest, LSTM deep learning, and Decision Tree models, respectively. The ARIMA model demonstrated the lowest accuracy. Additionally, all supplementary variables considered in the study significantly enhanced the accuracy of copper price forecasting models.

Keywords : ARIMA, Decision Tree, LSTM, Price forecasting, Random Forest, XGBoost

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

การค้นคว้าอิสระเรื่อง “การพยากรณ์ราคาทองแดงด้วยการเรียนรู้ของเครื่อง” ฉบับนี้ สำเร็จลุล่วงได้ด้วยความอนุเคราะห์และการสนับสนุนจากหลายฝ่าย ข้าพเจ้าขอกราบขอบพระคุณทุกท่านที่มีส่วนร่วมและให้ความช่วยเหลือในการดำเนินงานวิจัยครั้งนี้

ข้าพเจ้าขอขอบพระคุณ รศ.ดร.ณชญาดา กมลมิธิชม อาจารย์ที่ปรึกษาการค้นคว้าอิสระ ที่ให้คำแนะนำ ความรู้ และข้อเสนอแนะอันมีคุณค่าอย่างยิ่งตลอดระยะเวลาของการทำวิจัย ทำให้ข้าพเจ้าสามารถดำเนินงานวิจัยได้อย่างมีทิศทางและประสิทธิภาพ

ขอขอบคุณคณาจารย์ทุกท่านในศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ที่ถ่ายทอดองค์ความรู้และมอบโอกาสในการเรียนรู้เชิงลึกด้านวิทยาการข้อมูล ทำให้ข้าพเจ้าสามารถนำความรู้ไปประยุกต์ใช้ในการพัฒนาแบบจำลองทางวิทยาศาสตร์ข้อมูลได้อย่างมั่นใจ

ขอขอบคุณครอบครัวที่ให้กำลังใจและสนับสนุนข้าพเจ้าอย่างเต็มที่มาโดยตลอด ตลอดจนเพื่อน ๆ และผู้มีส่วนเกี่ยวข้องทุกท่านที่มีส่วนช่วยเหลือ ส่งเสริม และให้คำปรึกษาในทุกขั้นตอนของการทำวิจัยฉบับนี้จนสำเร็จลุล่วงด้วยดี

นาย วรัญชัย ไชยปัญญา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ค
สารบัญ	ง
สารบัญตาราง	ฉ
สารบัญรูป	ช
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	1
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	2
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	3
2.1 ทองแดง (Copper)	3
2.1.1 คุณสมบัติ	3
2.1.2 การผลิต	4
2.1.3 การประยุกต์ใช้งาน	4
2.1.4 แนวโน้มในอนาคต	4
2.2 อนุกรมเวลา (Time Series)	5
2.2.1 ความสำคัญของการวิเคราะห์อนุกรมเวลา	5
2.2.2 องค์ประกอบของอนุกรมเวลา	6
2.3 การเรียนรู้ของเครื่อง (Machine Learning)	6
2.4 แบบจำลอง ARIMA (Autoregressive Integrated Moving Average)	7
2.5 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)	8
2.6 แบบจำลอง XGBoost (Extreme Gradient Boosting)	9
2.7 แบบจำลองป่าสุ่ม (Random Forest)	10
2.8 แบบจำลอง LSTM (Long Short-Term Memory)	11
2.9 การตรวจสอบค่าสุดโต่ง (Outlier Detection)	12
2.9.1 การตรวจสอบค่าสุดโต่งด้วยวิธี Interquartile Range (IQR)	12
2.10 วิศวกรรมคุณลักษณะ (Feature Engineering)	13
2.10.1 การสร้างคุณลักษณะ (Feature Creation)	13
2.10.2 การแปลงคุณลักษณะ (Feature Transformation)	13
2.10.3 การสกัดคุณลักษณะ (Feature Extraction)	13
2.10.4 การเลือกคุณลักษณะ (Feature Selection)	14
2.10.5 การปรับสเกลคุณลักษณะ (Feature Scaling)	14
2.11 การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation))	14
2.11.1 การประเมินประสิทธิภาพสำหรับงานด้านจำแนกประเภท	15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.11.2 การประเมินประสิทธิภาพสำหรับงานด้านทำนายค่า 16
2.12 งานวิจัยที่เกี่ยวข้อง 17
บทที่ 3 วิธีการดำเนินงานวิจัย 20
3.1 ขั้นตอนการดำเนินงานวิจัย 21
3.2 เครื่องมือที่ใช้ในการวิจัย 21
3.3 การรวบรวมข้อมูล 21
3.4 การจัดเตรียมข้อมูล 22
3.4.1 การตรวจสอบและทำความสะอาดข้อมูล 22
3.4.2 วิศวกรรมคุณลักษณะ 23
3.5 การพัฒนาแบบจำลอง 27
3.5.1 แบบจำลองทางสถิติ 27
3.5.2 แบบจำลองการเรียนรู้ของเครื่อง 28
3.6 การวัดประสิทธิภาพแบบจำลอง 30
บทที่ 4 ผลการวิจัยและการอภิปรายผล 31
4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น 31
4.2 ผลการพัฒนาแบบจำลอง 32
4.2.1 ผลการพัฒนาแบบจำลอง ARIMA 32
4.2.2 ผลการพัฒนาแบบจำลองต้นไม้ตัดสินใจ 33
4.2.3 ผลการพัฒนาแบบจำลอง XGBoost 35
4.2.4 ผลการพัฒนาแบบจำลองป่าสุ่ม 36
4.2.5 ผลการพัฒนาแบบจำลองการเรียนรู้เชิงลึก LSTM 37
4.3 เปรียบเทียบคุณลักษณะที่มีผลต่อแบบจำลอง 38
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 40
5.1 สรุปผลการวิจัย 40
5.2 ข้อเสนอแนะ 40
เอกสารอ้างอิง 42
ภาคผนวก 44
ภาคผนวก ก 45
ประวัติผู้เขียน 67

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 แสดงการเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับงานวิจัยฉบับนี้	18
3.1 แสดงค่า Variance Inflation Factor กับ คุณลักษณะ	26
3.2 แสดงตัวแปรในการตั้งค่าแบบจำลอง	29
4.1 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง	38
4.1 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง XGBoost เพื่อศึกษาคุณลักษณะ	39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
2.1 แสดงภาพโครงสร้างของสายไฟฟ้าแรงดันสูง	4
2.2 แสดงองค์ประกอบของอนุกรมเวลา	5
2.3 แสดงภาพการฝึกสอนของการเรียนรู้ของเครื่องแบบมีผู้สอนและไม่มีผู้สอน	7
2.4 แสดงแผนภาพการทำงานของแบบจำลองต้นไม้ตัดสินใจ	8
2.5 แสดงแผนภาพการทำงานของแบบจำลอง XGBoost	9
2.6 แสดงแผนภาพการทำงานของแบบจำลองป่าสุ่ม	10
2.7 แสดงแผนภาพการทำงานของแบบจำลอง LSTM	11
2.8 แสดงแผนภาพการตรวจสอบค่าสุดโต่งด้วยวิธี IQR	12
2.9 แสดงแผนภาพการสร้างแบบจำลอง	14
2.10 แสดง ตาราง Confusion Matrix	15
3.1 แสดงขั้นตอนการดำเนินงาน	20
3.2 แสดงตัวอย่างข้อมูลชุดข้อมูลราคาทองแดงที่รวบรวม	22
3.3 แสดงข้อมูลรายปีของค่าผลิตภัณฑ์มวลรวมของประเทศที่คาดว่ามีผลต่อราคาทองแดง	22
3.4 แสดงกราฟการกระจายตัวของตัวแปรปริมาณการซื้อขายของโลหะทองแดง	23
3.5 แสดงแผนภาพความร้อนสหสัมพันธ์ของชุดข้อมูลโลหะทองแดง	23
3.6 แสดงแผนภาพความร้อนสหสัมพันธ์ของชุดข้อมูลโลหะทองแดงหลังจากสร้างตัวแปรความแตกต่างของราคาสูงสุด กับราคาต่ำสุดของโลหะทองแดง	24
3.7 แสดงแผนภาพความร้อนสหสัมพันธ์ของชุดข้อมูลสำหรับการสร้างแบบจำลอง	25
3.8 แสดงการตรวจสอบค่าสุดโต่งในชุดข้อมูล	26
3.9 แสดงกราฟ ACF	27
3.10 แสดงกราฟ PACF	28
4.1 แสดงกราฟราคาปิดโลหะทองแดงรายวัน	31
4.2 แสดงกราฟความถี่จาก Fast Fourier transform	31
4.3 แสดงกราฟข้อมูลหลังจากแยกส่วนประกอบอนุกรมเวลา	32
4.4 แสดง กราฟผลการพยากรณ์ของแบบจำลอง ARIMA เทียบกับข้อมูลจริง	33
4.5 แสดงกราฟผลการพยากรณ์ของแบบจำลอง ARIMA เทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ	33
4.6 แสดง กราฟผลการพยากรณ์ของแบบจำลองต้นไม้ตัดสินใจเทียบกับข้อมูลจริง	34
4.7 แสดงกราฟผลการพยากรณ์ของแบบจำลองต้นไม้ตัดสินใจเทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ	34
4.8 แสดง กราฟผลการพยากรณ์ของแบบจำลอง XGBoost เทียบกับข้อมูลจริง	35
4.9 แสดงกราฟผลการพยากรณ์ของแบบจำลอง XGBoost เทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ	35
4.10 กราฟผลการพยากรณ์ของแบบจำลองป่าสุ่มเทียบกับข้อมูลจริง	36
4.11 แสดงกราฟผลการพยากรณ์ของแบบจำลองป่าสุ่มเทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ	36
4.12 กราฟผลการพยากรณ์ของแบบจำลองการเรียนรู้เชิงลึก LSTM เทียบกับข้อมูลจริง	37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.13 แสดงกราฟผลการพยากรณ์ของแบบจำลองการเรียนรู้เชิงลึก LSTM เทียบกับข้อมูลจริง เฉพาะข้อมูลทดสอบ 37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ทองแดง เป็นหนึ่งในโลหะที่ได้รับความสำคัญในเศรษฐกิจโลก เนื่องจากคุณสมบัติการนำไฟฟ้าและความร้อนที่ยอดเยี่ยม จึงถูกนำไปใช้เป็นวัตถุดิบในอุตสาหกรรมต่างๆ เช่น พลังงาน อิเล็กทรอนิกส์ ยานยนต์ การก่อสร้าง และโครงข่ายโทรคมนาคม เป็นต้น โดยเฉพาะในยุคปัจจุบันที่มุ่งเน้นไปยังการเปลี่ยนผ่านเป็นพลังงานสะอาด ทองแดงจึงกลายเป็นวัตถุดิบสำคัญสำหรับโครงสร้างพื้นฐานด้านพลังงานหมุนเวียน

อุปกรณ์สำหรับการผลิตพลังงานสะอาดนอกจากแหล่งกำเนิดพลังงาน อย่างเช่น แผงโซลาร์เซลล์ กังหันลม และกังหันน้ำ เป็นต้น แล้วยังมีจำเป็นต้องมีอุปกรณ์อื่นด้วย โดยเฉพาะสายไฟฟ้าซึ่งถือเป็นส่วนประกอบที่สำคัญไม่แพ้กัน โดยวัตถุดิบหลักในอุตสาหกรรมสายไฟนั่นก็คือ ทองแดง โดยทองแดงจะใช้ผลิตเป็นตัวนำไฟฟ้า เพื่อคอยส่งผ่านไฟฟ้าจากแหล่งก่อกำเนิดไปส่วนถัดๆ ไปจนถึงผู้บริโภค นั่นจึงเป็นสาเหตุที่ความต้องการโลหะทองแดงมีปริมาณมากขึ้น แต่ด้วยหลายๆ ปัจจัย เช่น อุปสงค์ และอุปทานในตลาดโลก สภาวะเศรษฐกิจมหภาค นโยบายการเงินระหว่างประเทศ ความไม่แน่นอนทางการเมือง และภัยพิบัติทางธรรมชาติ เป็นต้น ความผันผวนของราคาทองแดงส่งผลกระทบโดยตรงต่อต้นทุนการผลิตในอุตสาหกรรม และกระทบต่อห่วงโซ่อุปทานทั่วโลก การพยากรณ์ราคาทองแดงที่แม่นยำจึงเป็นประโยชน์อย่างยิ่งต่อการวางแผนธุรกิจ การบริหารความเสี่ยง และการตัดสินใจลงทุน

อย่างไรก็ตาม การพยากรณ์ราคาทองแดงด้วยวิธีการดั้งเดิมมีข้อจำกัดหลายประการ เช่น การวิเคราะห์เชิงเทคนิคและการวิเคราะห์ปัจจัยพื้นฐานอาจไม่สามารถจับความสัมพันธ์ที่ซับซ้อนระหว่างปัจจัยต่างๆ ได้อย่างครบถ้วน และอาจมีอคติจากความคิดเห็นของผู้เชี่ยวชาญ นอกจากนี้ แบบจำลองทางสถิติแบบดั้งเดิมอาจไม่สามารถรองรับข้อมูลที่มีความซับซ้อนและมีมิติสูงได้อย่างมีประสิทธิภาพ

การใช้เทคนิคการพยากรณ์ราคาทองแดงด้วย การเรียนรู้ของเครื่อง (Machine Learning) เป็นวิธีการใหม่ในการแก้ไขปัญหาการพยากรณ์ราคาทองแดง ด้วยความสามารถในการวิเคราะห์ข้อมูลขนาดใหญ่ การค้นหารูปแบบที่ซับซ้อน และการปรับปรุงประสิทธิภาพของตัวเองอย่างต่อเนื่อง เทคนิคการเรียนรู้ของเครื่องมีศักยภาพในการพัฒนาแบบจำลองที่สามารถจับความสัมพันธ์ที่ซับซ้อนระหว่างปัจจัยต่างๆ ที่ส่งผลต่อราคาทองแดง ทำให้เกิดแบบจำลองที่มีความแม่นยำมากขึ้น ทำให้ลดผลกระทบทางธุรกิจจากการคาดการณ์ราคาทองแดงที่ผิดพลาด จนส่งผลให้เกิดต้นทุนการผลิตที่เกินจริง และการบริหารคลังสินค้าที่ขาดประสิทธิภาพ ส่งผลให้เสียเปรียบในด้านการแข่งขันทางธุรกิจ

1.2 วัตถุประสงค์ของงานวิจัย

1) เพื่อพัฒนาแบบจำลองสำหรับพยากรณ์ราคาทองแดง
2) เพื่อศึกษาและประเมินประสิทธิภาพของแบบจำลอง ด้วยเทคนิคการเรียนรู้ของเครื่อง และแบบจำลองทางสถิติ ในการพยากรณ์ราคาทองแดง
3) เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อการพยากรณ์ราคาทองแดง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ขอบเขตของงานวิจัย

1) ข้อมูลที่ใช้ในงานวิจัย คือ ข้อมูลปริมาณราคาทองแดง โดยอ้างอิงจากตลาด London Metal Exchange (LME) ปี พ.ศ. 2553 ถึง ปี พ.ศ. 2567 ในการพัฒนาแบบจำลอง
2) การศึกษาใช้แบบจำลองทางสถิติ ได้แก่ ARIMA และแบบจำลองการเรียนรู้ของเครื่อง ได้แก่ แบบจำลองต้นไม้ตัดสินใจ แบบจำลอง XGBoost แบบจำลองป่าสุ่ม และ แบบจำลองการเรียนรู้เชิงลึก LSTM สำหรับการพยากรณ์ราคาโลหะทองแดง
3) ประสิทธิภาพของแบบจำลองแต่ละประเภทจะถูกประเมินโดยใช้ค่าความคลาดเคลื่อน ได้แก่ ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE), ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) และค่าความคลาดเคลื่อนสัมพัทธ์เฉลี่ย (Mean Absolute Percentage Error: MAPE)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้รับแบบจำลองที่มีประสิทธิภาพสำหรับการพยากรณ์ราคาทองแดง
2) ได้ทราบถึงปัจจัยสำคัญที่มีอิทธิพลต่อการเปลี่ยนแปลงของราคาทองแดงจากการวิเคราะห์ตัวแปรต้นที่ใช้ในแบบจำลอง
3) ผลลัพธ์จากงานวิจัยสามารถใช้เป็นแนวทางหรือฐานข้อมูลสำหรับการพัฒนาแบบจำลองการพยากรณ์ราคาทองแดงที่มีความซับซ้อนและแม่นยำยิ่งขึ้นในอนาคต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาวิจัยเรื่อง “การพยากรณ์ราคาทองแดงด้วยการเรียนรู้ของเครื่อง” ผู้วิจัยได้รวบรวมแนวคิดและทฤษฎี จากเอกสารและงานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดของเนื้อหาดังนี้

2.1 ทองแดง (Copper)

2.2 อนุกรมเวลา (Time Series)

2.3 การเรียนรู้ของเครื่อง (Machine Learning)

2.4 แบบจำลอง ARIMA (Autoregressive Integrated Moving Average)

2.5 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

2.6 แบบจำลอง XGBoost (Extreme Gradient Boosting)

2.7 แบบจำลองป่าสุ่ม (Random Forest)

2.8 แบบจำลอง LSTM (Long Short-Term Memory)

2.9 การตรวจสอบค่าสุดโต่ง (Outlier Detection)

2.10 วิศวกรรมคุณลักษณะ (Feature Engineering)

2.11 การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation)

2.12 งานวิจัยที่เกี่ยวข้อง

2.1 ทองแดง (Copper)

ทองแดง เป็นธาตุเคมีที่มีสัญลักษณ์ Cu (มาจากภาษาละติน cuprum) และมีเลขอะตอม 29 เป็นโลหะที่มีคุณสมบัติโดดเด่น ได้แก่ ความอ่อน เหนียว ตีเป็นแผ่นได้ดี และมีค่าการนำความร้อนและไฟฟ้าสูงมาก เมื่อสัมผัสกับอากาศบริสุทธิ์ ทองแดงใหม่จะปรากฏเป็นสีส้มอมชมพู ซึ่งเป็นลักษณะเฉพาะตัวที่แยกออกจากโลหะทั่วไปที่มักมีสีเทาหรือเงิน

2.1.1 คุณสมบัติ

ทองแดงเป็นธาตุเพียงไม่กี่ชนิดที่มีสีธรรมชาติเป็นสีส้มแดง โดยเมื่อสัมผัสกับอากาศจะหมองคล้ำลงอย่างช้า ๆ จากการเกิดปฏิกิริยาออกซิเดชัน ซึ่งก่อให้เกิดชั้นบางของออกไซด์ที่ช่วยป้องกันการกัดกร่อนเพิ่มเติม ทองแดงยังมีค่าการนำไฟฟ้าและความร้อนรองจากเงิน ด้วยค่าความต้านทานไฟฟ้าที่ต่ำมาก จึงเหมาะสำหรับการใช้งานในระบบส่งกำลังและอุปกรณ์อิเล็กทรอนิกส์

อีกทั้งทองแดงยังมีความเหนียวสูง สามารถรีดเป็นเส้นลวดโดยไม่ขาด เหมาะอย่างยิ่งกับการใช้งานในสายไฟฟ้าและอุปกรณ์ที่ต้องการความยืดหยุ่น พร้อมทั้งมีความต้านทานต่อการคืบและความชื้น ทำให้เหมาะสมในงานติดตั้งถาวรในสภาพแวดล้อมที่หลากหลาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 การผลิต

ทองแดงส่วนใหญ่ถูกสกัดจากแร่ทองแดงซัลไฟด์ เช่น คาลโคไพไรต์ (chalcopyrite) และบอร์ไนต์ (bornite) ผ่านกระบวนการลอยตัวของแร่ ถลุง และอิเล็กโทรลิซิส เพื่อให้ได้ทองแดงบริสุทธิ์ในระดับ 99.99% ซึ่งเหมาะสำหรับการใช้งานทางไฟฟ้าในระดับสูง โดยประเทศผู้ผลิตสำคัญ ได้แก่ ชิลี เปรู และออสเตรเลียเป็นต้น

2.1.3 การประยุกต์ใช้งาน

ทองแดงเป็นวัสดุที่มีความสำคัญอย่างยิ่งในภาคอุตสาหกรรม โดยเฉพาะในกลุ่มพลังงานและอิเล็กทรอนิกส์ ด้วยคุณสมบัติการนำไฟฟ้าและความร้อนสูง รวมถึงความยืดหยุ่นและความทนทาน ทำให้ทองแดงเป็นส่วนประกอบสำคัญของระบบไฟฟ้าและเทคโนโลยีสมัยใหม่

ในอุตสาหกรรมไฟฟ้า ทองแดงถูกใช้ในสายไฟและสายเคเบิลเป็นสัดส่วนสูงถึง 60% ของการใช้งานทั้งหมด โดยสามารถพบเห็นในระบบไฟฟ้าแรงต่ำจนถึงแรงสูง เช่น สายไฟในอาคาร สถานีไฟฟ้าย่อย และหม้อแปลง ซึ่งล้วนต้องการประสิทธิภาพในการนำกระแสไฟฟ้าอย่างสูงสุด

ในด้านอิเล็กทรอนิกส์ ทองแดงมีบทบาทสำคัญในแผงวงจรพิมพ์และอุปกรณ์ที่ต้องการการระบายความร้อน เช่น ฮีทซิงค์และชิ้นส่วนภายในของคอมพิวเตอร์และโทรศัพท์มือถือ ทองแดงช่วยให้การส่งผ่านสัญญาณและพลังงานเป็นไปอย่างราบรื่น

นอกจากนั้น ยังมีการนำทองแดงมาใช้ในงานสถาปัตยกรรม เช่น หลังคา รางน้ำ และองค์ประกอบตกแต่งอาคาร อีกทั้งยังมีคุณสมบัติต้านจุลชีพ ซึ่งถูกนำไปใช้ในพื้นผิวสัมผัสของโรงพยาบาลหรือพื้นที่สาธารณะ

2.1.4 แนวโน้มในอนาคต

ในอนาคต ทองแดงยังคงเป็นวัสดุหลักในระบบส่งพลังงานและระบบอิเล็กทรอนิกส์ โดยเฉพาะในสภาพแวดล้อมที่มีข้อจำกัดด้านพื้นที่ ซึ่งต้องการวัสดุที่นำไฟฟ้าได้ดี ความบริสุทธิ์ในระดับ 99.99% ยังคงเป็นมาตรฐานของทองแดงที่ใช้ในระบบพลังงานระดับสูงดังรูปที่ 2.1 ที่แสดงถึงโครงสร้างของสายไฟฟ้าแรงสูงที่ใช้ทองแดงเป็นส่วนประกอบทั้งตัวนำไฟฟ้า และชั้นโล่ป้องกัน

รูปที่ 2.1 แสดงภาพโครงสร้างของสายไฟฟ้าแรงดันสูง
(ที่มา: www.pdcable.com/product/hxlp-cws-lat-69-115)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

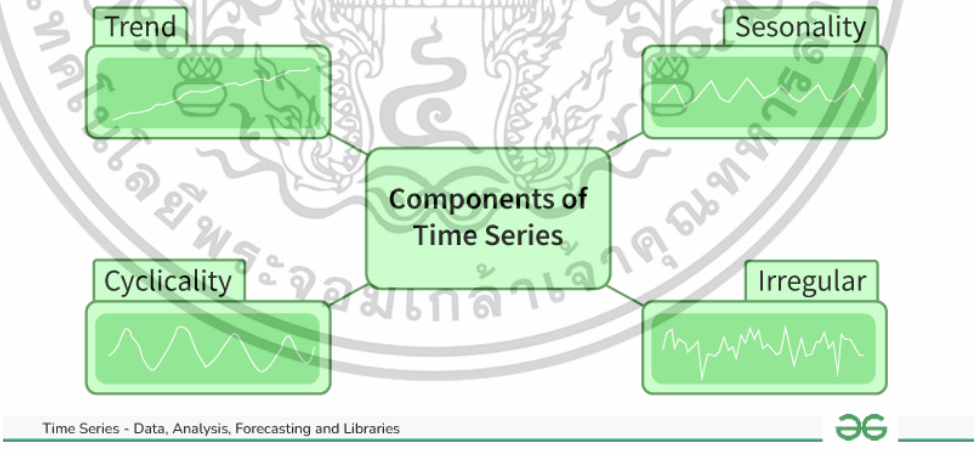
2.2 อนุกรมเวลา (Time Series)

อนุกรมเวลา คือชุดข้อมูลที่ถูกรวบรวมในช่วงเวลาที่ต่อเนื่องกัน โดยมีระยะห่างของเวลาที่สม่ำเสมอ เช่น รายวัน รายเดือน หรือรายปี ข้อมูลแต่ละจุดสะท้อนค่าที่วัดได้ ณ ช่วงเวลาหนึ่ง เช่น ราคาหุ้น อุณหภูมิ หรือยอดขายรายวัน โดยข้อมูลอนุกรมเวลามักแสดงเป็นกราฟเส้นที่มีเวลาอยู่บนแกนนอน (x-axis) และค่าของตัวแปรที่สนใจอยู่บนแกนตั้ง (y-axis)

2.2.1 ความสำคัญของการวิเคราะห์อนุกรมเวลา

การวิเคราะห์อนุกรมเวลา เป็นเครื่องมือสำคัญที่ช่วยให้สามารถเข้าใจพฤติกรรมของข้อมูลในอดีต และใช้ในการพยากรณ์อนาคต จุดเด่นของการวิเคราะห์ประเภทนี้ ได้แก่

1) การคาดการณ์แนวโน้มในอนาคต ช่วยให้ธุรกิจสามารถคาดการณ์ความต้องการของตลาด ราคาสินค้า หรือเหตุการณ์สำคัญในอนาคต
2) การตรวจจับรูปแบบและความผิดปกติ ทำให้สามารถตรวจพบพฤติกรรมที่ซ้ำกัน หรือเหตุการณ์ผิดปกติที่ควรได้รับการพิจารณา
3) การลดความเสี่ยง การรู้ล่วงหน้าช่วยให้สามารถจัดการกับความเสี่ยงได้อย่างมีประสิทธิภาพ
4) การวางแผนกลยุทธ์ สนับสนุนการตัดสินใจด้านการเงิน การขนส่ง การผลิต และทรัพยากร
5) การสร้างความได้เปรียบในการแข่งขัน ข้อมูลที่ถูกต้องแม่นยำช่วยในการปรับตัวตามแนวโน้มของตลาด



รูปที่ 2.2 แสดงองค์ประกอบของอนุกรมเวลา
(ที่มา: www.geeksforgeeks.org/time-series-analysis-and-forecasting)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.2 องค์ประกอบของอนุกรมเวลา

ข้อมูลอนุกรมเวลาประกอบด้วย 4 องค์ประกอบหลัก ดังที่แสงในรูปที่ 2.2 ได้แก่

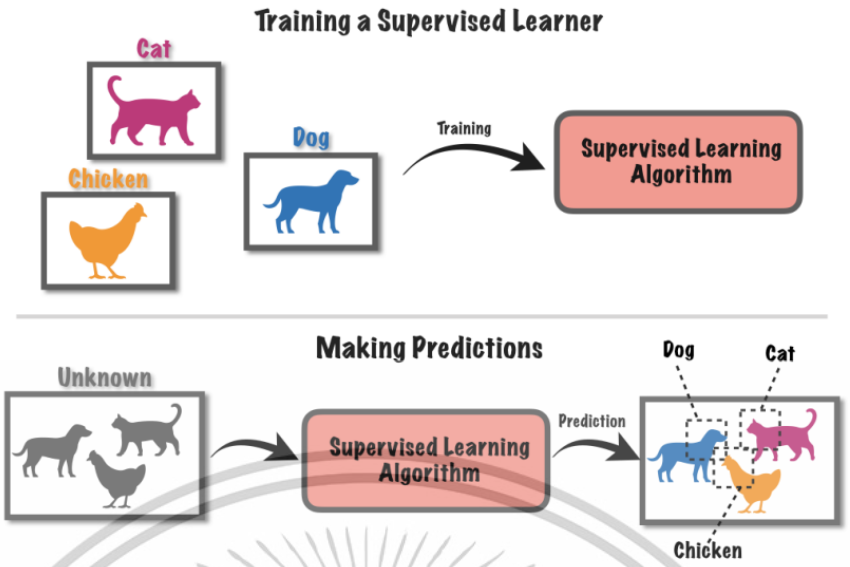
1) แนวโน้ม (Trend) การเปลี่ยนแปลงของข้อมูลในระยะยาว เช่น การเพิ่มขึ้นหรือลดลงของราคาอย่างต่อเนื่อง อาจเป็นแนวโน้มเชิงเส้นหรือแนวโน้มไม่เชิง
2) ฤดูกาล (Seasonality) ความผันผวนหรือรูปแบบที่เกิดซ้ำในช่วงเวลาที่แน่นอน เช่น ฤดูกาลขายดีในปลายปี วันหยุด หรือรอบปีทางธุรกิจ
3) วัฏจักร (Cyclicality) ความผันผวนระยะยาวที่ไม่สม่ำเสมอ เช่น วัฏจักรเศรษฐกิจที่อาจกินเวลาหลายปี และไม่ได้เกิดซ้ำแน่นอนเหมือนฤดูกาล
4) ความแปรปรวนแบบสุ่ม (Irregularity) ความผันผวนที่ไม่สามารถคาดการณ์ได้และไม่ได้เกิดจากแนวโน้ม ฤดูกาล หรือวัฏจักร เช่น ความผิดพลาดในการวัด หรือเหตุการณ์ที่ไม่คาดคิด

2.3 การเรียนรู้ของเครื่อง (Machine Learning)

การเรียนรู้ของเครื่อง คือสาขาหนึ่งของปัญญาประดิษฐ์ ที่ทำให้คอมพิวเตอร์สามารถ เรียนรู้จากข้อมูล ได้เอง โดยไม่ต้องถูกตั้งโปรแกรมไว้อย่างชัดเจนสำหรับทุกงาน แทนที่จะบอกคอมพิวเตอร์ว่าต้องทำอะไรทีละขั้นตอน เราจะให้ข้อมูลจำนวนมากและให้มันค้นหารูปแบบจากข้อมูลนั้นๆ ด้วยตัวเอง เพื่อนำไปใช้ในการทำนายหรือตัดสินใจในอนาคต โดยจะแบ่งออกเป็น 3 ประเภท ได้แก่

1) การเรียนรู้แบบมีผู้สอน (Supervised Learning) คือประเภทของการเรียนรู้ของเครื่องที่แบบจำลองจะเรียนรู้จากชุดข้อมูลที่ มีป้ายกำกับ ซึ่งหมายความว่าข้อมูลแต่ละรายการมี "คำตอบ" หรือผลลัพธ์ที่ถูกต้องอยู่แล้ว แบบจำลองจะพยายามเรียนรู้ความสัมพันธ์ระหว่างข้อมูลเข้า และผลลัพธ์ที่คาดหวัง เพื่อใช้ในการทำนายผลลัพธ์สำหรับข้อมูลใหม่ โดยมักจะใช้งานกับงานประเภท การจำแนกประเภทและการพยากรณ์
2) การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการเรียนรู้ของเครื่องที่แบบจำลองจะทำงานกับ ข้อมูลที่ไม่มีป้ายกำกับ โดยไม่มีคำตอบที่ถูกต้องให้ แบบจำลองจะพยายามค้นหารูปแบบที่ซ่อนอยู่ โครงสร้าง หรือความสัมพันธ์ในข้อมูลด้วยตัวเอง จุดประสงค์หลักคือการทำความเข้าใจโครงสร้างของข้อมูลหรือการจัดกลุ่มข้อมูลที่มีลักษณะคล้ายกันเข้าด้วยกัน โดยมักจะใช้งานกับงานประเภท การจัดกลุ่ม การลดมิติข้อมูล และ การหาความสัมพันธ์
3) การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นแนวคิดที่แตกต่างออกไป โดยเน้นที่การที่ เอเจนต์ (agent) เรียนรู้ที่จะตัดสินใจและกระทำในสภาพแวดล้อม เพื่อให้บรรลุเป้าหมายที่กำหนด เอเจนต์จะเรียนรู้ผ่านการลองผิดลองถูก (trial-and-error) และได้รับรางวัล (reward) สำหรับการกระทำที่นำไปสู่ผลลัพธ์ที่ดี และ การลงโทษ (penalty) สำหรับการกระทำที่ไม่ดี เป้าหมายของเอเจนต์คือการเรียนรู้กลยุทธ์ (policy) ที่จะเพิ่มรางวัลรวมที่ได้รับสูงสุดในระยะยาว โดยมักจะใช้งานกับหุ่นยนต์ เกม และ ระบบแนะนำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.3 แสดงภาพการฝึกสอนของการเรียนรู้ของเครื่องแบบมีผู้สอนและไม่มีผู้สอน (ที่มา: kasunprageethdissanayake.medium.com/artificial-intelligence-2-supervised-learning-unsupervised-learning-and-reinforcement-learning-7bf00c732e99)

2.4 แบบจำลอง ARIMA (Autoregressive Integrated Moving Average)

แบบจำลอง ARIMA เป็นแบบจำลองทางสถิติที่ใช้สำหรับวิเคราะห์และพยากรณ์ข้อมูลอนุกรมเวลา โดยอาศัยความสัมพันธ์ระหว่างค่าปัจจุบันกับค่าที่ผ่านมา รวมถึงการปรับให้ข้อมูลมีความนิ่ง (stationary) และการใช้ค่าเฉลี่ยเคลื่อนที่ของค่าความผิดพลาดจากอดีต เพื่อทำให้การพยากรณ์แม่นยำยิ่งขึ้น โดย ARIMA นิยามด้วยตัวแปร (p,d,q) ดังนี้

1) AR (Autoregressive) แทนที่ด้วยตัวแปร p คือจำนวนของค่าล่าช้าของตัวแปรเป้าหมาย ที่ใช้ในแบบจำลองเพื่อทำนายค่าปัจจุบันด้วยการถดถอยเชิงเส้น
2) I (Integrated) แทนที่ด้วยตัวแปร d คือจำนวนครั้งที่ทำ differencing เพื่อลบแนวโน้มหรือฤดูกาล เพื่อทำให้ข้อมูลอยู่ในรูปแบบนิ่ง
3) MA (Moving Average) แทนที่ด้วยตัวแปร q คือจำนวนของค่าล่าช้า ของค่าผิดพลาดจากอดีตที่ใช้ในการพยากรณ์ค่าปัจจุบัน

จุดแข็งของ ARIMA คือเป็นแบบจำลองที่ใช้ข้อมูลในอดีตเพียงอย่างเดียวโดยไม่จำเป็นต้องมีตัวแปรต้นจากภายนอก ทำให้เหมาะสำหรับกรณีที่มีข้อมูลจำกัด และสามารถจัดการกับข้อมูลที่ไม่นิ่งได้ด้วยการปรับให้นิ่งผ่านการ differencing นอกจากนี้ ยังเหมาะสำหรับการพยากรณ์ระยะสั้นและมีพื้นฐานทางสถิติที่ชัดเจน ช่วยให้สามารถวิเคราะห์ผลลัพธ์และสมมติฐานได้อย่างเป็นระบบ

อย่างไรก็ตาม ARIMA ก็มีข้อจำกัดที่สำคัญ เช่น ไม่เหมาะกับการพยากรณ์ระยะยาวหรือข้อมูลที่มีรูปแบบฤดูกาล ซึ่งจะต้องพัฒนาไปใช้แบบจำลอง SARIMA แทน หรือใช้ SARIMAX แทนหากมีข้อมูลตัวแปรต้นหลายตัว นอกจากนี้ ยังไม่สามารถคาดการณ์จุดเปลี่ยนแนวโน้ม ได้ดี และต้องใช้ข้อมูลที่มีความต่อเนื่องโดยไม่มีค่าว่าง

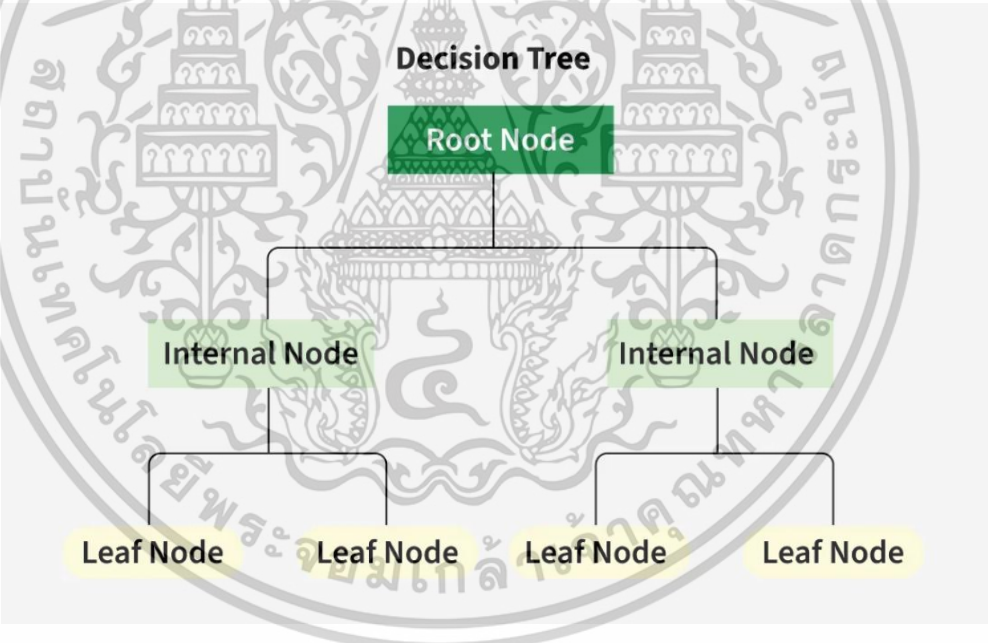
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 แบบจำลองต้นไม้ตัดสินใจ (Decision Tree)

แบบจำลองต้นไม้ตัดสินใจ เป็นอัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน โดยอาศัยแนวคิดที่เข้าใจง่ายและสามารถตีความได้อย่างชัดเจน ซึ่งโครงสร้างของแบบจำลองต้นไม้ตัดสินใจ มีลักษณะเป็นลำดับชั้นของเงื่อนไขที่แตกแขนงเป็นกิ่งและใบ เพื่อแบ่งกลุ่มข้อมูลตามเกณฑ์ที่กำหนด

องค์ประกอบของต้นไม้ตัดสินใจจะประกอบด้วย

1) โหนดราก (Root Node) เป็นโหนดเริ่มต้นของต้นไม้ ซึ่งแทนกลุ่มข้อมูลทั้งหมด และจะแตกออกเป็นกิ่งตามเงื่อนไขที่ดีที่สุดในการแบ่งแยกข้อมูล
2) โหนดตัดสินใจ (Decision Nodes) เป็นโหนดที่อยู่ระหว่างทาง ซึ่งแสดงการตัดสินใจหรือเงื่อนไขที่ใช้แบ่งข้อมูลออกเป็นกลุ่มย่อย
3) โหนดใบ (Leaf Nodes) เป็นโหนดสุดท้ายที่ไม่สามารถแบ่งได้อีก แสดงผลลัพธ์หรือการพยากรณ์
4) กิ่ง เป็นเส้นทางเชื่อมระหว่างโหนด ซึ่งแทนการไหลของเงื่อนไข



รูปที่ 2.4 แสดงแผนภาพการทำงานของแบบจำลองต้นไม้ตัดสินใจ
(ที่มา: www.geeksforgeeks.org/decision-tree)

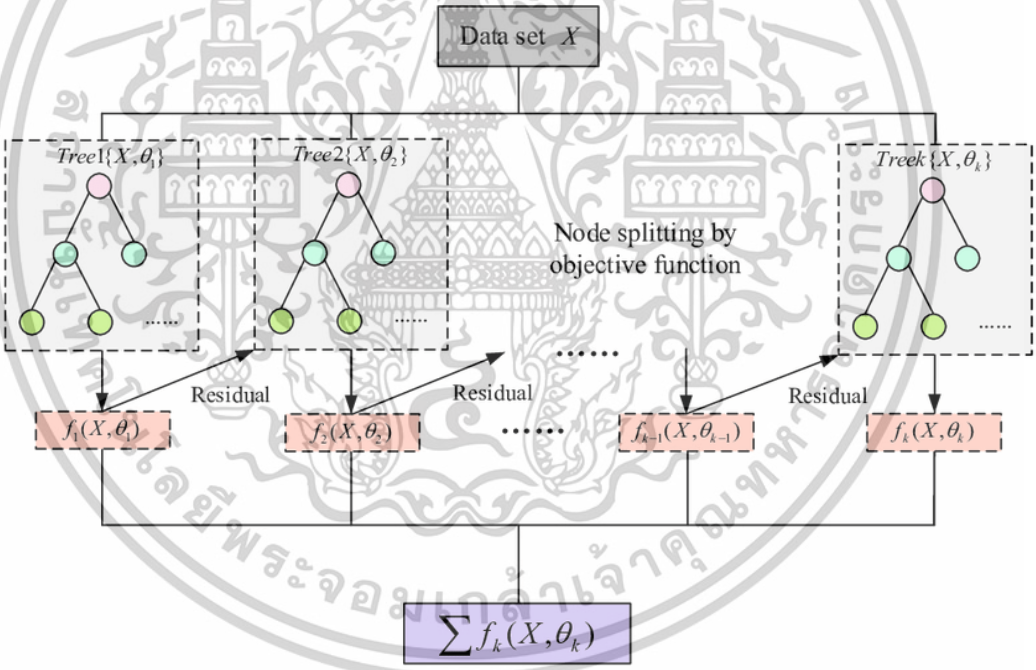
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6 แบบจำลอง XGBoost (Extreme Gradient Boosting)

แบบจำลอง XGBoost เป็นอัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน ที่อยู่บนพื้นฐานของการสร้างจากแบบจำลองการตัดสินใจแบบต้นไม้หลายชุด โดยนำแนวคิดของ Gradient Boosting มาพัฒนาเพิ่มเติมเพื่อให้สามารถฝึกแบบจำลองได้รวดเร็วและแม่นยำยิ่งขึ้น

หลักการทำงานของ XGBoost คือการฝึกแบบจำลองใหม่ทีละรอบ โดยแต่ละรอบจะสร้างต้นไม้การตัดสินใจ ที่เรียนรู้จากข้อผิดพลาดของแบบจำลองก่อนหน้า โดยใช้ Gradient ของค่าความผิดพลาด เพื่อคำนวณทิศทางและขนาดของการปรับค่า การเรียนรู้จึงค่อย ๆ ปรับค่าทำนายให้ใกล้เคียงค่าจริงมากยิ่งขึ้นในแต่ละรอบ

XGBoost มีจุดเด่นที่สำคัญคือความสามารถในการควบคุมปัญหา overfitting ด้วยเทคนิค regularization ที่ช่วยลดความซับซ้อนของแบบจำลอง รองรับการประมวลผลข้อมูลที่มีค่าหายไปโดยไม่จำเป็นต้องเติมค่าล่วงหน้า และสามารถประมวลผลแบบขนานได้อย่างรวดเร็ว ทำให้เหมาะสมสำหรับข้อมูลขนาดใหญ่และซับซ้อน นอกจากนี้ยังสามารถปรับแต่งพารามิเตอร์ได้หลากหลายเพื่อเพิ่มประสิทธิภาพในการเรียนรู้ ส่งผลให้ XGBoost เป็นแบบจำลองที่ได้รับความนิยมในวงการวิทยาศาสตร์ข้อมูลและการวิเคราะห์เชิงพยากรณ์อย่างแพร่หลาย



รูปที่ 2.5 แสดงแผนภาพการทำงานของแบบจำลอง XGBoost
(ที่มา: www.geeksforgeeks.org/xgboost)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7 แบบจำลองป่าสุ่ม (Random Forest)

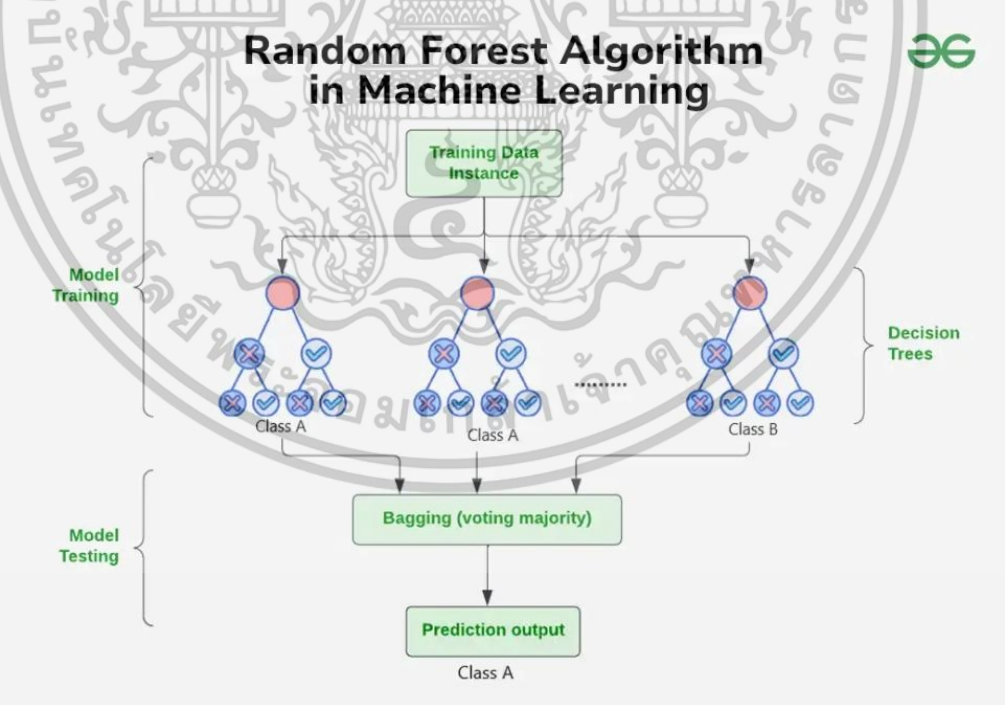
แบบจำลองป่าสุ่ม เป็นอัลกอริทึมการเรียนรู้ของเครื่องแบบมีผู้สอน ซึ่งถูกพัฒนาขึ้นจากแนวคิดของการรวมหลายแบบจำลอง โดยเฉพาะการรวมของต้นไม้ตัดสินใจ หลายต้นเข้าด้วยกันเพื่อให้การพยากรณ์หรือการจัดประเภทมีความแม่นยำมากขึ้นและลดความลำเอียงของแบบจำลองเดี่ยว

หลักการทำงานของแบบจำลองป่าสุ่ม คือการสร้างต้นไม้หลายต้นจากชุดข้อมูลฝึกที่แตกต่างกันเล็กน้อย (ผ่านการสุ่มตัวอย่างแบบ Bootstrap) และในระหว่างการสร้างแต่ละต้นไม้ จะมีการสุ่มเลือกเฉพาะบางคุณลักษณะ มาใช้ในการตัดสินใจแต่ละโหนด ซึ่งช่วยลดความสัมพันธ์ระหว่างต้นไม้แต่ละต้น และเพิ่มความสามารถในการทั่วไปของแบบจำลอง

เมื่อทำการพยากรณ์ ผลลัพธ์จากต้นไม้ทุกต้นจะถูกรวบรวมมารวมกัน โดยในกรณีของการถดถอย จะใช้ค่าเฉลี่ยของผลลัพธ์ และในกรณีของการจำแนกประเภท จะใช้การโหวตเสียงข้างมาก ซึ่งช่วยเพิ่มความแม่นยำและความเสถียรของการพยากรณ์เมื่อเปรียบเทียบกับการใช้ต้นไม้เดี่ยว

จุดแข็งของแบบจำลองป่าสุ่ม คือความสามารถในการจัดการกับข้อมูลที่มีมิติสูง, การป้องกัน overfitting ได้ดีในหลายกรณี, และไม่จำเป็นต้องปรับแต่งตัวแปรมากนัก นอกจากนี้ยังสามารถจัดการกับข้อมูลที่มีค่าว่าง ได้อย่างมีประสิทธิภาพ และสามารถวัดความสำคัญของตัวแปร เพื่อการตีความเชิงวิเคราะห์เพิ่มเติมได้อีกด้วย

ด้วยคุณสมบัติเหล่านี้ แบบจำลองป่าสุ่ม จึงเป็นหนึ่งในแบบจำลองยอดนิยมสำหรับงานพยากรณ์และการวิเคราะห์ข้อมูลในสาขาต่าง ๆ รวมถึงการพยากรณ์ราคาสินค้าโภคภัณฑ์ เช่น ราคาทองแดง



รูปที่ 2.6 แสดงแผนภาพการทำงานของแบบจำลองป่าสุ่ม
(ที่มา: www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8 แบบจำลอง LSTM (Long Short-Term Memory)

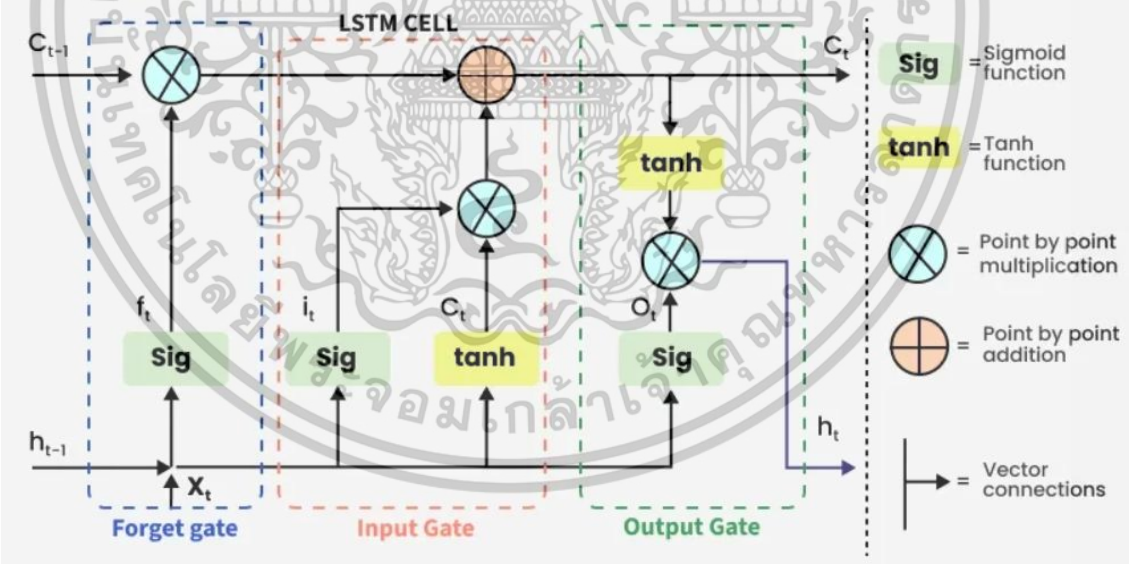
แบบจำลอง LSTM เป็นโครงข่ายประสาทเทียมแบบลึก (Deep Neural Network) ที่อยู่ในกลุ่มของ Recurrent Neural Networks (RNNs) ซึ่งได้รับการออกแบบมาเพื่อจัดการกับปัญหาที่ RNNs แบบดั้งเดิมไม่สามารถจัดการได้อย่างมีประสิทธิภาพ เช่น การลืมข้อมูลในลำดับเวลาที่ยาวนาน

โครงสร้างของแบบจำลอง LSTM ได้รับการพัฒนาขึ้นมาโดยมีการเพิ่ม “หน่วยความจำ” (memory cell) ซึ่งสามารถรักษาข้อมูลสำคัญไว้ได้ในช่วงเวลาที่ยาวนาน และมีการควบคุมด้วยโครงสร้างประตู (gates) ได้แก่

1) Forget Gate ทำหน้าที่ตัดสินใจว่าจะลบข้อมูลใดออกจากหน่วยความจำ
2) Input Gate ควบคุมว่าข้อมูลใหม่ใดควรถูกเก็บเข้าไปในหน่วยความจำ
3) Output Gate กำหนดว่าข้อมูลส่วนใดในหน่วยความจำควรถูกนำมาใช้เพื่อส่งออกในขั้นตอนปัจจุบัน

ด้วยกลไกเหล่านี้แบบจำลอง LSTM จึงสามารถจดจำรูปแบบจากข้อมูลอนุกรมเวลาได้อย่างมีประสิทธิภาพ ทั้งในลำดับเวลาสั้นและยาว โดยไม่เกิดปัญหา Gradient Vanishing และ Gradient Exploding ที่มักพบใน RNN ทั่วไป

แบบจำลอง LSTM นิยมใช้ในการวิเคราะห์และพยากรณ์ข้อมูลที่มีลักษณะตามลำดับเวลา เช่น การพยากรณ์หุ้น การวิเคราะห์ข้อความ และการรู้จำเสียง สำหรับงานพยากรณ์ราคาทองแดง แบบจำลอง LSTM ถือเป็นแบบจำลองที่มีความสามารถสูงในการเรียนรู้จากข้อมูลราคาในอดีต เพื่อทำนายแนวโน้มในอนาคต โดยเฉพาะเมื่อต้องจัดการกับข้อมูลที่มีความซับซ้อนหรือมีความแปรปรวนสูง



รูปที่ 2.7 แสดงแผนภาพการทำงานของแบบจำลอง LSTM
(ที่มา: www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.9 การตรวจสอบค่าสุดโต่ง (Outlier Detection)

ในการวิเคราะห์ข้อมูลเชิงสถิติและการพัฒนาแบบจำลองทางการเรียนรู้ของเครื่อง ค่าสุดโต่ง หรือ ค่าผิดปกติ คือค่าข้อมูลที่เบี่ยงเบนอย่างมากจากแนวโน้มหลักของชุดข้อมูล อาจเกิดจากข้อผิดพลาดในการวัด การบันทึกข้อมูล หรือในบางกรณีอาจสะท้อนถึงเหตุการณ์ที่เกิดขึ้นจริงแต่ไม่ปกติ ตัวอย่างเช่น ราคาสินทรัพย์ที่ผันผวนเฉียบพลัน หรือยอดขายที่เพิ่มขึ้นผิดปกติในช่วงเทศกาล

ค่าสุดโต่ง มีผลอย่างมากต่อการวิเคราะห์ทางสถิติ เช่น ทำให้ค่าเฉลี่ยบิดเบือน หรือนำไปสู่แบบจำลองที่มีประสิทธิภาพต่ำหากไม่จัดการอย่างเหมาะสม

2.9.1 การตรวจสอบค่าสุดโต่งด้วยวิธี Interquartile Range (IQR)

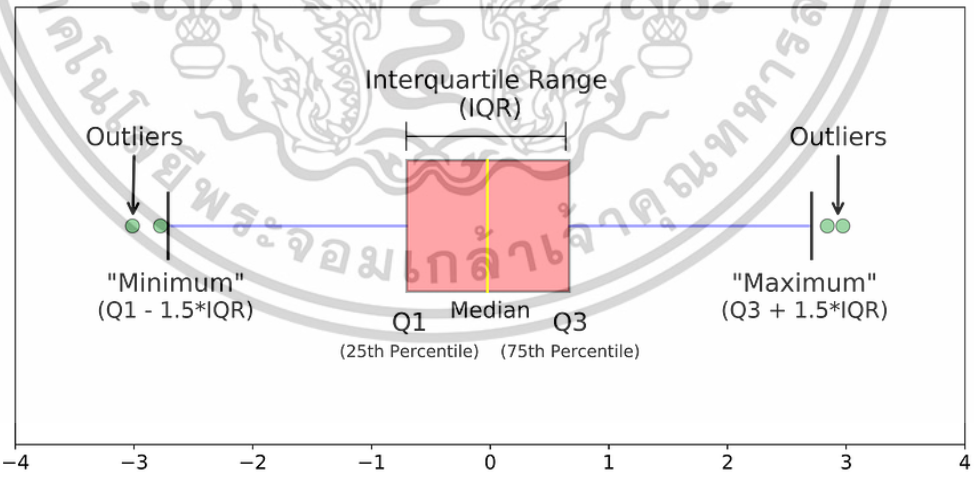
เป็นเทคนิคที่ใช้วัดการกระจายของข้อมูลโดยไม่อ่อนไหวต่อค่าที่เบี่ยงเบนมากเกินไป ทำให้เหมาะสมกับข้อมูลที่มีการแจกแจงไม่สมมาตรหรือข้อมูลที่มีความเบ้

วิธี IQR ใช้ควอไทล์ที่ 1 (Q_1) หรือค่าที่แบ่งข้อมูลส่วนล่าง 25% และควอไทล์ที่ 3 (Q_3) หรือค่าค่าที่แบ่งข้อมูลส่วนบน 75% ซึ่งเป็นค่าที่แบ่งข้อมูลออกเป็นสี่ส่วนอย่างเท่า ๆ กัน โดยช่วงระหว่างควอไทล์นี้สามารถใช้ในการประเมินค่าที่เบี่ยงเบนออกไปจากค่ากลางโดย IQR คือค่าความกว้างระหว่างควอไทล์ ใช้เป็นเกณฑ์วัดการกระจายของข้อมูลส่วนกลางได้ดังสมการ

$$IQR = Q_3 - Q_1$$

จากนั้น กำหนดช่วงค่าปกติของข้อมูลไว้ในช่วงขอบล่าง (Lower Bound) และขอบบน (Upper Bound) ดังสมการ

$$Lower\ Bound = Q_1 - 1.5 \times IQR$$
$$Upper\ Bound = Q_3 + 1.5 \times IQR$$



รูปที่ 2.8 แสดงแผนภาพการตรวจสอบค่าสุดโต่งด้วยวิธี IQR

(ที่มา: blog.alliedoffsets.com/beyond-the-norm-how-outlier-detection-transforms-data-analysis)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.10 วิศวกรรมคุณลักษณะ (Feature Engineering)

วิศวกรรมคุณลักษณะ คือกระบวนการสร้างคุณลักษณะใหม่หรือการแปลงคุณลักษณะที่มีอยู่เพื่อปรับปรุงประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง โดยเกี่ยวข้องกับการเลือกข้อมูลที่เกี่ยวข้องจากข้อมูลดิบและแปลงเป็นรูปแบบที่แบบจำลองสามารถเข้าใจได้ง่าย เป้าหมายคือการปรับปรุงความแม่นยำของแบบจำลองโดยการให้ข้อมูลที่มีความหมายและเกี่ยวข้องมากขึ้น กระบวนการที่เกี่ยวข้องในการวิศวกรรมคุณลักษณะ ได้แก่

2.10.1 การสร้างคุณลักษณะ (Feature Creation)

เป็นกระบวนการสร้างคุณลักษณะใหม่จากข้อมูลเดิม โดยอาศัยความรู้เฉพาะทางหรือการสังเกตรูปแบบที่ซ่อนอยู่ในข้อมูล ประเภทของการสร้างคุณลักษณะ ได้แก่:

1) ตามความรู้เฉพาะด้าน (Domain-Driven) ใช้ความเข้าใจเกี่ยวกับโดเมนเพื่อสร้างคุณลักษณะที่คาดว่าจะมีประโยชน์ต่อแบบจำลอง
2) ขับเคลื่อนโดยข้อมูล (Data-Driven) ใช้วิธีการเชิงวิเคราะห์เพื่อค้นหารูปแบบหรือความสัมพันธ์ที่มีนัยสำคัญ
3) การสังเคราะห์ (Synthesis) เป็นการรวมคุณลักษณะที่มีอยู่หลายตัวเข้าด้วยกัน เช่น การบวก ลบ หรือคำนวณอัตราส่วน

2.10.2 การแปลงคุณลักษณะ (Feature Transformation)

เป็นการปรับเปลี่ยนคุณลักษณะให้เหมาะสมกับการเรียนรู้ของโมเดล โดยไม่เปลี่ยนปริมาณข้อมูล ประเภทที่พบบ่อย ได้แก่:

1) การปรับมาตรฐาน (Normalization): เช่น การปรับค่าคุณลักษณะให้อยู่ในช่วง 0 ถึง 1
2) การปรับสเกล (Scaling) เช่น Standard Scaling เพื่อให้ข้อมูลมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1
3) การเข้ารหัส (Encoding) เช่น การแปลงข้อมูลประเภทหมวดหมู่เป็นตัวเลข
4) การแปลงเชิงคณิตศาสตร์ (Mathematical Transformation) เช่น การใช้ log, square root, หรือ power transformation

2.10.3 การสกัดคุณลักษณะ (Feature Extraction)

หมายถึงการแปลงชุดคุณลักษณะเดิมให้กลายเป็นชุดคุณลักษณะใหม่ที่สรุปสาระสำคัญของข้อมูลได้ดีขึ้น ประเภทของการสกัดคุณลักษณะได้แก่:

1) การลดมิติ (Dimensionality Reduction) เช่น PCA (Principal Component Analysis)
2) การรวมคุณลักษณะ (Feature Combination) เช่น การรวมค่าคุณลักษณะหลายตัวเข้าด้วยกัน
3) การรวมกลุ่ม (Feature Aggregation) เช่น การหาค่าเฉลี่ย ค่ามัธยฐาน หรือค่ารวมในช่วงเวลาต่าง ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.10.4 การเลือกคุณลักษณะ (Feature Selection)

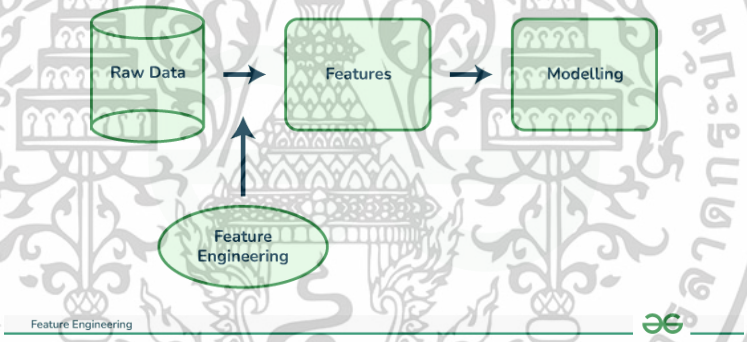
เป็นการคัดเลือกเฉพาะคุณลักษณะที่มีความเกี่ยวข้องกับตัวแปรเป้าหมาย โดยช่วยลดมิติข้อมูลและป้องกัน overfitting ประเภทของเทคนิคที่ใช้ได้แก่:

1) Filter Method เลือกคุณลักษณะโดยใช้เกณฑ์ทางสถิติ เช่น ค่า correlation
2) Wrapper Method ใช้แบบจำลองเพื่อประเมินประสิทธิภาพของชุดคุณลักษณะย่อย
3) Embedded Method: เป็นส่วนหนึ่งของการฝึกแบบจำลอง เช่น Lasso Regression ที่เลือกคุณลักษณะโดยอัตโนมัติผ่านการถ่วงน้ำหนัก

2.10.5 การปรับสเกลคุณลักษณะ (Feature Scaling)

เป็นการทำให้คุณลักษณะทั้งหมดมีช่วงค่าที่ใกล้เคียงกัน ซึ่งช่วยให้แบบจำลองที่อ่อนไหวต่อสเกลทำงานได้ดีขึ้น ประเภทที่นิยมใช้ได้แก่:

1) Min-Max Scaling ปรับค่าคุณลักษณะให้อยู่ในช่วง 0 ถึง 1
2) Standard Scaling ปรับให้มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1
3) Robust Scaling ใช้ค่ามัธยฐานและ IQR เพื่อให้ทนต่อค่าผิดปกติ (outliers)



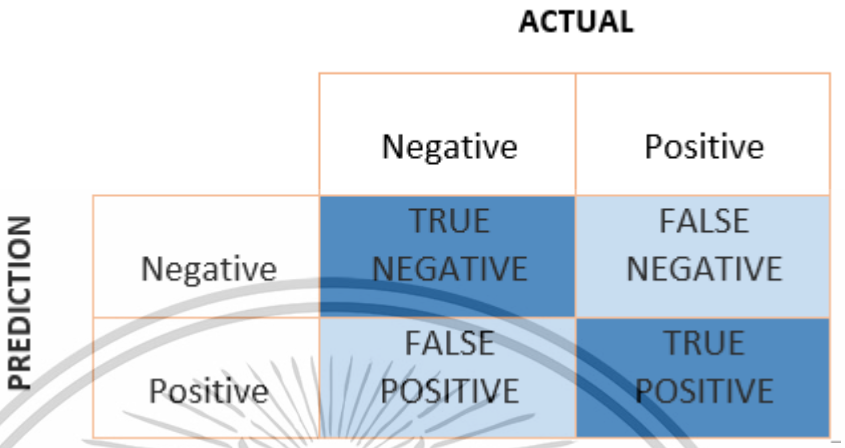
รูปที่ 2.9 แสดงแผนภาพการสร้างแบบจำลอง
(ที่มา: www.geeksforgeeks.org/what-is-feature-engineering)

2.11 การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation)

การประเมินประสิทธิภาพนั้นจำเป็นกับการทำแบบจำลองอย่างมาก เพราะเป็นสิ่งที่เอาไว้วัดว่าแบบจำลองนั้นๆ ทำงานเป็นอย่างไร ประเมินผลว่าทำงานได้น่าพึงพอใจหรือไม่ รวมถึงยังไว้ใช้เปรียบเทียบระหว่างแบบจำลองเพื่อตอบคำถามที่ว่าเราควรเลือกใช้โมเดลไหนดีได้อีกด้วย ซึ่งการประเมินสำหรับแบบจำลองนั้นจะนำเสนอด้วยกัน 2 ประเภท คือ การประเมินประสิทธิภาพสำหรับงานด้านการจำแนกประเภท และการประเมินประสิทธิภาพสำหรับงานด้านการทำนายค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.11.1 การประเมินประสิทธิภาพสำหรับงานด้านจำแนกประเภท



รูปที่ 2.10 แสดงตาราง Confusion Matrix
(ที่มา: towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262)

การประเมินประสิทธิภาพสำหรับงานด้านการจำแนกประเภทจำเป็นต้องเข้าใจ Confusion Matrix หรือ ตารางขนาด 2x2 ที่ใช้แสดงผลการทำนายของแบบจำลองเปรียบเทียบกับค่าจริง โดยมีองค์ประกอบหลัก 4 ส่วน ได้แก่

- True Positive (TP) คือ กรณีที่แบบจำลองทำนายว่าเป็นบวก และในความเป็นจริงก็เป็นบวก
- True Negative (TN) คือ กรณีที่แบบจำลองทำนายว่าเป็นลบ และในความเป็นจริงก็เป็นลบ
- False Positive (FP) คือ กรณีที่แบบจำลองทำนายว่าเป็นบวก แต่ในความเป็นจริงเป็นลบ
- False Negative (FN) คือ กรณีที่แบบจำลองทำนายว่าเป็นลบ แต่ในความเป็นจริงเป็นบวก

เมื่อทราบองค์ประกอบหลักทั้ง 4 ส่วนแล้ว จะนำค่าเหล่านี้มาพิจารณาเพื่อประเมินประสิทธิภาพของแบบจำลองได้ดังสมการต่อไปนี้

1) ความแม่นยำ (Accuracy) คือ อัตราส่วนของจำนวนครั้งที่แบบจำลองทำนายได้ถูกต้อง (ทั้งกรณีบวกและลบ) ต่อจำนวนตัวอย่างทั้งหมดในชุดข้อมูล

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2) ความเที่ยงตรง (Precision) คือ อัตราส่วนของจำนวนครั้งที่แบบจำลองทำนายว่าเป็นบวกแล้วถูกต้อง ต่อจำนวนครั้งทั้งหมดที่แบบจำลองทำนายว่าเป็นบวก

$$Precision = \frac{TP}{TP + FP}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) ค่าการเรียกคืน (Recall) คือ อัตราส่วนของจำนวนค่าบวกจริงที่แบบจำลองสามารถทำนายได้ถูกต้อง ต่อจำนวนค่าบวกจริงทั้งหมดในข้อมูล

$$Recall = \frac{TP}{TP + FN}$$

4) คะแนนเอฟวัน (F1 Score) คือ ค่าเฉลี่ยฮาร์มอนิกของความเที่ยงตรงและค่าการเรียกคืน ซึ่งใช้เพื่อประเมินความสามารถของแบบจำลองในการทำนายค่าบวกได้อย่างแม่นยำและครอบคลุม

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

2.11.2 การประเมินประสิทธิภาพสำหรับงานด้านทำนายค่า

ในการประเมินประสิทธิภาพของแบบจำลองที่ใช้สำหรับงานด้านการพยากรณ์หรือการทำนายค่า นิยมใช้ตัวชี้วัดที่เปรียบเทียบค่าที่ทำนายได้กับค่าจริง เพื่อวัดความคลาดเคลื่อนของแบบจำลอง โดยตัวชี้วัดที่นิยมใช้มีดังนี้

1) ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE)

คือ ค่าเฉลี่ยของค่าความคลาดเคลื่อนในแต่ละจุดข้อมูล โดยไม่สนใจทิศทางของความคลาดเคลื่อน MAE เหมาะสำหรับวัดค่าความคลาดเคลื่อนโดยรวมที่เข้าใจง่ายและมีหน่วยเดียวกับข้อมูลที่ทำนาย

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

2) ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสอง (Root Mean Squared Error: RMSE)

คือ ค่ารากที่สองของค่าเฉลี่ยของค่าความคลาดเคลื่อนยกกำลังสอง โดยจะให้น้ำหนักกับค่าความคลาดเคลื่อนที่มากกว่าปกติ RMSE จะตอบสนองต่อ outlier ได้มากกว่า MAE เนื่องจากยกกำลังสอง

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

3) ค่าความคลาดเคลื่อนสัมพัทธ์เฉลี่ยร้อยละ (Mean Absolute Percentage Error: MAPE)

คือ ค่าเฉลี่ยของสัดส่วนค่าความคลาดเคลื่อนเมื่อเทียบกับค่าจริง โดยแสดงในรูปเปอร์เซ็นต์ MAPE ช่วยให้สามารถตีความความคลาดเคลื่อนได้ง่ายในแง่ของเปอร์เซ็นต์ แต่อาจมีปัญหาเมื่อค่าจริงใกล้ศูนย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

4) ค่าสัมประสิทธิ์การตัดสินใจ (R-squared: R^2)

คือ สัดส่วนของความแปรปรวนในข้อมูลจริงที่สามารถอธิบายได้ด้วยแบบจำลอง ยิ่งเข้าใกล้ 1 แสดงว่าแบบจำลองสามารถอธิบายข้อมูลได้ดี

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

2.12 งานวิจัยที่เกี่ยวข้อง

Sánchez Lasheras, et al. (2015) ศึกษาการพยากรณ์ราคาทองแดงในตลาด COMEX โดยเปรียบเทียบแบบจำลอง ARIMA กับแบบจำลองโครงข่ายประสาทเทียม MLP และ Elman RNN โดยใช้ข้อมูลรายวันระหว่างปี ค.ศ. 2002–2014 ผลพบว่าแบบจำลอง Elman RNN ให้ผลแม่นยำที่สุด รองลงมาคือ MLP และ ARIMA ตามลำดับโดยใช้ MAPE เป็นตัวประสิทธิภาพ พร้อมเสนอให้พัฒนาแบบจำลองการเรียนรู้แบบลูกผสมเพื่อเพิ่มประสิทธิภาพการพยากรณ์

Luo, et al. (2022) ศึกษาการพยากรณ์ราคาทองแดงแบบหลายช่วงเวลาโดยการปรับปรุงแบบจำลอง LSTM โดยใช้ข้อมูลรายเดือนระหว่างปี ค.ศ. 1991-2021 โดยใช้ตัวแปรตามราคาทองแดง และตัวแปรต้นเช่น โลหะทองคำ และเหล็ก และพลังงานเช่น น้ำมัน และถ่านหิน พร้อมระบบแก้ไขความผิดพลาด (error correction) เพื่อพยากรณ์ราคาทองแดงล่วงหน้า พบว่าแบบจำลอง F-TS-GA-LSTM-EC มีความแม่นยำสูงที่สุดเมื่อเทียบกับแบบจำลอง BPNN, ELM และ LSTM ทั่วไป

Hu, J. (2023) ศึกษาการพยากรณ์ราคาทองแดงในระยะสั้นด้วยข้อมูลขนาดเล็ก โดยใช้ข้อมูลรายวันระหว่างเดือนพฤศจิกายน ค.ศ. 2021 ถึงเดือนธันวาคม ค.ศ. 2022 รวม 273 วัน โดยใช้ราคาทองแดงเป็นตัวแปรตาม และใช้ตัวแปรต้น เช่น ราคาน้ำมันดิบ ราคาโลหะอื่น ดัชนีราคาหุ้น และอัตราแลกเปลี่ยนเงินตราต่างประเทศ งานวิจัยได้เปรียบเทียบแบบจำลองทางสถิติ ได้แก่ AR และ VAR กับแบบจำลองการเรียนรู้ของเครื่อง XGBoost พบว่า XGBoost ให้ค่าความคลาดเคลื่อนต่ำที่สุดในช่วงพยากรณ์ 1–3 วัน ขณะที่ VAR มีความสามารถแข่งขันได้ในกรณีการพยากรณ์ระยะสั้นเพียง 1 วัน นอกจากนี้ยังเสนอว่าการเลือกค่าความล่าช้าที่เหมาะสมจากโมเดล VAR สามารถช่วยเพิ่มประสิทธิภาพของ XGBoost ได้อย่างมีนัยสำคัญ

Li, et al. (2024) ศึกษาการพยากรณ์ราคาทองแดง โดยพัฒนาแบบจำลองการเรียนรู้แบบผสมผสานด้วยการรวมแบบจำลอง DELM, XGBoost และ LSTM เข้าด้วยกันโดยแบบจำลอง CNN เป็น meta-model และปรับพารามิเตอร์ด้วยวิธีการ Sparrow Search Algorithm (SSA) ร่วมกับเทคนิค Adversarial Interpretive Structural Model (AISM) วิเคราะห์หาตัวแปรต้นที่ส่งผลมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่สุด โดยใช้ข้อมูลรายเดือนตั้งแต่เดือนเมษายน ค.ศ. 1996 ถึงเดือนมิถุนายน ค.ศ. 2022 พบว่าแบบจำลอง DELM เป็นแบบจำลองที่แม่นยำที่สุดโดย MAPE มีค่า 4.45% และเมื่อผสมผสานแบบจำลองแล้ว ค่าความคลาดเคลื่อนต่ำลงเหลือ 3.59%

ตารางที่ 2.1 แสดงการเปรียบเทียบงานวิจัยที่เกี่ยวข้องกับงานวิจัยฉบับนี้

ผู้วิจัย	ข้อมูลที่ใช้	เทคนิคที่ใช้	การวัดประสิทธิภาพ	ผลการวิจัย
Sánchez Lasheras, et al. (2015)	ข้อมูลรายวัน ค.ศ. 2002-2014	ARIMA, MLP, และ Elman RNN	MAPE	Elman RNN ให้ผลแม่นยำที่สุด รองลงมาคือ MLP และ ARIMA ตามลำดับ
Luo, et al. (2022)	ข้อมูลรายเดือน ค.ศ. 1991-2021	GA-LSTM	MAE, RMSE, MAPE, TIC, และ R^2	F-TS-GA-LSTM-EC มีความแม่นยำที่สุดเมื่อเทียบกับแบบจำลอง BPNN, ELM และ LSTM ทั่วไป
Hu, J. (2023)	ข้อมูลรายวัน ค.ศ. 2021-2022	AR, VAR, และ XGBoost	MAE, MSE, และ RMSE	XGBoost ให้ค่าความคลาดเคลื่อนต่ำที่สุดในช่วงพยากรณ์ 1–3 วัน ขณะที่ VAR มีความสามารถแข่งขันได้ในการพยากรณ์ 1 วัน นอกจากนี้ยังเสนอว่าการเลือกค่าความล่าช้าที่เหมาะสมจากโมเดล VAR สามารถช่วยเพิ่มประสิทธิภาพของ XGBoost
Li, et al. (2024)	ข้อมูลรายเดือน ค.ศ. 1996–2022	DELM, XGBoost, LSTM, และ CNN ที่รวม DELM, XGBoost, และ LSTM เข้าด้วยกัน	MAE, RMSE, MAPE, และ R^2	DELM ให้ค่าความคลาดเคลื่อนต่ำที่สุดสำหรับแบบจำลองเดี่ยว แต่เมื่อผสมผสานแบบจำลองส่งผลให้สามารถทำนายได้มีประสิทธิภาพยิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

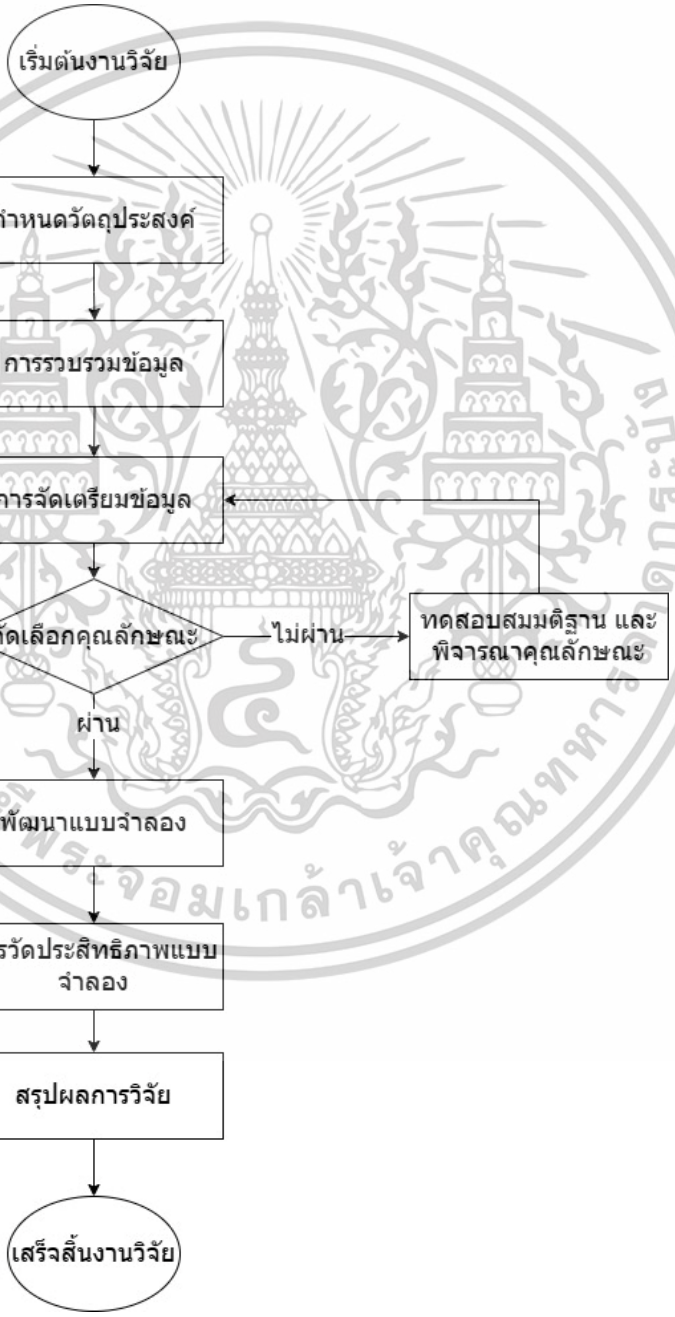
งานวิจัยฉบับนี้	ข้อมูลรายวัน ค.ศ. 2010-2024	ARIMA, ต้นไม้ตัดสินใจ, XGBoost, ป่าสุ่ม, และ LSTM	MAE, RMSE, และ MAPE	XGBoost มีประสิทธิภาพสูงที่สุด รองลงมาเป็น ป่าสุ่ม, LSTM, ต้นไม้ตัดสินใจ, และ ARIMA ตามลำดับ โดยตัวแปรตามทั้งหมดส่งผลให้สามารถทำนายได้อย่างมีประสิทธิภาพยิ่งขึ้น

งานวิจัยฉบับนี้เมื่อเปรียบเทียบกับงานวิจัยที่เกี่ยวข้องจะมีความเหมือน และความแตกต่างดังที่แสดงในตารางที่ 2.1 โดยภาพรวมแล้วจะเห็นได้ว่าการพยากรณ์ราคาทองแดงจะใช้ข้อมูลที่เป็นอนุกรมเวลา แต่ความถี่ของชุดข้อมูลจะขึ้นกับวัตถุประสงค์ และความสนใจของผู้วิจัย นอกจากนี้เมื่อพิจารณาถึงเทคนิคที่ใช้ในการวิจัยจะเห็นได้ว่า เทคนิคชนิดเดียวกันมักจะมีผลการวิจัยไปในแนวทางที่สอดคล้องกันโดยเฉพาะแบบจำลอง XGBoost ซึ่งงานวิจัยนี้เลือกใช้ แสดงให้เห็นถึงความน่าเชื่อถือ และสอดคล้องกับงานวิจัยก่อนหน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อ พัฒนา และเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่อง และแบบจำลองทางสถิติ สำหรับพยากรณ์ราคาทองแดง พร้อมทั้งศึกษาปัจจัยที่มีอิทธิพลต่อการพยากรณ์ราคาทองแดง โดยผู้วิจัยได้นำความรู้ ทฤษฎี และงานวิจัยที่เกี่ยวข้องมาวางแผนขั้นตอนในการทำวิจัย



รูปที่ 3.1 แสดงขั้นตอนการดำเนินงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1 ขั้นตอนการดำเนินงานวิจัย

การศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์ราคาทองแดงโดยใช้เทคนิคการเรียนรู้ของเครื่อง ซึ่งกระบวนการวิจัยประกอบด้วยลำดับขั้นตอนอย่างเป็นระบบ เพื่อให้เกิดความเข้าใจทั้งในด้านการเตรียมข้อมูล การเลือกตัวแปร การสร้างแบบจำลอง และการประเมินผลความแม่นยำของการพยากรณ์ โดยสามารถสรุปเป็นขั้นตอนหลักได้ตามแผนภาพในรูปที่ 3.1

3.2 เครื่องมือที่ใช้ในการวิจัย

การวิจัยฉบับนี้ใช้ภาษาคอมพิวเตอร์ไพธอน 3 (Python 3) ผ่านแพลตฟอร์ม Google Colab เป็นเครื่องมือหลักในการศึกษาวิจัย ทั้งรวบรวมข้อมูล จัดเตรียมข้อมูล พัฒนาแบบจำลอง และวัดประสิทธิภาพของแบบจำลอง

3.3 การรวบรวมข้อมูล

ข้อมูลที่ใช้ในงานวิจัย รวบรวมจากหน่วยงานที่เกี่ยวข้องที่เผยแพร่ในอินเตอร์เน็ต โดยทำการรวบรวมชุดข้อมูลราคาทองแดงในอดีตตั้งแต่วันที่ 4 มกราคม พ.ศ. 2553 ถึง 31 ธันวาคม พ.ศ. 2567 ประกอบด้วยข้อมูล 3,685 แถว หรือ วัน และตัวแปร 7 คุณลักษณะ ได้แก่

1) วันที่ ที่ทำการเก็บข้อมูล
2) ราคาปิดรายวันของโลหะทองแดง
3) ราคาเปิดรายวันของโลหะทองแดง
4) ราคาสูงสุดของโลหะทองแดง ณ วันที่ทำการเก็บข้อมูล
5) ราคาต่ำสุดของโลหะทองแดง ณ วันที่ทำการเก็บข้อมูล
6) ปริมาณการซื้อขายของโลหะทองแดง ณ วันที่เก็บข้อมูล
7) อัตราการเปลี่ยนแปลงเมื่อเทียบกับวันก่อนหน้า

จากนั้นได้พิจารณาจากความรู้ของทางผู้จัดทำและการศึกษาการวิจัยที่เกี่ยวข้อง จึงได้รวบรวมข้อมูลที่คาดว่าจะส่งผลต่อความผันผวนของราคาทองแดงเพิ่มเติมประกอบด้วยข้อมูลรายปีของค่าผลิตภัณฑ์มวลรวมในประเทศรายปีตั้งแต่ปี พ.ศ. 2553 ถึง ปี พ.ศ. 2567 ของประเทศที่มีมูลค่าการส่งออกแร่ทองแดงอันดับต้นๆ ของโลก อย่างประเทศ ชีลี เปรู และออสเตรเลีย และประเทศมหาอำนาจที่มีผลต่อเศรษฐกิจโลกอย่างประเทศสหรัฐอเมริกา และสาธรณรัฐประชาชนจีน และค่าข้อมูลรายวันของ อัตราแลกเปลี่ยนค่าเงินปอนด์อังกฤษต่อดอลลาร์สหรัฐ และราคาปิดรายวันของโลหะทองคำ อันเนื่องมาจากราคาโลหะทองแดงมาจาก London Metal Exchange และทองคำเป็นโลหะที่มีมูลค่าทางเศรษฐกิจเป็นอย่างมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	A	B	C	D	E	F	G
1	Date	Price	Open	High	Low	Vol.	Change %
2	12/31/2024	8,768.00	8,840.00	8,830.00	8,821.50	33.80K	-1.59%
3	12/30/2024	8,910.00	8,955.00	8,946.00	8,950.00	28.11K	-0.80%
4	12/27/2024	8,982.00	8,968.00	8,957.50	8,958.00	25.08K	0.35%
5	12/24/2024	8,950.50	8,965.00	8,960.00	8,962.00	21.38K	0.43%
6	12/23/2024	8,912.00	8,950.00	8,949.50	8,950.00	26.92K	-0.32%
7	12/20/2024	8,941.00	8,910.00	8,905.00	8,905.00	36.34K	0.65%
8	12/19/2024	8,883.00	8,920.00	8,913.00	8,915.00	63.60K	-1.61%
9	12/18/2024	9,028.50	9,025.00	9,017.00	9,020.00	34.45K	0.43%
10	12/17/2024	8,989.50	8,980.00	8,979.00	8,979.00	42.29K	-0.81%
11	12/16/2024	9,062.50	9,055.00	9,050.00	9,050.00	39.26K	0.11%
12	12/13/2024	9,052.50	9,110.00	9,098.00	9,100.00	43.96K	-0.43%
13	12/12/2024	9,091.50	9,160.00	9,152.00	9,157.00	57.21K	-1.09%
14	12/11/2024	9,192.00	9,205.00	9,200.00	9,200.00	56.86K	-0.27%
15	12/10/2024	9,216.50	9,195.00	9,185.00	9,181.00	35.52K	-0.17%
16	12/9/2024	9,232.00	9,215.00	9,215.00	9,215.00	51.05K	1.20%
17	12/6/2024	9,122.50	9,155.00	9,147.00	9,149.00	46.13K	0.53%

รูปที่ 3.2 แสดงตัวอย่างข้อมูลชุดข้อมูลราคาทองแดงที่รวบรวม

	A	B	C	D	E	F
1	Year	Chile (USD Billions)	Peru (USD Billions)	Australia (USD Billions)	USA (USD Billions)	China (USD Billions)
2	2010	254.77	155.93	1,142.84	14,964.38	6,087.16
3	2011	276.31	171.78	1,389.98	15,599.73	7,551.50
4	2012	267.12	181.76	1,538.49	16,253.97	8,532.23
5	2013	278.27	192.75	1,565.37	16,843.19	9,570.41
6	2014	260.06	202.52	1,489.96	17,550.68	10,430.71
7	2015	242.51	190.22	1,349.09	18,206.03	11,061.55
8	2016	257.86	197.78	1,250.41	18,695.11	11,233.28
9	2017	277.29	212.17	1,326.75	19,495.39	12,143.50
10	2018	297.77	222.44	1,428.25	20,527.14	13,608.15
11	2019	282.31	227.24	1,392.68	21,374.37	14,292.28
12	2020	254.04	206.31	1,330.36	21,323.80	14,688.16
13	2021	315.52	222.08	1,559.27	23,594.76	17,820.23
14	2022	302.12	236.28	1,693.15	25,744.26	17,882.30
15	2023	335.53	246.4	1,724.00	27721	17,795.06
16	2024	343.82	303.29	1880	30510	19230

รูปที่ 3.3 แสดงข้อมูลรายปีของค่าผลิตภัณฑ์มวลรวมของประเทศที่คาดว่ามีผลต่อราคาทองแดง

3.4 การจัดเตรียมข้อมูล

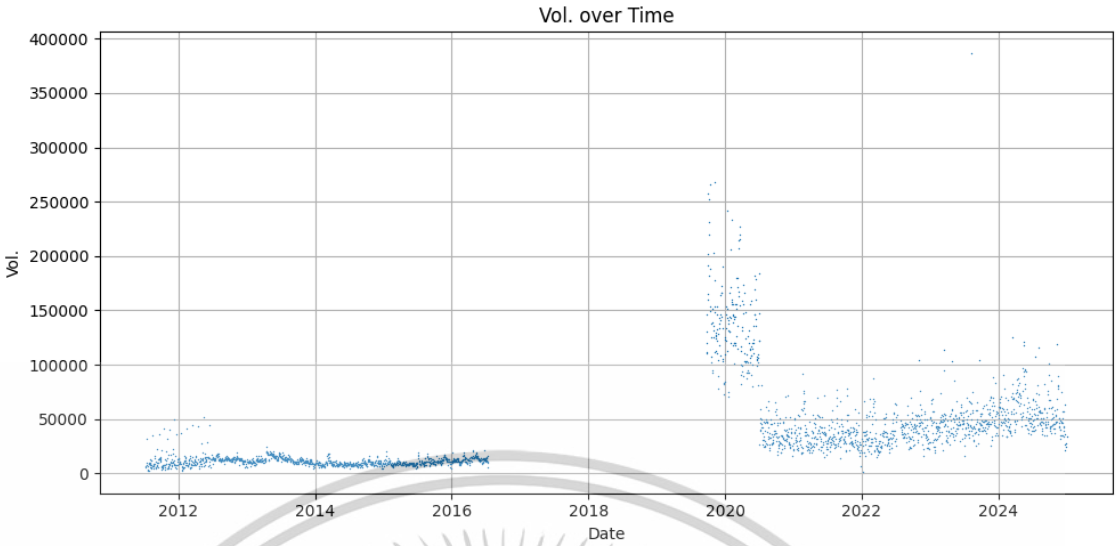
การวิจัยนี้ได้มีการจัดเตรียมข้อมูล เพื่อเพิ่มประสิทธิภาพในการฝึกสอนแบบจำลอง ให้สามารถทำนายผลลัพธ์ได้อย่างแม่นยำ โดยขั้นตอนดังนี้

3.4.1 การตรวจสอบและทำความสะอาดข้อมูล

การตรวจสอบและทำความสะอาดข้อมูลสำหรับงานวิจัยนี้ทำตามขั้นตอนดังต่อไปนี้

1) ตรวจสอบคุณภาพของข้อมูล เพื่อตรวจสอบโครงสร้าง ชนิดข้อมูล และค่าที่ขาดหาย
2) แปลงคอลัมน์วันที่ (Date) ที่อยู่ในรูปแบบข้อความ เป็นชนิด datetime เพื่อให้ง่ายต่อการวิเคราะห์ในรูปแบบอนุกรมเวลา
3) แปลงค่าตัวเลขที่อยู่ในรูปแบบข้อความ ให้เป็น float
4) ตรวจสอบค่าว่างในชุดข้อมูลพบ ตัวแปรปริมาณการซื้อขายของโลหะทองแดง มีค่าว่าง 1169 จากทั้งหมด 3729 จึงตัดสินใจลบตัวแปรนี้ออกจากชุดข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



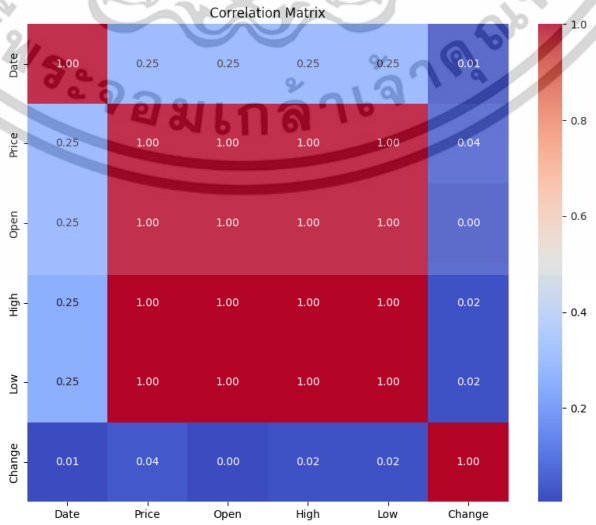
รูปที่ 3.4 แสดงกราฟการกระจายตัวของตัวแปรปริมาณการซื้อขายของโลหะทองแดง

3.4.2 วิศวกรรมคุณลักษณะ

ในการจัดเตรียมข้อมูลให้มีคุณภาพสำหรับการทำแบบจำลอง ได้มีการใช้หลากหลายวิธีการดังที่กล่าวไว้เบื้องต้นในหัวข้อที่ 2.10 เข้ามาช่วยในการจัดทำชุดข้อมูลดังนี้

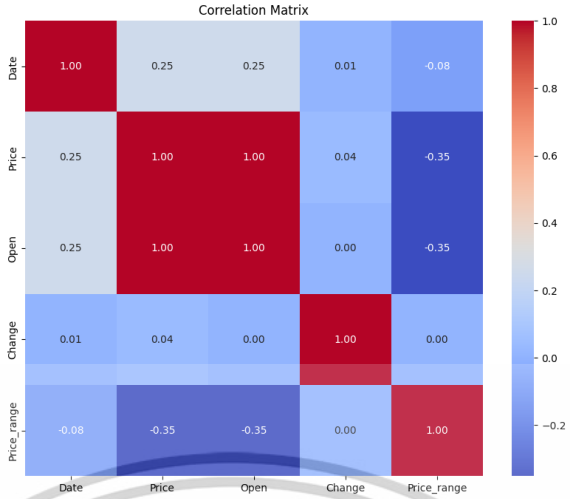
1) การสร้างคุณลักษณะใหม่จากภายในชุดข้อมูล

เมื่อพิจารณาค่าสหสัมพันธ์จากแผนภาพความร้อนสหสัมพันธ์ในชุดข้อมูลโลหะทองแดงพบว่าตัวแปร ราคาเปิดรายวันของโลหะทองแดง ราคาสูงสุดของโลหะทองแดง ณ วันที่ทำการเก็บข้อมูล และราคาต่ำสุดของโลหะทองแดง ณ วันที่ทำการเก็บข้อมูล มีความสัมพันธ์กันสูง จึงได้สร้างแปรตัวใหม่ขึ้นมาคือ ความแตกต่างของราคาสูงสุด กับราคาต่ำสุดของโลหะทองแดง โดยเมื่อพิจารณาจากค่าค่าสหสัมพันธ์อีกครั้งพบว่าไม่มีความสัมพันธ์กับตัวแปรอื่น



รูปที่ 3.5 แสดงแผนภาพความร้อนสหสัมพันธ์ของชุดข้อมูลโลหะทองแดง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.6 แสดงแผนภาพความร้อนสหสัมพันธ์ของชุดข้อมูลโลหะทองแดงหลังจากสร้างตัวแปร ความแตกต่างของราคาสูงสุด กับราคาต่ำสุดของโลหะทองแดง

2) การสร้างคุณลักษณะใหม่จากการรวมชุดข้อมูลภายนอก

เนื่องจากชุดข้อมูลที่ใช้ในการวิจัยนี้มีการนำข้อมูลจากแหล่งภายนอกมาเป็นตัวแปรต้นร่วมด้วย จึงจำเป็นต้องดำเนินการรวมข้อมูลจากหลายแหล่งเข้าด้วยกัน โดยข้อมูลที่มีความถี่รายวันจะใช้ตัวแปรวันที่เป็นเงื่อนไขหลักในการรวมข้อมูล ส่วนข้อมูลที่มีความถี่รายปีจะต้องสร้างตัวแปร "ปี" ขึ้นมาเพื่อใช้เป็นเกณฑ์ในการจับคู่ข้อมูลร่วมกับข้อมูลหลัก เมื่อดำเนินการรวมข้อมูลเสร็จสิ้นแล้ว จึงทำการลบตัวแปร "ปี" ออกจากชุดข้อมูล เพื่อให้ข้อมูลมีความพร้อมสำหรับการวิเคราะห์ในขั้นตอนต่อไป

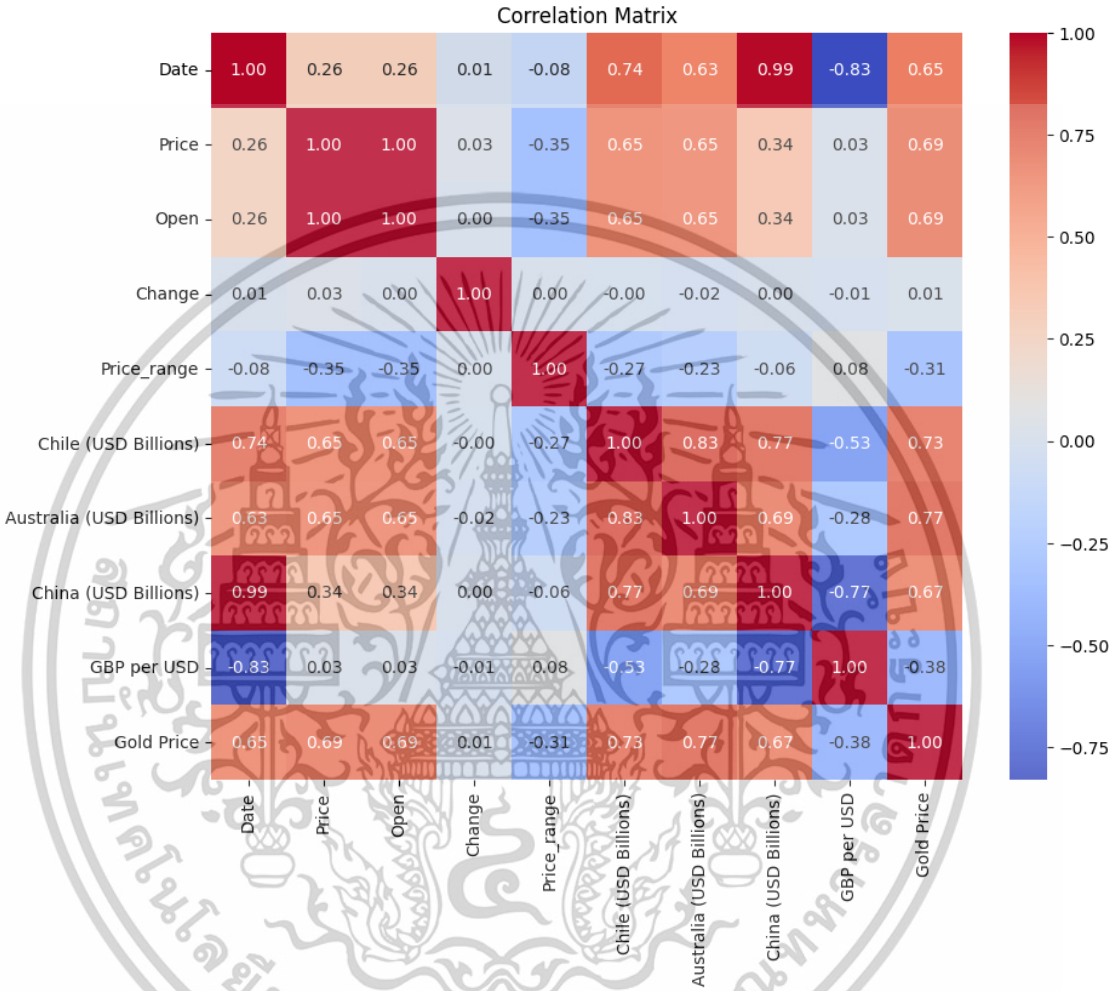
3) การคัดเลือกคุณลักษณะ

เพื่อให้ได้แบบจำลองที่มีประสิทธิภาพสูงสุดและหลีกเลี่ยงปัญหา multicollinearity หรือ overfitting ในการศึกษาครั้งนี้ ได้ดำเนินการคัดเลือกตัวแปรโดยพิจารณาจากมุมมองเชิงเศรษฐศาสตร์และผลการวิเคราะห์เชิงสถิติจากค่าสหสัมพันธ์จากแผนภาพความร้อนสหสัมพันธ์ และ ค่า Variance Inflation Factor (VIF) พบว่าค่าผลิตภัณฑ์มวลรวมในประเทศรายปีของประเทศชิลี และเปรู มีความสัมพันธ์กันสูงจึงได้ตัดค่าผลิตภัณฑ์มวลรวมในประเทศรายปีของประเทศเปรู เนื่องจากประเทศชิลีมีมูลค่าการส่งออกทองแดงเป็นอันดับหนึ่งของโลกในปัจจุบัน นอกจากนี้ยังพบว่าค่าผลิตภัณฑ์มวลรวมในประเทศรายปีของประเทศสหรัฐอเมริกาเมื่อพิจารณาจากค่า VIF พบว่ามีค่ามากกว่า 10 ซึ่งหมายถึงพบปัญหา multicollinearity ชุดข้อมูลที่เตรียมพร้อมสำหรับการทำแบบจำลองจะประกอบด้วยตัวแปร 10 คุณลักษณะ ได้แก่

1) วันที่ ที่ทำการเก็บข้อมูล
2) ราคาปิดรายวันของโลหะทองแดง
3) ราคาเปิดรายวันของโลหะทองแดง
4) อัตราการเปลี่ยนแปลงเมื่อเทียบกับวันก่อนหน้า
5) ความแตกต่างของราคาสูงสุด กับราคาต่ำสุด
6) ค่าผลิตภัณฑ์มวลรวมในประเทศรายปี ของประเทศชิลี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7) ค่าผลิตภัณฑ์มวลรวมในประเทศรายปี ของประเทศออสเตรเลีย
8) ค่าผลิตภัณฑ์มวลรวมในประเทศรายปี ของประเทศจีน
9) อัตราแลกเปลี่ยนค่าเงินปอนด์อังกฤษต่อดอลลาร์สหรัฐ
10) ราคาปิดรายวันของโลหะทองคำ



รูปที่ 3.7 แสดงแผนภาพความร้อนสหสัมพันธ์ของชุดข้อมูลสำหรับการสร้างแบบจำลอง

โดยคุณลักษณะ วันที่ ที่ทำการเก็บข้อมูล คือดัชนีของชุดข้อมูล และ ราคาปิดรายวันของโลหะทองแดง คือตัวแปรตาม หรือตัวแปรเป้าหมาย โดยพิจารณาคุณภาพของข้อมูลจากค่า Variance Inflation Factor จะพบว่าชุดข้อมูลนี้แต่ละตัวแปรต้นมีค่าไม่เกิน 10 ดังตารางที่ 3.1 แสดงถึงชุดข้อมูลนี้ไม่มีปัญหา multicollinearity หรือ มีปัญหา multicollinearity ที่น้อย

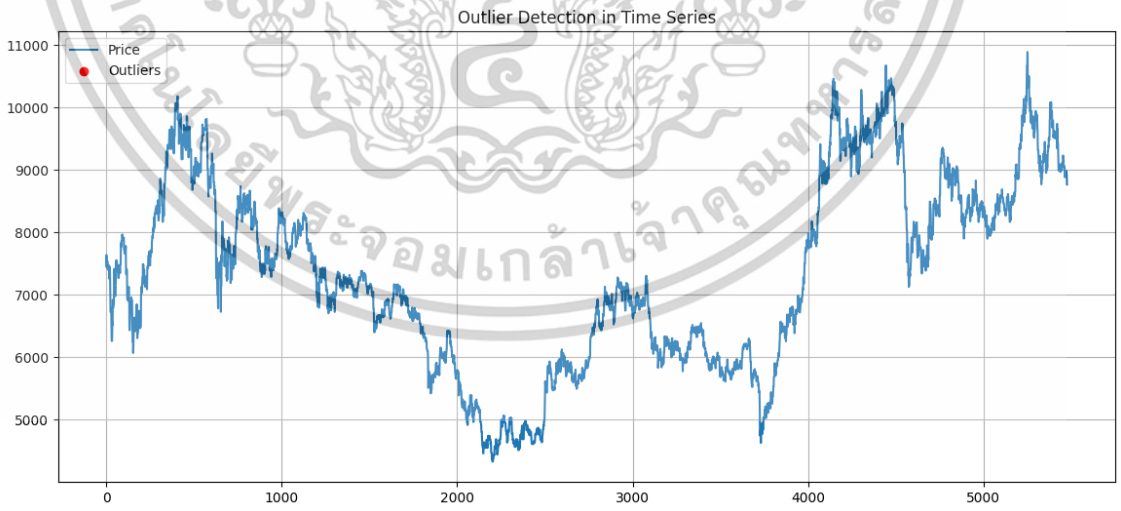
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.1 แสดงค่า Variance Inflation Factor กับ คุณลักษณะ

คุณลักษณะ	Variance Inflation Factor (VIF)
ราคาเปิดรายวันของโลหะทองแดง	3.698092
อัตราการเปลี่ยนแปลงเมื่อเทียบกับวันก่อนหน้า	1.002226
ความแตกต่างของราคาสูงสุด กับราคาต่ำสุดของโลหะทองแดง	1.289904
ค่าผลิตภัณฑ์มวลรวมในประเทศรายปี ของประเทศชิลี	6.620013
ค่าผลิตภัณฑ์มวลรวมในประเทศรายปี ของประเทศออสเตรเลีย	5.707418
ค่าผลิตภัณฑ์มวลรวมในประเทศรายปี ของประเทศจีน	7.094062
อัตราแลกเปลี่ยนค่าเงินปอนด์อังกฤษต่อดอลลาร์สหรัฐ	5.707418
ราคาปิดรายวันของโลหะทองคำ	3.866263

4) การตรวจสอบค่าสุดโต่ง

การจัดเตรียมข้อมูลนี้ได้มีการตรวจสอบค่าสุดโต่งด้วยวิธี IQR ดังที่กล่าวไว้ในหัวข้อที่ 2.9.1 พบว่าไม่มีค่าสุดโต่งแสดงออกมาดังรูปที่ 3.7



รูปที่ 3.8 แสดงการตรวจสอบค่าสุดโต่งในชุดข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5) การจัดการค่าว่าง

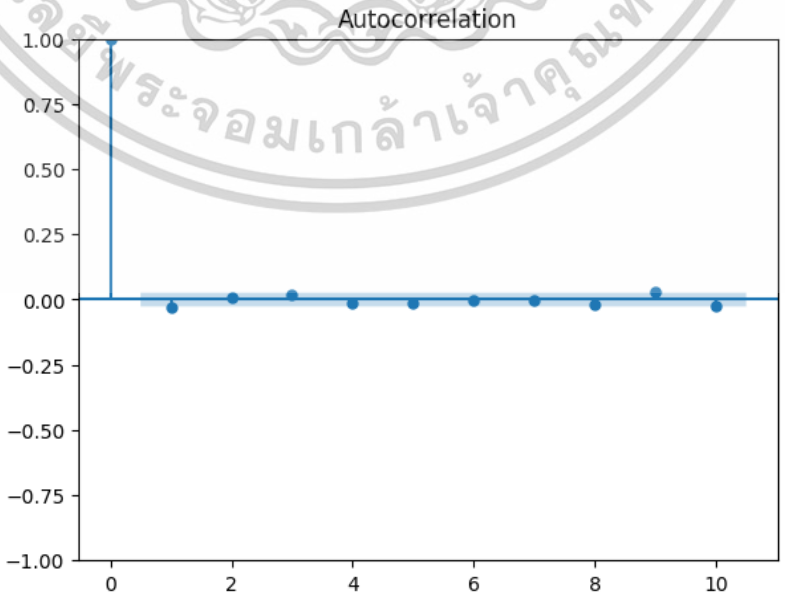
เนื่องจากชุดข้อมูลราคาทองแดงที่ใช้ในการวิจัยไม่สามารถบันทึกข้อมูลได้ครบทุกวัน ซึ่งอาจเกิดจากหลายปัจจัย เช่น ไม่มีการซื้อขายในวันหยุดราชการหรือวันหยุดสุดสัปดาห์ เป็นต้น ส่งผลให้เกิดช่องว่างของข้อมูลในลำดับเวลา เพื่อให้ได้ชุดข้อมูลที่เป็นอนุกรมเวลาที่สมบูรณ์ จึงจำเป็นต้องดำเนินการเติมค่าข้อมูลในวันที่หายไป โดยใช้วิธีเติมค่าด้วยข้อมูลจากวันก่อนหน้า (forward fill) เพื่อรักษาความต่อเนื่องของข้อมูลสำหรับการวิเคราะห์ในขั้นตอนถัดไป

3.5 การพัฒนาแบบจำลอง

การวิจัยนี้ใช้แบบจำลองทั้งหมด 5 แบบ ซึ่งสามารถแยกออกได้เป็น 2 กลุ่มได้แก่

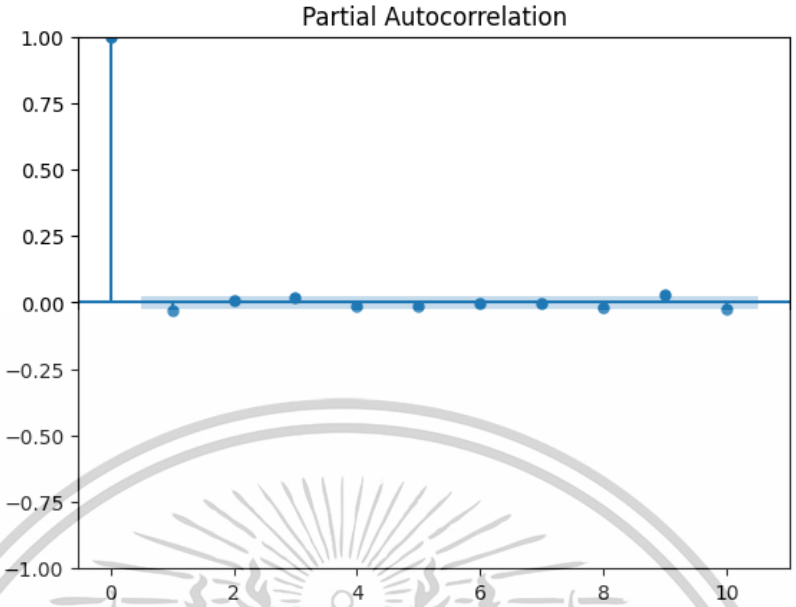
3.5.1 แบบจำลองทางสถิติ

1) ARIMA (AutoRegressive Integrated Moving Average)
   1) ตรวจสอบและปรับให้ชุดข้อมูลเป้าหมายให้มีความนิ่งโดยใช้ Augmented Dickey-Fuller (ADF) Test พบว่าค่า p-values มากกว่า 0.05 ที่ 0.3336 แสดงว่าชุดข้อมูลไม่มีความนิ่ง และเมื่อทำ differencing 1 ครั้ง และตรวจสอบค่า p-values พบว่ามีค่าน้อยกว่า 0.05 ที่ 0 แสดงว่าค่า d สำหรับชุดข้อมูลนี้มีค่าเท่ากับ 1
   2) พิจารณากราฟ ACF และ PACF เพื่อกำหนดค่า p และ q พบว่าค่าเข้าใกล้ 0 อย่างรวดเร็วที่ข้อมูลล่าช้าที่ 1 ทั้งกราฟ ACF และ PACF หมายความว่าไม่มี partial autocorrelation อย่างชัดเจน และไม่มี moving average เกิดขึ้น แสดงค่า p และ q ของชุดข้อมูลนี้มีค่าเท่ากับ 0
   3) สร้างและฝึกแบบจำลองด้วยชุดข้อมูลฝึกสอนโดยแบ่งเป็นอัตราส่วน 80 จากชุดข้อมูลทั้งหมด
   4) ทำนายข้อมูลชุดทดสอบด้วยอัตราส่วน 20 จากชุดข้อมูลทั้งหมด และประเมินประสิทธิภาพของแบบจำลอง



รูปที่ 3.9 แสดงกราฟ ACF

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 แสดงกราฟ PACF

3.5.2 แบบจำลองการเรียนรู้ของเครื่อง

1) แบบจำลองต้นไม้ตัดสินใจ
    1) เตรียมชุดข้อมูล โดยกำหนดดัชนีของชุดข้อมูลด้วยวันที่ จากนั้นแยกตัวแปรเป้าหมายและตัวแปรต้น
    2) สร้างตัวแปรทางเวลาเพิ่มเติมโดยกำหนดเป็น 1 วัน 7 (1 สัปดาห์) วัน 15 (ครึ่งเดือน) วัน 30 วัน (1 เดือน) 120 วัน (1 ไตรมาส) 180 วัน (6 เดือนหรือครึ่งปี) และ 365 วัน (1 ปี)
    3) สร้างและฝึกแบบจำลองด้วยชุดข้อมูลฝึกสอนโดยแบ่งเป็นอัตราส่วน 80 จากชุดข้อมูลทั้งหมด และกำหนดตัวแปรดังตารางที่ 3.2
    4) ทำนายข้อมูลชุดทดสอบด้วยอัตราส่วน 20 จากชุดข้อมูลทั้งหมด และประเมินประสิทธิภาพของแบบจำลอง
2) แบบจำลอง XGBoost
    1) เตรียมชุดข้อมูล โดยกำหนดดัชนีของชุดข้อมูลด้วยวันที่ จากนั้นแยกตัวแปรเป้าหมายและตัวแปรต้น
    2) สร้างตัวแปรทางเวลาเพิ่มเติมโดยกำหนดเป็น 1 วัน 7 (1 สัปดาห์) วัน 15 (ครึ่งเดือน) วัน 30 วัน (1 เดือน) 120 วัน (1 ไตรมาส) 180 วัน (6 เดือนหรือครึ่งปี) และ 365 วัน (1 ปี)
    3) สร้างและฝึกแบบจำลองด้วยชุดข้อมูลฝึกสอนโดยแบ่งเป็นอัตราส่วน 80 จากชุดข้อมูลทั้งหมด และกำหนดตัวแปรดังตารางที่ 3.2
    4) ทำนายข้อมูลชุดทดสอบด้วยอัตราส่วน 20 จากชุดข้อมูลทั้งหมด และประเมินประสิทธิภาพของแบบจำลอง
3) แบบจำลองป่าสุ่ม
    1) เตรียมชุดข้อมูล โดยกำหนดดัชนีของชุดข้อมูลด้วยวันที่ จากนั้นแยกตัวแปรเป้าหมายและตัวแปรต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) สร้างตัวแปรทางเวลาเพิ่มเติมโดยกำหนดเป็น 1 วัน 7 (1 สัปดาห์) วัน 15 (ครึ่งเดือน) วัน 30 วัน (1 เดือน) 120 วัน (1 ไตรมาส) 180 วัน (6 เดือนหรือครึ่งปี) และ 365 วัน (1 ปี)
3) สร้างและฝึกแบบจำลองด้วยชุดข้อมูลฝึกสอนโดยแบ่งเป็นอัตราส่วน 80 จากชุดข้อมูลทั้งทั้งหมด และกำหนดตัวแปรดังตารางที่ 3.2
4) ทำนายข้อมูลชุดทดสอบด้วยอัตราส่วน 20 จากชุดข้อมูลทั้งหมด และประเมินประสิทธิภาพของแบบจำลอง

4) แบบจำลองการเรียนรู้เชิงลึก LSTM (Long Short-Term Memory)
1) เตรียมชุดข้อมูล โดยกำหนดดัชนีของชุดข้อมูลด้วยวันที่ จากนั้นแยกตัวแปรเป้าหมายและตัวแปรต้น
2) ทำการปรับขนาดข้อมูลด้วย MinMaxScaler
3) สร้างลำดับเวลาสำหรับแบบจำลอง
4) สร้างเลเยอร์ LSTM โดยกำหนดตัวแปรดังตารางที่ 3.2 และ Dense สำหรับโครงข่ายประสาทเทียม
5) สร้างและฝึกแบบจำลองด้วยชุดข้อมูลฝึกสอนโดยแบ่งเป็นอัตราส่วน 80 จากชุดข้อมูลทั้งหมด
6) ทำนายข้อมูลชุดทดสอบด้วยอัตราส่วน 20 จากชุดข้อมูลทั้งหมด และประเมินประสิทธิภาพของแบบจำลอง

ตารางที่ 3.2 แสดงตัวแปรในการตั้งค่าแบบจำลอง

แบบจำลอง	**ตัวแปรที่กำหนด**
แบบจำลอง ARIMA	(p=0,d=1,q=0)
แบบจำลองต้นไม้ตัดสินใจ	max_depth=5, min_samples_split=10, min_samples_leaf=5, random_state=42
แบบจำลอง XGBoost	n_estimators=100, learning_rate=0.05, max_depth=6, subsample=0.8, colsample_bytree=0.8, random_state=42
แบบจำลองป่าสุ่ม	n_estimators=100, max_depth=5, min_samples_leaf=5, random_state=42
แบบจำลองการเรียนรู้เชิงลึก LSTM	units=128, activation='tanh', return_sequences=False, input_shape=(TIME_STEPS, X.shape[1]))

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6 การวัดประสิทธิภาพแบบจำลอง

งานวิจัยนี้ได้ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองแต่ละตัวด้วยตัวชี้วัดสามวิธีได้แก่

1) ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) คือค่าเฉลี่ยของ ค่าความคลาดเคลื่อน (Error) แบบสัมบูรณ์ระหว่างค่าที่คาดการณ์กับค่าจริง โดยมีหน่วยเดียวกันกับค่าที่ทำนาย โดยยิ่งค่าเข้าใกล้ศูนย์มากเท่าใด ยิ่งแสดงว่าแบบจำลองสามารถทำนายได้แม่นยำมากขึ้น
2) ค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Root Mean Squared Error: RMSE) คือค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนยกกำลังสองระหว่างค่าที่คาดการณ์กับค่าจริง โดยมีหน่วยเดียวกับค่าที่ทำนาย ค่านี้ให้ความสำคัญกับความคลาดเคลื่อนที่สูงเป็นพิเศษ ยิ่งค่า เข้าใกล้ศูนย์มากเท่าใด ยิ่งแสดงว่าแบบจำลองมีความแม่นยำในการทำนายมากขึ้น
3) ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) คือค่าเฉลี่ยของค่าความคลาดเคลื่อนสัมบูรณ์ในรูปแบบเปอร์เซ็นต์เมื่อเทียบกับค่าจริง ยิ่งค่าต่ำเท่าใด ยิ่งแสดงว่าแบบจำลองสามารถทำนายได้แม่นยำ โดยเหมาะสำหรับการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลองในข้อมูลที่มีหน่วยต่างกัน

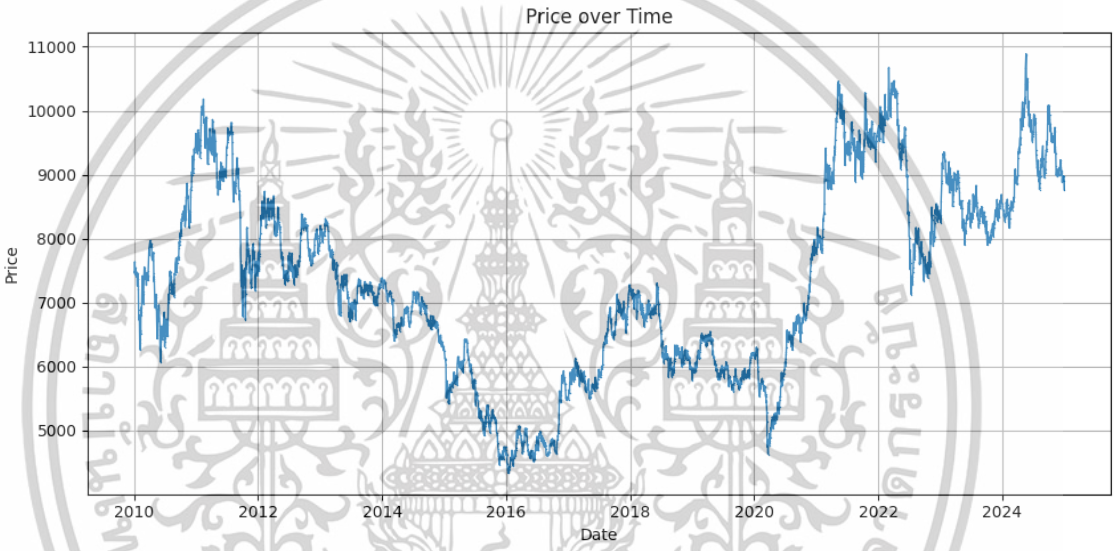
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการวิจัยและการอภิปรายผล

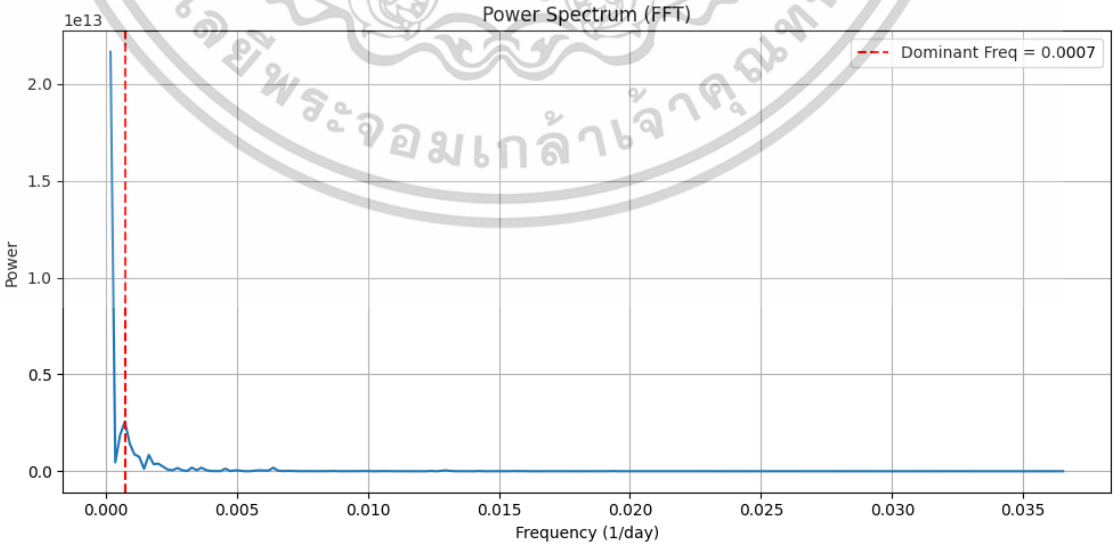
ในบทนี้จะกล่าวถึงผลการดำเนินงานจากการสร้างแบบจำลองการเรียนรู้ของเครื่องจักรในการการพยากรณ์ราคาทองแดง ตามวิธีดำเนินงานวิจัยที่แสดงไว้ในบทที่ 3

4.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น

เมื่อพิจารณาจากรูปที่ 4.1 จะเห็นว่าไม่สามารถระบุลักษณะฤดูกาล หรือวัฏจักร ของอนุกรมเวลาได้อย่างชัดเจน แต่สามารถสังเกตเห็นแนวโน้มในบางช่วงของเวลา



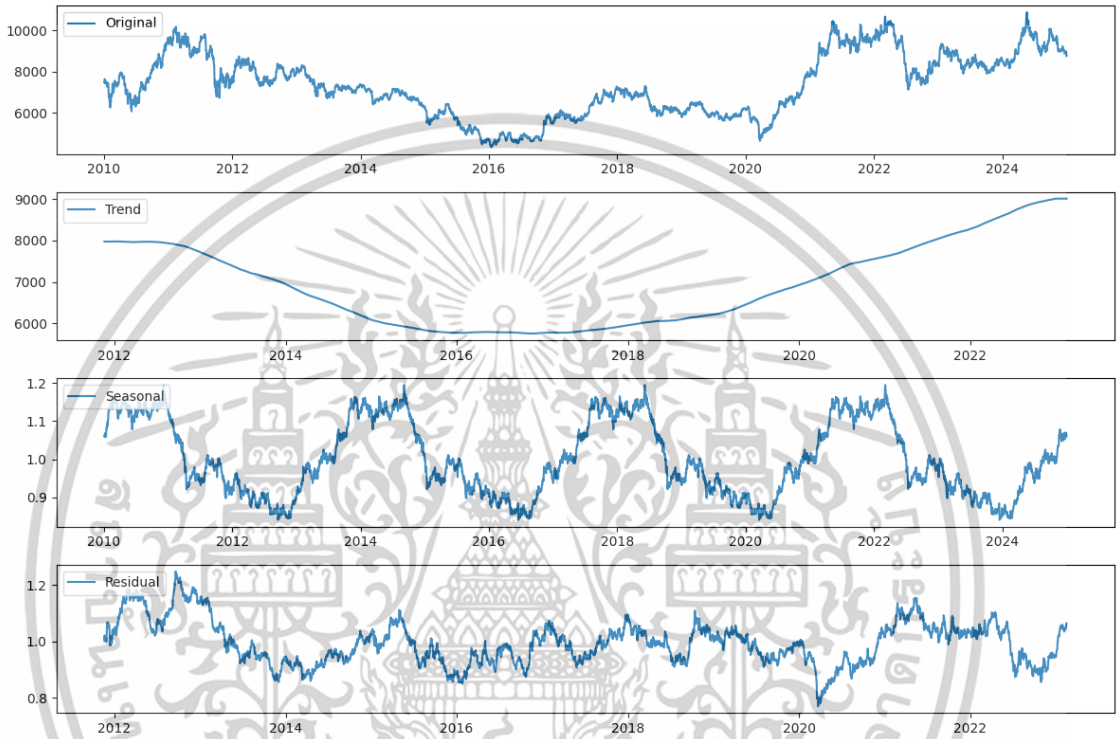
รูปที่ 4.1 แสดงกราฟราคาปิดโลหะทองแดงรายวัน



รูปที่ 4.2 แสดงกราฟความถี่จาก Fast Fourier transform

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อไม่สามารถพิจารณาได้ด้วยสายตา จึงนำกระบวนการทางคณิตศาสตร์อย่าง Fast Fourier transform มาวิเคราะห์เพื่อค้นหาวัฏจักรที่โดดเด่นที่สุดสำหรับชุดข้อมูลอนุกรมเวลาที่มีความผันผวนสูง ดังรูปที่ 4.2 จะเห็นได้ว่าความถี่ที่ได้จากกระบวนการ Fast Fourier transform มีค่าสูงสุดที่ 0.0007 แสดงว่าค่าความถี่นี้เด่นชัดมากที่สุดในชุดข้อมูลนี้ หรือเมื่อแปลงเป็นคาบเวลาจะได้ที่ 1369 วัน หรือประมาณ 2 ปี 9 เดือน และเมื่อกำจัดค่าความแปรปรวนสุ่ม โดยการแยกส่วนประกอบของอนุกรมเวลา จะเห็นได้ว่าชุดข้อมูลมีวัฏจักรที่เด่นชัดที่เด่นชัดมากยิ่งขึ้นดังรูปที่ 4.3



รูปที่ 4.3 แสดงกราฟข้อมูลหลังจากแยกส่วนประกอบอนุกรมเวลา

4.2 ผลการพัฒนาแบบจำลอง

การวิจัยนี้ได้ดำเนินการพัฒนาแบบจำลองพยากรณ์ราคาทองแดงจำนวน 5 แบบ ได้แก่ แบบจำลอง ARIMA, แบบจำลองต้นไม้ตัดสินใจ, แบบจำลอง XGBoost, แบบจำลองป่าสุ่ม และแบบจำลองการเรียนรู้เชิงลึก LSTM โดยใช้ข้อมูลที่ได้จากการรวบรวมข้อมูลดังที่แสดงในหัวข้อที่ 3.3 และจัดเตรียมข้อมูลดังที่แสดงในหัวข้อที่ 3.4

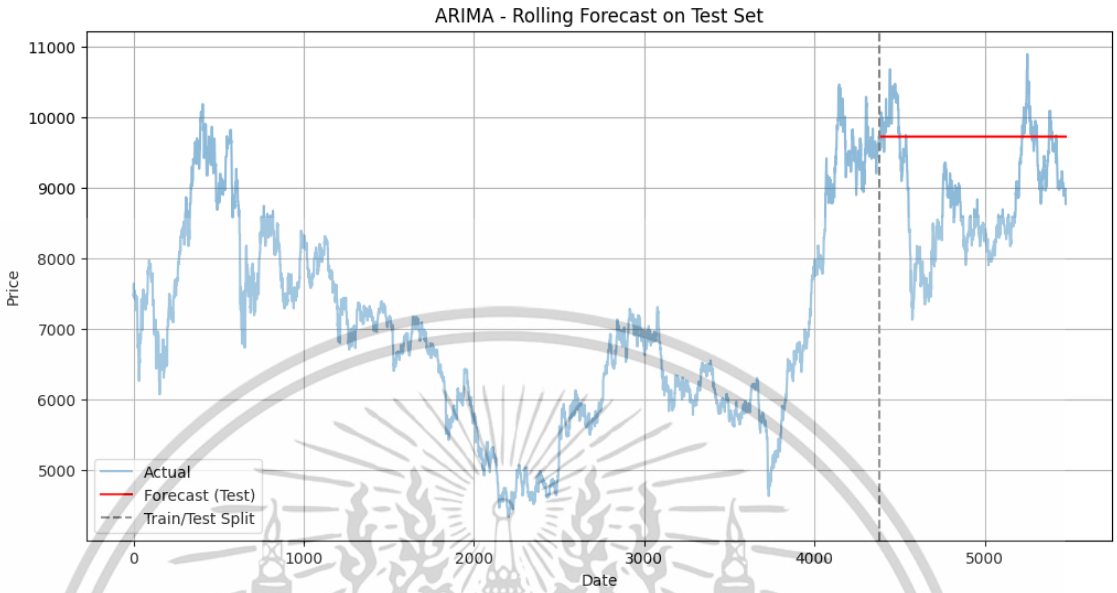
4.2.1 ผลการพัฒนาแบบจำลอง ARIMA

แบบจำลอง ARIMA ได้คัดเลือกตัวแปรด้วยหลักการทางสถิติดังที่แสดงไว้ในหัวข้อที่ 3.5.1 โดยได้กำหนดค่าตัวแปรเป็น (0,1,0)

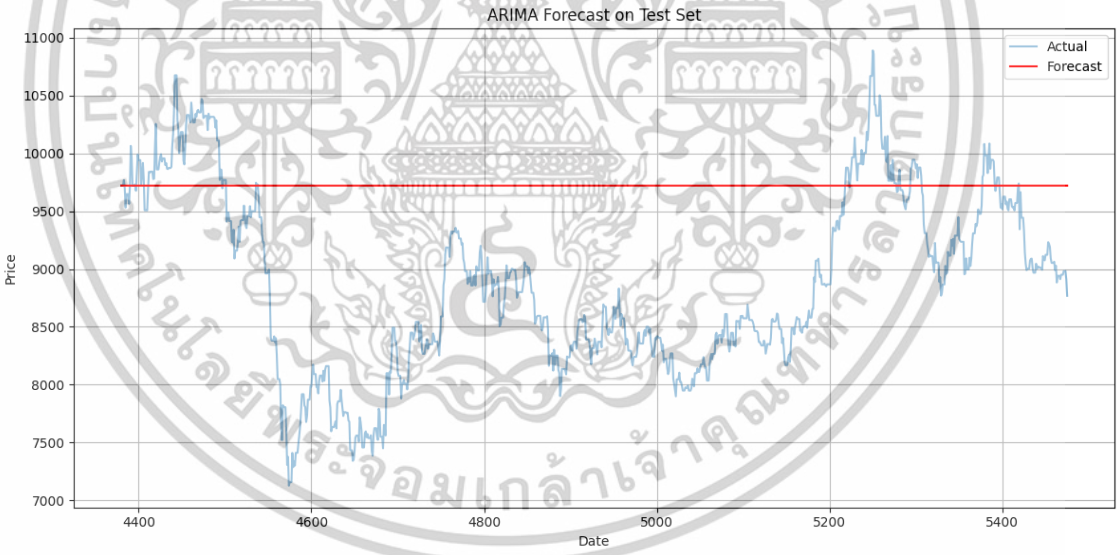
ผลการพยากรณ์ของ ARIMA พบว่าแบบจำลองไม่สามารถจับแนวโน้มของข้อมูลได้ มีความคลาดเคลื่อนสูงเมื่อพิจารณาจากดังรูปที่ 4.4 และ 4.5 จะเห็นได้ว่าค่าที่ได้จากการพยากรณ์จะมีรูปแบบนิ่ง และเมื่อพิจารณาจากตัวชี้วัดทั้งสามค่าจะเห็นได้ว่าค่า MAE คลาดเคลื่อน 969.6957 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 1,149.8894 ดอลลาร์สหรัฐต่อตัน และค่า MAPE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คลาดเคลื่อน 11.59% เมื่อพิจารณาประกอบกันแล้วแบบจำลอง ARIMA ไม่เหมาะสมสำหรับการพยากรณ์ข้อมูลที่มีความแปรปรวนสุ่มมาก



รูปที่ 4.4 แสดงกราฟผลการพยากรณ์ของแบบจำลอง ARIMA เทียบกับข้อมูลจริง



รูปที่ 4.5 แสดงกราฟผลการพยากรณ์ของแบบจำลอง ARIMA เทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ

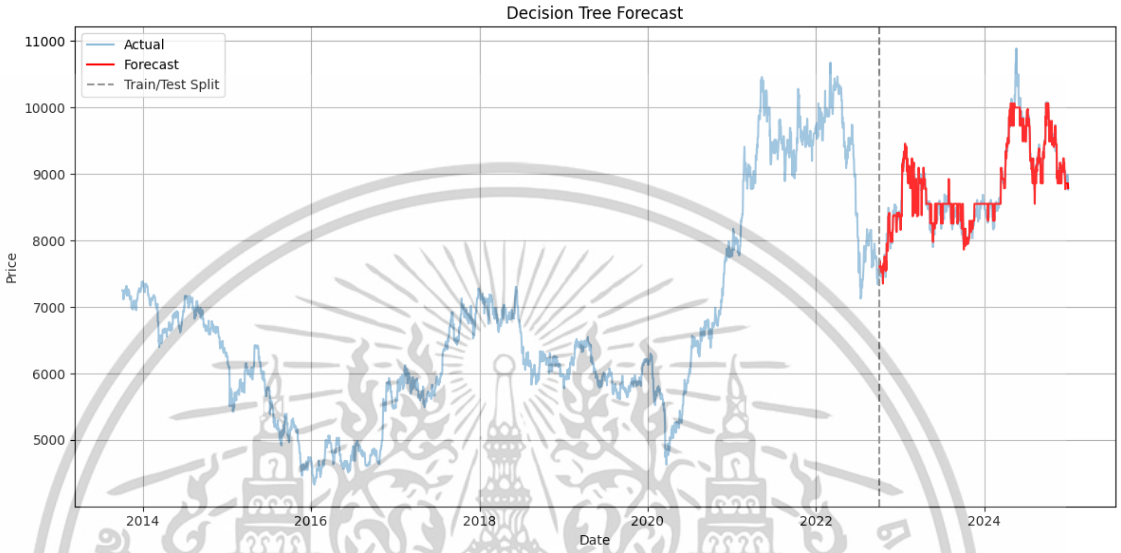
4.2.2 ผลการพัฒนาแบบจำลองต้นไม้ตัดสินใจ

แบบจำลองต้นไม้ตัดสินใจเป็นแบบจำลองที่ไม่สามารถเข้าใจในลำดับเวลาได้ จึงได้มีการสร้างค่าล่าช้าเพื่อให้แบบจำลองสามารถเข้าใจลำดับเวลาได้ โดยกำหนดทั้งหมดดังนี้ 1 วัน, 7 (1 สัปดาห์), วัน 15 (ครึ่งเดือน), วัน 30 วัน (1 เดือน), 120 วัน (1 ไตรมาส), 180 วัน (6 เดือนหรือครึ่งปี) และ 365 วัน (1 ปี)

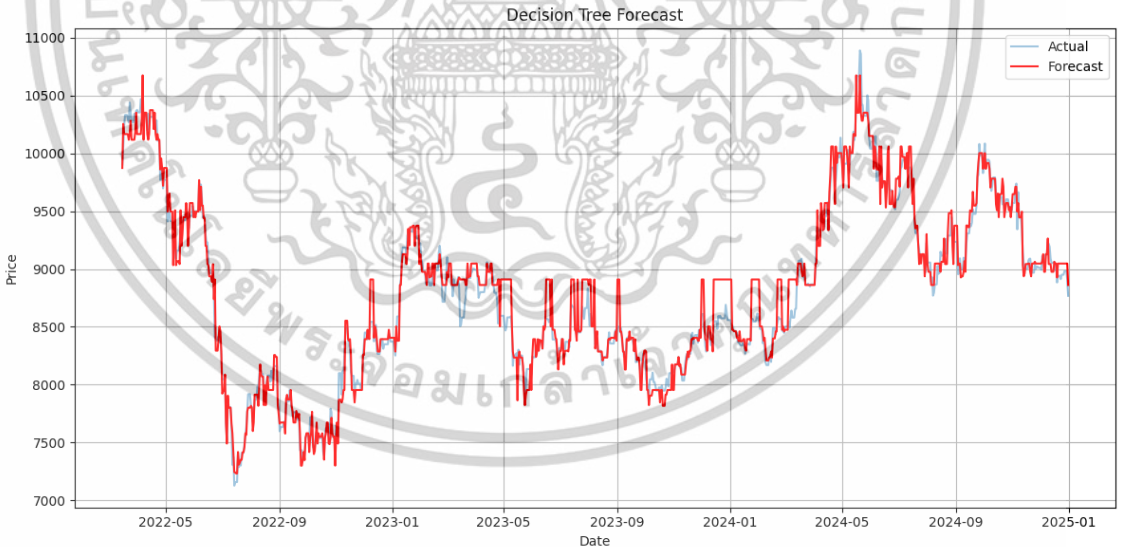
ผลการพยากรณ์ของแบบจำลองต้นไม้ตัดสินใจ พบว่าแบบจำลองสามารถจับแนวโน้มของข้อมูลได้ในระดับนึง เมื่อพิจารณาจากดังรูปที่ 4.6 และ 4.7 จะเห็นได้ว่าค่าที่ได้จากการพยากรณ์จะมี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แนวโน้มไปในทิศทางเดียวกันกับข้อมูลจริงที่ใช้ทดสอบแต่ยังไม่ราบลื่นมากนัก และเมื่อพิจารณาจากตัวชี้วัดทั้งสามค่าจะเห็นได้ว่าค่า MAE คลาดเคลื่อน 97.2573 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 137.9294 ดอลลาร์สหรัฐต่อตัน และค่า MAPE คลาดเคลื่อน 1.10% เมื่อพิจารณาประกอบกันแล้วถึงแม้จะมีค่าความคลาดเคลื่อนที่ต่ำแต่ในบางช่วงมีการทำนายที่ผิดพลาด อาจจำเป็นต้องนำไปพัฒนาต่อเพื่อให้ได้ผลที่น่าพอใจกว่านี้ก่อนจะนำไปใช้งานจริง



รูปที่ 4.6 แสดง กราฟผลการพยากรณ์ของแบบจำลองต้นไม้ตัดสินใจเทียบกับข้อมูลจริง



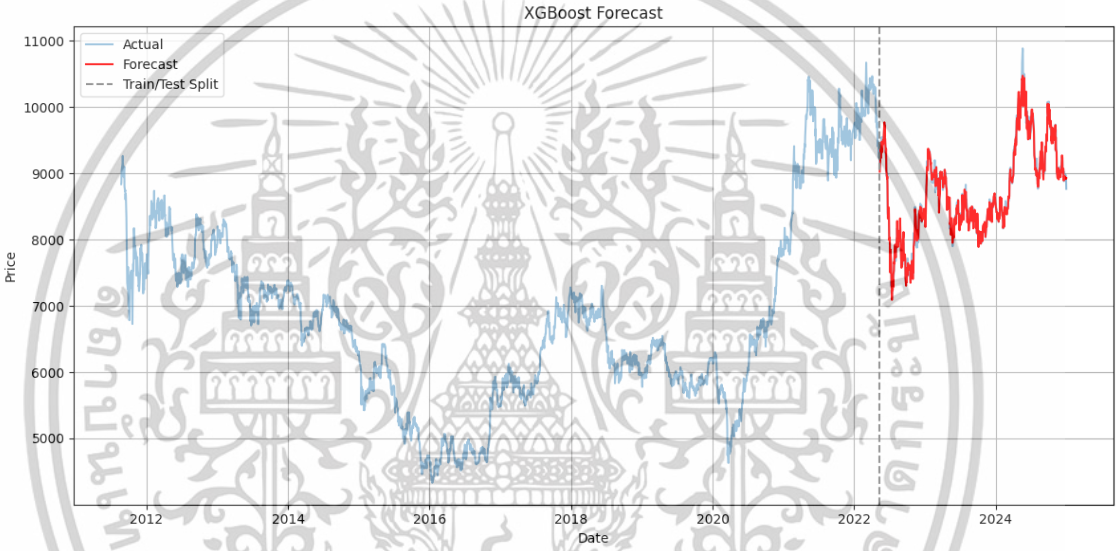
รูปที่ 4.7 แสดงกราฟผลการพยากรณ์ของแบบจำลองต้นไม้ตัดสินใจเทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

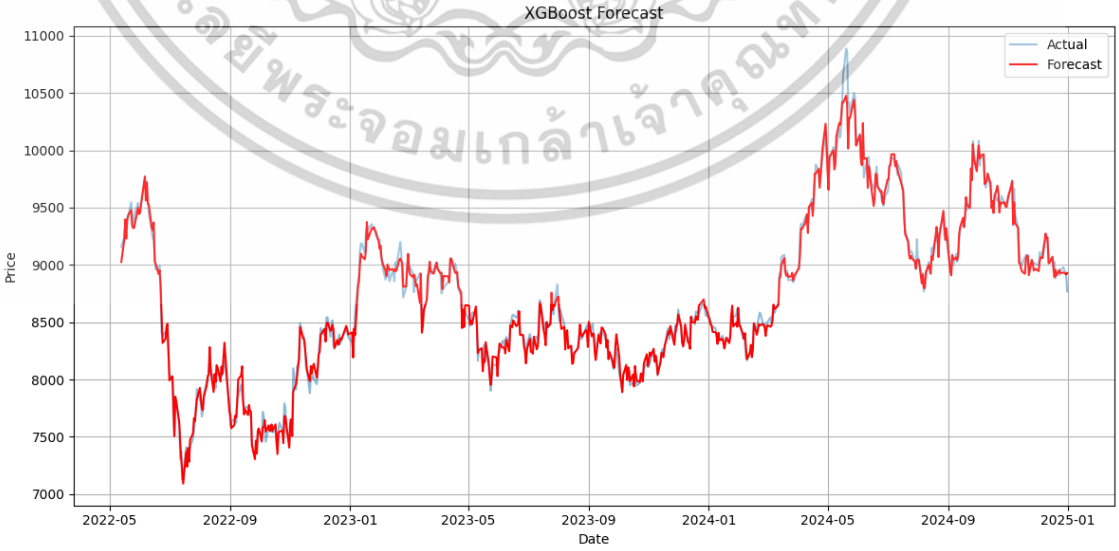
4.2.3 ผลการพัฒนาแบบจำลอง XGBoost

แบบจำลอง XGBoost เป็นแบบจำลองที่ไม่สามารถเข้าใจในลำดับเวลาได้เนื่องจากมีพื้นฐานมาจากแบบจำลองต้นไม้ตัดสินใจ จึงได้มีการสร้างค่าล่าช้าเพื่อให้แบบจำลองสามารถเข้าใจดำลับเวลาได้เช่นเดียวกับแบบจำลองต้นไม้ตัดสินใจ

ผลการพยากรณ์ของ XGBoost พบว่าแบบจำลองสามารถจับแนวโน้มของข้อมูลได้ เมื่อพิจารณาจากดังรูปที่ 4.8 และ 4.9 จะเห็นได้ว่าค่าที่ได้จากการพยากรณ์จะมีแนวโน้มไปในทิศทางเดียวกันกับข้อมูลจริงที่ใช้ทดสอบ และเมื่อพิจารณาจากตัวชี้วัดทั้งสามค่าจะเห็นได้ว่าค่า MAE คลาดเคลื่อน 52.3623 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 70.6940 ดอลลาร์สหรัฐต่อตัน และค่า MAPE คลาดเคลื่อน 0.59% เมื่อพิจารณาประกอบกันแล้วแบบจำลองนี้สามารถใช้ในการการพยากรณ์ราคาทองแดงได้ถึงแม้ว่าจะไม่สามารถทำนายค่าในช่วงกลางปี พ.ศ. 2567 ได้อย่างแม่นยำ



รูปที่ 4.8 แสดง กราฟผลการพยากรณ์ของแบบจำลอง XGBoost เทียบกับข้อมูลจริง



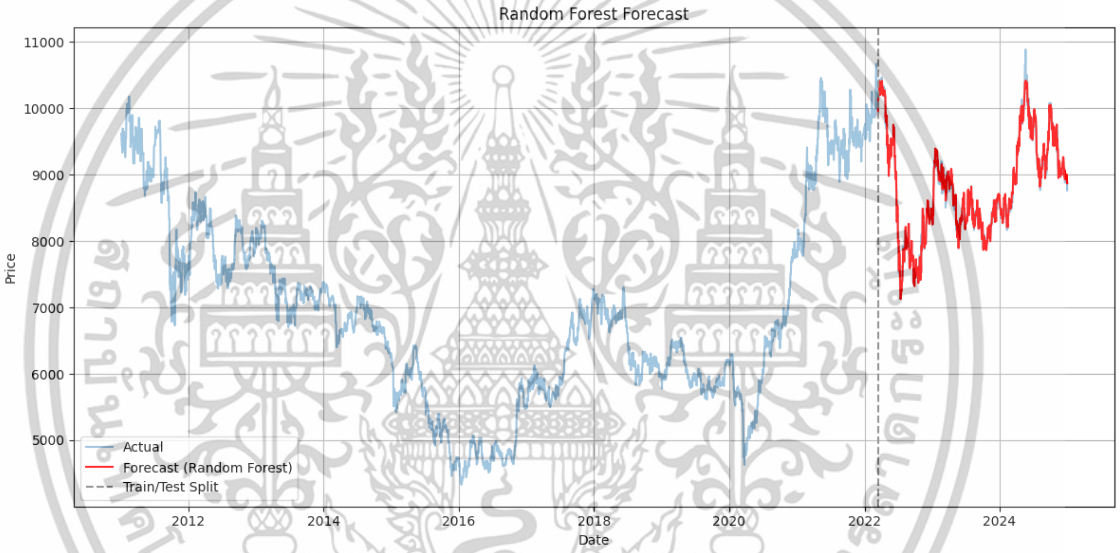
รูปที่ 4.9 แสดงกราฟผลการพยากรณ์ของแบบจำลอง XGBoost เทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

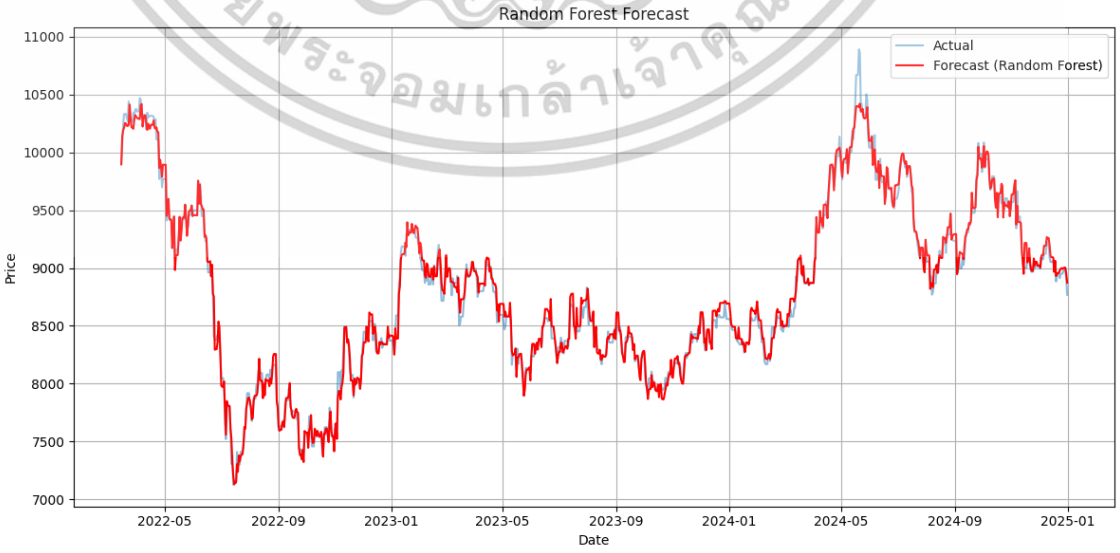
4.2.4 ผลการพัฒนาแบบจำลองป่าสุ่ม

แบบจำลองป่าสุ่มเป็นแบบจำลองที่ไม่สามารถเข้าใจในลำดับเวลาได้เนื่องจากมีพื้นฐานมาจากแบบจำลองต้นไม้ตัดสินใจเช่นเดียวกับแบบจำลอง XGBoost จึงได้มีการสร้างค่าล่าช้าเพื่อให้แบบจำลองสามารถเข้าใจดำลับเวลาได้เช่นเดียวกัน

ผลการพยากรณ์ของแบบจำลองป่าสุ่มพบว่าแบบจำลองสามารถจับแนวโน้มของข้อมูลได้ เมื่อพิจารณาจากดังรูปที่ 4.10 และ 4.11 จะเห็นได้ว่าค่าที่ได้จากการพยากรณ์จะมีแนวโน้มไปในทิศทางเดียวกันกับข้อมูลจริงที่ใช้ทดสอบ และเมื่อพิจารณาจากตัวชี้วัดทั้งสามค่าจะเห็นได้ว่าค่า MAE คลาดเคลื่อน 53.0240 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 71.3360 ดอลลาร์สหรัฐต่อตัน และค่า MAPE คลาดเคลื่อน 0.60% เมื่อพิจารณาประกอบกันแล้วแบบจำลองนี้สามารถใช้ในการการพยากรณ์ราคาทองแดงได้ถึงแม้ว่าจะไม่สามารถทำนายค่าในช่วงกลางปี พ.ศ. 2567 ได้อย่างแม่นยำเช่นเดียวกันกับแบบจำลอง XGBoost



รูปที่ 4.10 กราฟผลการพยากรณ์ของแบบจำลองป่าสุ่มเทียบกับข้อมูลจริง

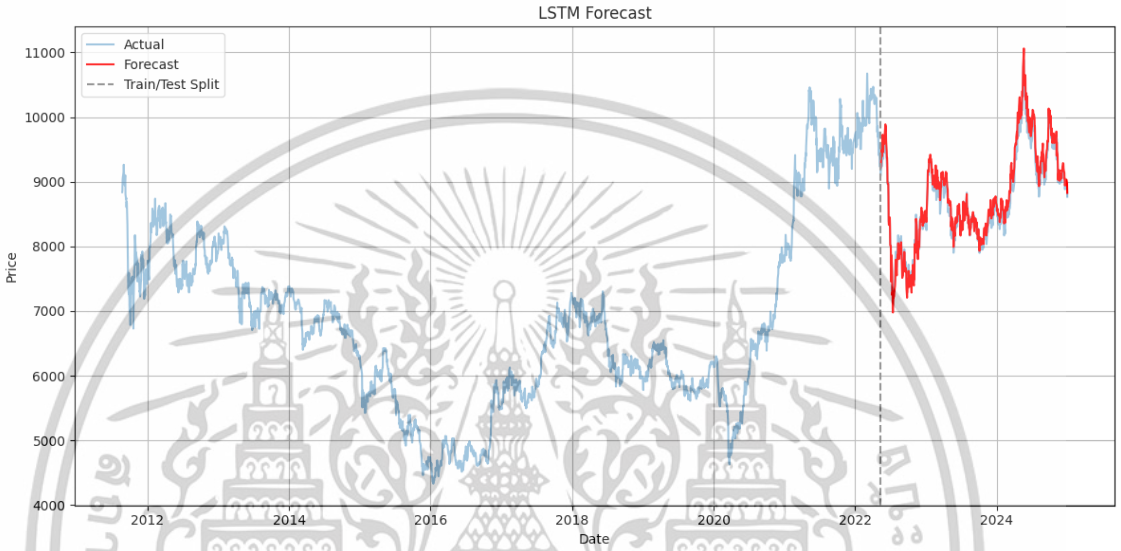


รูปที่ 4.11 แสดงกราฟผลการพยากรณ์ของแบบจำลองป่าสุ่มเทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ

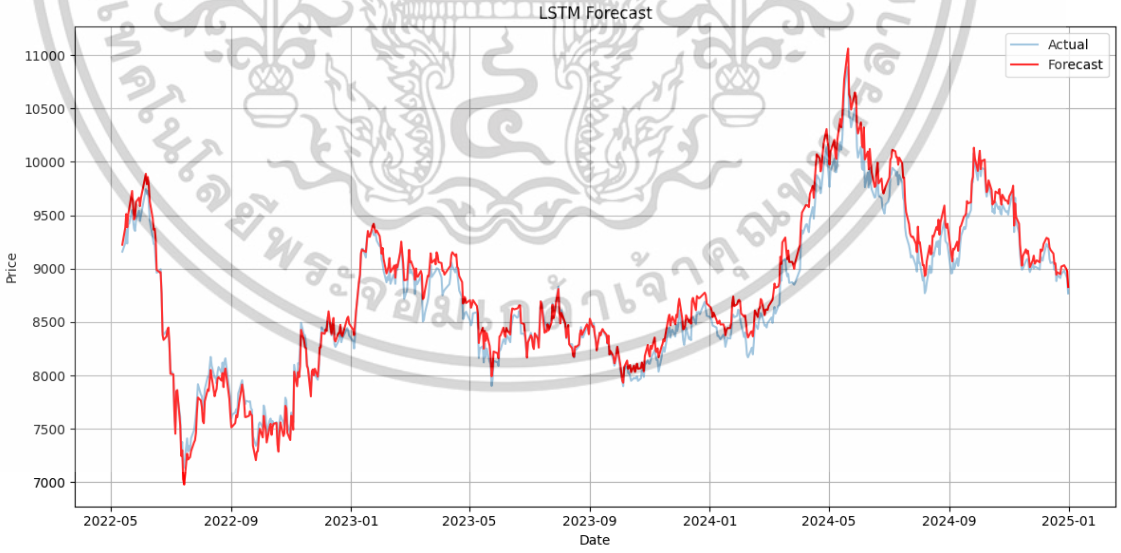
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.5 ผลการพัฒนาแบบจำลองการเรียนรู้เชิงลึก LSTM

ผลการพยากรณ์ของ LSTM พบว่าแบบจำลองสามารถจับแนวโน้มของข้อมูลได้ เมื่อพิจารณาจากดังรูปที่ 4.12 และ 4.13 จะเห็นได้ว่าค่าที่ได้จากการพยากรณ์จะมีแนวโน้มไปในทิศทางเดียวกันกับข้อมูลจริงที่ใช้ทดสอบ และเมื่อพิจารณาจากตัวชี้วัดทั้งสามค่าจะเห็นได้ว่าค่า MAE คลาดเคลื่อน 87.4854 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 117.5612 ดอลลาร์สหรัฐต่อตัน และค่า MAPE คลาดเคลื่อน 0.99% เมื่อพิจารณารวมกันพบว่าแบบจำลองนี้ถึงแม้จะจับแนวโน้มของข้อมูลได้แต่ยังมีความคลาดเคลื่อนจากค่าจริงพอสมควร



รูปที่ 4.12 กราฟผลการพยากรณ์ของแบบจำลองการเรียนรู้เชิงลึก LSTM เทียบกับข้อมูลจริง



รูปที่ 4.13 แสดงกราฟผลการพยากรณ์ของแบบจำลองการเรียนรู้เชิงลึก LSTM เทียบกับข้อมูลจริงเฉพาะข้อมูลทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง

แบบจำลอง	MAE (ดอลลาร์สหรัฐต่อตัน)	RMSE (ดอลลาร์สหรัฐต่อตัน)	MAPE (%)
ARIMA	969.6957	1,149.8894	11.59%
ต้นไม้ตัดสินใจ	97.2573	137.9294	1.10%
XGBoost	52.3623	70.6940	0.59%
ป่าสุ่ม	53.0240	71.3360	0.60%
LSTM	87.4854	117.5612	0.99%

เมื่อพิจารณาจากตารางที่ 4.1 จะเห็นได้ว่าแบบจำลอง XGBoost มีประสิทธิภาพในการทำนายได้ดีที่สุด แบบจำลองป่าสุ่มมีประสิทธิภาพในการทำนายเป็นอันดับที่สอง แบบจำลอง LSTM มีประสิทธิภาพในการทำนายเป็นอันดับที่ 3 และ แบบจำลอง ARIMA ประสิทธิภาพต่ำที่สุดอย่างเห็นได้ชัด แต่เมื่อพิจารณาโดยรวมแล้วมีเพียงแบบจำลอง XGBoost และ ป่าสุ่มที่ควรจะนำไปทดลองใช้งาน ถึงแม้จะไม่สามารถค่าสูงสุดในช่วงกลางปี พ.ศ. 2567 ได้

4.3 เปรียบเทียบคุณลักษณะที่มีผลต่อแบบจำลอง

จากผลการทดลองที่ 4.2 พบว่าแบบจำลอง XGBoost เป็นแบบจำลองที่ดีที่สุดในงานวิจัยนี้จึงได้พิจารณาเป็นแบบจำลองตัวแทนในการศึกษาคุณลักษณะของชุดข้อมูล

เมื่อพิจารณาตัดคุณลักษณะผลิตภัณฑ์มวลรวมในประเทศรายปี เพื่อศึกษาว่าชุดข้อมูลรายปีส่งผลต่อการพัฒนาแบบจำลองหรือไม่พบว่าค่าความคลาดเคลื่อน MAE คลาดเคลื่อน 53.0757 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 71.4272 ดอลลาร์สหรัฐต่อตัน และค่า MAPE คลาดเคลื่อน 0.60%

เมื่อพิจารณาตัดคุณลักษณะอัตราแลกเปลี่ยนค่าเงินปอนด์อังกฤษต่อดอลลาร์สหรัฐ พบว่าค่าความคลาดเคลื่อน MAE คลาดเคลื่อน 61.4311 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 84.3387 ดอลลาร์สหรัฐต่อตัน และค่า MAPE คลาดเคลื่อน 0.70%

เมื่อพิจารณาตัดคุณลักษณะราคาปิดโลหะทองคำ พบว่าค่าความคลาดเคลื่อน MAE คลาดเคลื่อน 54.1957 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 73.8112 ดอลลาร์สหรัฐต่อตัน และค่า MAPE คลาดเคลื่อน 0.60%

เมื่อพิจารณาตัดทุกคุณลักษณะยกเว้นราคาเปิดโลหะทองแดง พบว่าค่าความคลาดเคลื่อน MAE คลาดเคลื่อน 62.8452 ดอลลาร์สหรัฐต่อตัน ค่า RMSE คลาดเคลื่อน 82.0912 ดอลลาร์สหรัฐต่อตัน และค่า MAPE คลาดเคลื่อน 0.71%

เมื่อพิจารณาจากตารางที่ 4.2 พบว่าทุกการทดลองตัดคุณลักษณะจะส่งผลให้ค่าความคลาดเคลื่อนของแบบจำลองมีค่าสูงขึ้น โดยที่การตัดคุณลักษณะอัตราแลกเปลี่ยนค่าเงินปอนด์อังกฤษต่อดอลลาร์สหรัฐ ส่งผลมากที่สุด และการตัดคุณลักษณะผลิตภัณฑ์มวลรวมในประเทศรายปีส่งผลน้อยที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 แสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง XGBoost เพื่อศึกษาคุณลักษณะ

แบบจำลอง	**MAE (ดอลลาร์สหรัฐต่อตัน)**	**RMSE (ดอลลาร์สหรัฐต่อตัน)**	**MAPE (%)**
ไม่ตัดคุณลักษณะ	52.3623	70.6940	0.59%
ตัดผลิตภัณฑ์มวลรวมในประเทศรายปี	53.0757	71.4272	0.60%
ตัดอัตราแลกเปลี่ยนค่าเงินปอนด์อังกฤษต่อดอลลาร์สหรัฐ	61.4311	84.3387	0.70%
ตัดราคาปิดโลหะทองคำ	54.1957	73.8112	0.60%
ตัดทุกคุณลักษณะเหลือแค่ราคาเปิดโลหะทองแดง	62.8452	82.0912	0.71%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองสำหรับการพยากรณ์ราคาทองแดง โดยใช้ข้อมูลอนุกรมเวลาราคาปิดทองแดงรายวันจากตลาด London Metal Exchange (LME) และข้อมูลภายนอกจากการรวบรวมข้อมูล ในหัวข้อที่ 3.3 และเปรียบเทียบประสิทธิภาพของแบบจำลองการเรียนรู้ของเครื่องแบบมีผู้สอน 5 รูปแบบ ดังทฤษฎีตามที่อ้างในหัวข้อที่ 2.3 ถึง 2.8 ได้แก่ แบบจำลอง ARIMA, แบบจำลองต้นไม้ตัดสินใจ, แบบจำลอง XGBoost, แบบจำลองป่าสุ่ม และแบบจำลองการเรียนรู้เชิงลึก LSTM โดยใช้ค่าความคลาดเคลื่อนในการประเมินประสิทธิภาพแบบจำลองตามทฤษฎีที่อ้างอิงถึงในหัวข้อที่ 2.11.2 ได้แก่ MAE, RMSE และ MAPE เป็นตัวชี้วัด

จากผลการพัฒนาแบบจำลองในหัวข้อที่ 4.2 พบว่าแบบจำลอง XGBoost ให้ผลการพยากรณ์ที่แม่นยำที่สุด โดยมีค่า MAPE เพียง 0.59% รองลงมาคือแบบจำลองป่าสุ่ม, แบบจำลองการเรียนรู้เชิงลึก LSTM, และแบบจำลองต้นไม้ตัดสินใจ ตามลำดับ ขณะที่แบบจำลอง ARIMA ให้ผลการพยากรณ์คลาดเคลื่อนมากที่สุด โดยมีค่า MAPE สูงถึง 11.59%

ผลการเปรียบเทียบจากการพัฒนาแบบจำลองนี้แสดงให้เห็นว่าแบบจำลองการเรียนรู้ของเครื่อง โดยเฉพาะ แบบจำลอง XGBoost และแบบจำลองป่าสุ่ม สามารถจัดการกับข้อมูลอนุกรมเวลาที่มีความผันผวนได้ดีกว่าแบบจำลองทางสถิติแบบดั้งเดิมอย่าง ARIMA อันเนื่องมาจากข้อจำกัดของแบบจำลองที่เหมาะสมกับอนุกรมเวลาที่นิ่ง และมีรูปแบบที่ชัดเจน ส่งผลให้การพยากรณ์สำหรับชุดข้อมูลที่มีความผันผวนสูงเกิดคลาดเคลื่อนเป็นอย่างมาก

แม้ว่าแบบจำลอง LSTM จะเป็นแบบจำลองการเรียนรู้เชิงลึก ที่ทราบโดยทั่วกันว่ามีศักยภาพสูง แต่ประสิทธิภาพในการพยากรณ์ของ LSTM ในการวิจัยครั้งนี้ยังด้อยกว่าแบบจำลอง XGBoost และแบบจำลองป่าสุ่ม ซึ่งอาจเกิดจากข้อจำกัดในการปรับแต่งไฮเปอร์พารามิเตอร์ของผู้วิจัย อย่างไรก็ตามพบว่ามีเพียงแบบจำลองการเรียนรู้เชิงลึก LSTM ที่สามารถทำนายแนวโน้มของราคาทองแดงในช่วงกลางปี พ.ศ. 2567 หรือช่วงที่ราคาปิดทองแดงสูงที่สุดในชุดข้อมูลนี้ได้

เมื่อพิจารณาผลการทดลองการตัดคุณลักษณะในหัวข้อที่ 4.3 พบว่าทุกการทดลองส่งผลให้ค่าความคลาดเคลื่อนของแบบจำลอง XGBoost มีค่าเพิ่มมากขึ้นโดยการตัดคุณลักษณะอัตราแลกเปลี่ยนค่าเงินปอนด์อังกฤษต่อดอลลาร์สหรัฐ ส่งผลมากที่สุด การตัดคุณลักษณะราคาปิดโลหะทองคำ และการตัดผลิตภัณฑ์มวลรวมในประเทศรายปีมีค่าใกล้เคียงกันและส่งผลรองลงมา แสดงถึงตัวแปรตามที่นำมาพิจารณาเพิ่มเติมล้วนช่วยเพิ่มความแม่นยำของแบบจำลองในการพยากรณ์ราคาทองแดง ได้อย่างมีนัยสำคัญ

5.2 ข้อเสนอแนะ

1) ควรใช้วิธีการค้นหาค่าพารามิเตอร์ที่เหมาะสม เช่น วิธี Grid Search เพื่อปรับค่าพารามิเตอร์ของแบบจำลองให้มีประสิทธิภาพสูงสุด และลดข้อจำกัดที่เกิดจากความไม่ชำนาญของผู้ใช้งานในการตั้งค่าด้วยตนเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) ควรพิจารณานำแบบจำลองการเรียนรู้ของเครื่องรูปแบบอื่นๆ มาใช้ในการพยากรณ์ราคาทองแดงเพิ่มเติม เพื่อเปรียบเทียบและเลือกแบบจำลองที่เหมาะสมกับลักษณะของข้อมูล
3) ควรพิจารณาเพิ่มตัวแปรต้นอื่นๆ ที่อาจมีผลกระทบต่อราคาทองแดง เช่น ราคาโลหะชนิดอื่น, ราคาพลังงาน, ดัชนีหุ้น, และตัวชี้วัดทางเศรษฐกิจต่างๆ เพื่อช่วยเพิ่มความแม่นยำในการพยากรณ์
4) ควรเพิ่มการพิจารณาค่าความสำคัญของคุณลักษณะ (Feature Importance) เพื่อใช้ประกอบในการคัดเลือกตัวแปรต้นที่เหมาะสมในการพัฒนาแบบจำลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

Copper. (2025, June 15). In *Wikipedia*. https://en.wikipedia.org/wiki/Copper

GeeksforGeeks. (2024, August 13). *Time Series Analysis and Forecasting*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/time-series-analysis-and-forecasting/

Ben Salem, H. (2023, January 26). *Supervised vs Unsupervised vs Reinforcement learning: A comparative overview*. Medium. https://medium.com/@bensalemh300/supervised-vs-unsupervised-vs-reinforcement-learning-a3e7bcf1dd23

IBM (2025, May). *Supervised vs unsupervised learning*. IBM. https://www.ibm.com/think/topics/supervised-vs-unsupervised-learning

Techplayon (2025, May 15). *Machine Learning – Supervised, Unsupervised and Reinforcement*. Techplayon. https://www.techplayon.com/machine-learning-supervised-unsupervised-reinforcement/

Investopedia. (n.d.). *Autoregressive Integrated Moving Average (ARIMA)*. Investopedia. https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp

GeeksforGeeks. (2021, August 9). *Model Selection for ARIMA*. GeeksforGeeks. https://www.geeksforgeeks.org/r-language/model-selection-for-arima/

GeeksforGeeks. (2023, February 28). *Decision Tree in Machine Learning*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/decision-tree/

GeeksforGeeks. (2023, July 21). *XGBoost in Machine Learning*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/xgboost/

GeeksforGeeks. (2023, June 22). *Random Forest Algorithm in Machine Learning*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/random-forest-algorithm-in-machine-learning/

GeeksforGeeks. (2023, October 9). *Introduction to Long Short Term Memory (LSTM)*. GeeksforGeeks. https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

GeeksforGeeks. (2023, July 3). *Interquartile Range to Detect Outliers in Data*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/interquartile-range-to-detect-outliers-in-data/

Brownlee, J. (2019, November 20). *Understanding Boxplots*. KDnuggets. https://www.kdnuggets.com/2019/11/understanding-boxplots.html

GeeksforGeeks. (2023, July 17). *What is Feature Engineering?*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/what-is-feature-engineering/

Suraj, S. (2021, July 29). *Performance Metrics: Confusion Matrix, Precision, Recall and F1 Score*. Towards Data Science. https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262/

Sánchez Lasheras, et al. (2015). Forecasting the COMEX copper spot price by means of neural networks and ARIMA models. *Resources Policy, 45*, 37–43

Luo, et al. (2022). Multi-step-ahead copper price forecasting using a two-phase architecture based on an improved LSTM with novel input strategy and error correction. *Resources Policy, 79*, 102962

Hu, J. (2023). *Forecasting copper price using VAR and the XGBoost model: An experiment with a relatively small dataset* (Master's thesis, Lund University, School of Economics and Management).

Li, et al. (2024). A novel copper price forecasting ensemble method using adversarial interpretive structural model and sparrow search algorithm. *Resources Policy, 91*, 104892

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

การศึกษาวิจัยเรื่อง “การพยากรณ์ราคาทองแดงด้วยการเรียนรู้ของเครื่อง” เขียนโดยภาษาคอมพิวเตอร์ไพธอน 3 โดยมีคำสั่งดังนี้

1. การทำความสะอาดและจัดเตรียมข้อมูล

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from google.colab import drive
drive.mount('/content/drive')

df = pd.read_csv('/content/drive/My Drive/IS/Copper Futures Historical Data.csv')
df.head()
df.info()
print(df.isnull().sum())
df['Date'] = pd.to_datetime(df['Date'])

cols_to_clean = ['Price', 'Open', 'High', 'Low']
for col in cols_to_clean:
    df[col] = df[col].str.replace(',', '', regex=False).astype(float)
df['Change %'] = df['Change %'].str.replace('%', '', regex=False).astype(float) / 100
df['Vol.'] = df['Vol.'].str.replace('K', '', regex=False).astype(float) * 1000
df = df.rename(columns={'Change %': 'Change'})
df = df.sort_values('Date')

df.info()
df.describe()


plt.figure(figsize=(10, 5))
plt.plot(df['Date'], df['Price'], marker='.', linestyle='-',linewidth=1 ,  markersize=1,
markeredgewidth=0.5)
plt.title('Price over Time')
plt.xlabel('Date')
plt.ylabel('Price')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.grid(True)
plt.tight_layout()
plt.show()

plt.figure(figsize=(10, 5))
plt.plot(df['Date'], df['Vol.'], marker='.', linestyle='None',linewidth=1 ,  markersize=1,
markeredgewidth=0.5)
plt.title('Vol. over Time')
plt.xlabel('Date')
plt.ylabel('Vol.')
plt.grid(True)
plt.tight_layout()
plt.show()

df.drop(columns=['Vol.'], inplace=True)

correlation_matrix = df.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

from statsmodels.stats.outliers_influence import variance_inflation_factor
from statsmodels.tools.tools import add_constant


X = df[['Open', 'High', 'Low', 'Change']]
X = add_constant(X)

vif = pd.DataFrame()
vif['Variable'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

print(vif)

df["Price_range"] = df["High"] - df["Low"]
df = df.drop(columns=['High', 'Low'])
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
correlation_matrix = df.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

X = df[['Open', 'Change', 'Price_range']]
X = add_constant(X)

vif = pd.DataFrame()
vif['Variable'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

print(vif)

df['Year'] = df['Date'].dt.year

df_GDP = pd.read_csv('/content/drive/My Drive/IS/GDP.csv')

df_GDP.info()

cols_to_clean = ['Australia (USD Billions)', 'USA (USD Billions)', 'China (USD Billions)']
for col in cols_to_clean:
    df_GDP[col] = df_GDP[col].str.replace(',', '', regex=False).astype(float)

df_GDP.info()
df = pd.merge(df, df_GDP, on='Year', how='inner')

correlation_matrix = df.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

print(df.columns)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
X = df[['Open', 'Change', 'Price_range',
        'Chile (USD Billions)', 'Peru (USD Billions)',
        'Australia (USD Billions)', 'USA (USD Billions)',
        'China (USD Billions)']]
X = add_constant(X)

vif = pd.DataFrame()
vif['Variable'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

print(vif)

df2 = df.drop(columns=['China (USD Billions)', 'Peru (USD Billions)'])

correlation_matrix = df2.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

X = df2[['Open', 'Change', 'Price_range',
        'Chile (USD Billions)',
        'Australia (USD Billions)', 'USA (USD Billions)']]
X = add_constant(X)

vif = pd.DataFrame()
vif['Variable'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

print(vif)

df3 = df.drop(columns=['USA (USD Billions)', 'Peru (USD Billions)'])

correlation_matrix = df3.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.title('Correlation Matrix')
plt.show()

X = df3[['Open', 'Change', 'Price_range',
        'Chile (USD Billions)',
        'Australia (USD Billions)', 'China (USD Billions)']]
X = add_constant(X)
vif = pd.DataFrame()
vif['Variable'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

print(vif)

df_GBP = pd.read_csv('/content/drive/My Drive/IS/GBP_USD.csv')
df_Gold = pd.read_csv('/content/drive/My Drive/IS/Gold.csv')

df_GBP['Date'] = pd.to_datetime(df_GBP['Date'])
df_Gold['Date'] = pd.to_datetime(df_Gold['Date'])

df = pd.merge(df, df_GBP, on='Date', how='inner')
df = pd.merge(df, df_Gold, on='Date', how='inner')

df['Gold Price'] = df['Gold Price'].str.replace(',', '', regex=False).astype(float)

correlation_matrix = df.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

X = df[['Open', 'Change', 'Price_range',
        'Chile (USD Billions)', 'Australia (USD Billions)',
        'USA (USD Billions)', 'GBP per USD', 'Gold Price']]
X = add_constant(X)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
vif = pd.DataFrame()
vif['Variable'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

print(vif)

X = df[['Open', 'Change', 'Price_range',
       'Chile (USD Billions)', 'Australia (USD Billions)',
       'China (USD Billions)', 'GBP per USD', 'Gold Price']]
X = add_constant(X)

vif = pd.DataFrame()
vif['Variable'] = X.columns
vif['VIF'] = [variance_inflation_factor(X.values, i) for i in range(X.shape[1])]

print(vif)

df = df.drop(columns=['Year'])
df = df.drop(columns=['USA (USD Billions)', 'Peru (USD Billions)'])

correlation_matrix = df.corr()

plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

full_range = pd.date_range(start=df.index.min(), end=df.index.max(), freq='D')
df = df.reindex(full_range).ffill()

df.index.name = 'Date'
df = df.reset_index()
df.to_csv('/content/drive/My Drive/IS/Cleaned_Copper_Data.csv', index=False)
price = df['Price']

from scipy.fft import fft
import numpy as np
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
Q1 = df['Price'].quantile(0.25)
Q3 = df['Price'].quantile(0.75)
IQR = Q3 - Q1

outliers = df[(df['Price'] < Q1 - 1.5*IQR) | (df['Price'] > Q3 + 1.5*IQR)]

import matplotlib.pyplot as plt

plt.figure(figsize=(14, 6))
plt.plot(df.index, df['Price'], label='Price')
plt.scatter(outliers.index, outliers['Price'], color='red', label='Outliers')
plt.legend()
plt.title('Outlier Detection in Time Series')
plt.grid(True)
plt.show()
```

2. การหาคาบด้วย Fast Fourier transform

```
f = fft(price.to_numpy())
power = np.abs(f)**2
frequencies = np.fft.fftfreq(5476, d=1)

pos_mask = frequencies > 0
frequencies = frequencies[pos_mask]
power = power[pos_mask]

peak_index = np.argmax(power[1:100]) + 1
dominant_frequency = frequencies[peak_index]
print(dominant_frequency)

dominant_period = 1 / dominant_frequency
print(dominant_period)

plt.figure(figsize=(10, 5))
plt.plot(frequencies[:200], power[:200])
plt.axvline(x=dominant_frequency, color='red', linestyle='--', label=f'Dominant Freq =
{dominant_frequency:.4f}')
plt.title('Power Spectrum (FFT)')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.xlabel('Frequency (1/day)')
plt.ylabel('Power')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

3. การแยกส่วนประกอบอนุกรมเวลา

```
from statsmodels.tsa.seasonal import seasonal_decompose

result = seasonal_decompose(df['Price'], model='multiplicative', period = 1369)

plt.figure(figsize=(12, 8))
plt.subplot(4, 1, 1)
plt.plot(df['Price'], label='Original')
plt.legend(loc='upper left')

plt.subplot(4, 1, 2)
plt.plot(result.trend, label='Trend')
plt.legend(loc='upper left')

plt.subplot(4, 1, 3)
plt.plot(result.seasonal, label='Seasonal')
plt.legend(loc='upper left')

plt.subplot(4, 1, 4)
plt.plot(result.resid, label='Residual')
plt.legend(loc='upper left')

plt.tight_layout()
plt.show()
```

4. การพัฒนาแบบจำลอง

ARIMA

```
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.stattools import adfuller
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
df = pd.read_csv('/content/drive/My Drive/IS/Cleaned_Copper_Data.csv')
df['Date'] = pd.to_datetime(df['Date'])

y = df['Price']

result = adfuller(y.dropna())
print('ADF Statistic:', result[0])
print('p-value:', result[1])

y_diff = y.diff().dropna()result = adfuller(y_diff.dropna())
print("ADF Statistic:", result[0])
print("p-value:", result[1])

from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
plot_acf(y_diff, lags=10)
plot_pacf(y_diff, lags=10)
plt.show()
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_absolute_error, mean_squared_error

df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)

split_index = int(len(y) * 0.8)
train, test = y.iloc[:split_index], y.iloc[split_index:]

order = (0, 1, 0)

history = list(train)
predictions = []

for t in range(len(test)):
    model = ARIMA(history, order=order)
    model_fit = model.fit()
    yhat = model_fit.forecast()[0]
    predictions.append(yhat)
    history.append(yhat)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
forecast_series = pd.Series(predictions, index=test.index)

compare = pd.DataFrame({'y': test, 'y_pred': predictions})

mae  = mean_absolute_error(compare['y'], compare['y_pred'])
mse  = mean_squared_error(compare['y'], compare['y_pred'])
rmse = np.sqrt(mse)
mape = np.mean(np.abs((compare['y'] - compare['y_pred']) / compare['y'])) * 100

print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")

plt.figure(figsize=(12, 6))
plt.plot(y, label='Actual', alpha=0.5)
plt.plot(forecast_series, label='Forecast (Test)', color='red')
plt.axvline(test.index[0], color='gray', linestyle='--', label='Train/Test Split')
plt.legend()
plt.xlabel('Date')
plt.ylabel('Price')
plt.title('ARIMA - Rolling Forecast on Test Set')
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
test.plot(label='Actual', alpha=0.5)
forecast_series.plot(label='Forecast', color='red')  # ค่าทำนายจาก ARIMA
plt.legend()
plt.title('ARIMA Forecast on Test Set')
plt.xlabel('Date')
plt.ylabel('Price')
plt.grid(True)
plt.tight_layout()
plt.show()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Decision Tree

```
from sklearn.tree import DecisionTreeRegressor

df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)

lags = [1, 7, 14, 30, 120, 180, 365]
for col in df.columns:
    for lag in lags:
        df[f'{col}_lag_{lag}'] = df[col].shift(lag)

df = df.dropna()

X = df.drop(columns=['Price'])
y = df['Price']

train_size = int(len(df) * 0.8)
X_train, X_test = X.iloc[:train_size], X.iloc[train_size:]
y_train, y_test = y.iloc[:train_size], y.iloc[train_size:]

model = DecisionTreeRegressor(
    max_depth=5,
    min_samples_split=10,
    min_samples_leaf=5,
    random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
y_pred_series = pd.Series(y_pred, index=y_test.index, name='y_pred')

compare = pd.DataFrame({'y': y_test, 'y_pred': y_pred_series})
mae  = mean_absolute_error(compare['y'], compare['y_pred'])
mse  = mean_squared_error(compare['y'], compare['y_pred'])
rmse = np.sqrt(mse)
mape = np.mean(np.abs((compare['y'] - compare['y_pred']) / compare['y'])) * 100
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")

plt.figure(figsize=(12, 6))
plt.plot(y, label='Actual', alpha=0.5)
plt.plot(y_pred_series, label='Forecast', color='red')
plt.axvline(y_test.index[0], color='gray', linestyle='--', label='Train/Test Split')
plt.title('Decision Tree Forecast')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.tight_layout()
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
plt.plot(y_test, label='Actual', alpha=0.5)
plt.plot(y_pred_series, label='Forecast', color='red')
plt.title('Decision Tree Forecast')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.tight_layout()
plt.grid(True)
plt.show()
```

XGBoost

```
from xgboost import XGBRegressor
import matplotlib.pyplot as plt

df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)

lags = [1, 7, 14, 30, 120, 180, 365]
for col in df.columns:
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    for lag in lags:
        df[f'{col}_lag_{lag}'] = df[col].shift(lag)

df = df.dropna()

X = df.drop(columns=['Price'])
y = df['Price']

train_size = int(len(df) * 0.8)
X_train, X_test = X.iloc[:train_size], X.iloc[train_size:]
y_train, y_test = y.iloc[:train_size], y.iloc[train_size:]

model = XGBRegressor(
    n_estimators=100,
    learning_rate=0.05,
    max_depth=6,
    subsample=0.8,
    colsample_bytree=0.8,
    random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
y_pred_series = pd.Series(y_pred, index=y_test.index, name='y_pred')

compare = pd.DataFrame({'y': y_test, 'y_pred': y_pred_series})
mae  = mean_absolute_error(compare['y'], compare['y_pred'])
mse  = mean_squared_error(compare['y'], compare['y_pred'])
rmse = np.sqrt(mse)
mape = np.mean(np.abs((compare['y'] - compare['y_pred']) / compare['y'])) * 100

print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")

plt.figure(figsize=(12, 6))
plt.plot(y, label='Actual', alpha=0.5)
plt.plot(y_pred_series, label='Forecast', color='red')
plt.axvline(y_test.index[0], color='gray', linestyle='--', label='Train/Test Split')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.title('XGBoost Forecast')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.tight_layout()
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
plt.plot(y_test, label='Actual', alpha=0.5)
plt.plot(y_pred_series, label='Forecast', color='red')
plt.title('XGBoost Forecast')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.tight_layout()
plt.grid(True)
plt.show()
```

Random Forrest

```
from sklearn.ensemble import RandomForestRegressor

df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)

lags = [1, 7, 14, 30, 120, 180, 365]
for col in df.columns:
    for lag in lags:
        df[f'{col}_lag_{lag}'] = df[col].shift(lag)

df = df.dropna()

X = df.drop(columns=['Price'])
y = df['Price']

train_size = int(len(df) * 0.8)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
X_train, X_test = X.iloc[:train_size], X.iloc[train_size:]
y_train, y_test = y.iloc[:train_size], y.iloc[train_size:]

model = RandomForestRegressor(
    n_estimators=100,
    max_depth=5,
    min_samples_leaf=5,
    random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
y_pred_series = pd.Series(y_pred, index=y_test.index, name='y_pred')

mae = mean_absolute_error(y_test, y_pred_series)
rmse = np.sqrt(mean_squared_error(y_test, y_pred_series))
mape = np.mean(np.abs((y_test - y_pred_series) / y_test)) * 100

print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")


plt.figure(figsize=(12, 6))
plt.plot(y, label='Actual', alpha=0.5)
plt.plot(y_pred_series, label='Forecast (Random Forest)', color='red')
plt.axvline(y_test.index[0], color='gray', linestyle='--', label='Train/Test Split')
plt.title('Random Forest Forecast')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.tight_layout()
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
plt.plot(y_test, label='Actual', alpha=0.5)
plt.plot(y_pred_series, label='Forecast (Random Forest)', color='red')
plt.title('Random Forest Forecast')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.tight_layout()
plt.grid(True)
plt.show()
```

LSTM

```
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout

df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)

X = df.drop(columns=['Price'])
y = df['Price']

x_scaler = MinMaxScaler()
y_scaler = MinMaxScaler()

X_scaled = x_scaler.fit_transform(X)
y_scaled = y_scaler.fit_transform(y.values.reshape(-1, 1))

def create_sequences(X, y, time_steps=365):
    Xs, ys = [], []
    for i in range(len(X) - time_steps):
        Xs.append(X[i:i+time_steps])
        ys.append(y[i+time_steps])
    return np.array(Xs), np.array(ys)

TIME_STEPS = 365
X_seq, y_seq = create_sequences(X_scaled, y_scaled, TIME_STEPS)

train_size = int(len(X_seq) * 0.8)
X_train, X_test = X_seq[:train_size], X_seq[train_size:]
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
y_train, y_test = y_seq[:train_size], y_seq[train_size:]

from tensorflow.keras.callbacks import EarlyStopping

model = Sequential()
model.add(LSTM(units=128, activation='tanh', return_sequences=False,
input_shape=(TIME_STEPS, X.shape[1])))
model.add(Dropout(0.2))
model.add(Dense(1))
model.compile(optimizer='adam', loss='mse')


early_stop = EarlyStopping(
    monitor='val_loss',
    patience=10,
    restore_best_weights=True
)

history = model.fit(
    X_train, y_train,
    validation_data=(X_test, y_test),
    epochs=50,
    batch_size=32,
    callbacks=[early_stop],
    verbose=1
)

y_pred_scaled = model.predict(X_test)
y_pred = y_scaler.inverse_transform(y_pred_scaled)
y_true = y_scaler.inverse_transform(y_test)

index_test = df.index[TIME_STEPS + train_size:]
y_pred_series = pd.Series(y_pred.flatten(), index=index_test, name='y_pred')
y_true_series = pd.Series(y_true.flatten(), index=index_test, name='y_true')

mae = mean_absolute_error(y_true_series, y_pred_series)
rmse = np.sqrt(mean_squared_error(y_true_series, y_pred_series))
mape = np.mean(np.abs((y_true_series - y_pred_series) / y_true_series)) * 100
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")

plt.figure(figsize=(12, 6))
plt.plot(y, label='Actual', alpha=0.5)
plt.plot(y_pred_series, label='Forecast', color='red')
plt.axvline(y_true_series.index[0], color='gray', linestyle='--', label='Train/Test Split')
plt.title('LSTM Forecast')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.tight_layout()
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
plt.plot(y_true_series, label='Actual', alpha=0.5)
plt.plot(y_pred_series, label='Forecast', color='red')
plt.title('LSTM Forecast')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
plt.tight_layout()
plt.grid(True)
plt.show()
```

ทดสอบตัดตัวแปร

```
df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)

lags = [1, 7, 14, 30, 120, 180, 365]
for col in df.columns:
    for lag in lags:
        df[f'{col}_lag_{lag}'] = df[col].shift(lag)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
df = df.dropna()

X = df.drop(columns=['Price','Chile (USD Billions)', 'Australia (USD Billions)',
        'China (USD Billions)'])
y = df['Price']

train_size = int(len(df) * 0.8)
X_train, X_test = X.iloc[:train_size], X.iloc[train_size:]
y_train, y_test = y.iloc[:train_size], y.iloc[train_size:]

model = XGBRegressor(objective='reg:squarederror', n_estimators=100,
random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
y_pred_series = pd.Series(y_pred, index=y_test.index, name='y_pred')

compare = pd.DataFrame({'y': y_test, 'y_pred': y_pred_series})
mae  = mean_absolute_error(compare['y'], compare['y_pred'])
mse  = mean_squared_error(compare['y'], compare['y_pred'])
rmse = np.sqrt(mse)
mape = np.mean(np.abs((compare['y'] - compare['y_pred']) / compare['y'])) * 100

print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")

df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)

lags = [1, 7, 14, 30, 120, 180, 365]
for col in df.columns:
    for lag in lags:
        df[f'{col}_lag_{lag}'] = df[col].shift(lag)

df = df.dropna()
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
X = df.drop(columns=['Price','GBP per USD'])
y = df['Price']

train_size = int(len(df) * 0.8)
X_train, X_test = X.iloc[:train_size], X.iloc[train_size:]
y_train, y_test = y.iloc[:train_size], y.iloc[train_size:]

model = XGBRegressor(objective='reg:squarederror', n_estimators=100,
random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
y_pred_series = pd.Series(y_pred, index=y_test.index, name='y_pred')

compare = pd.DataFrame({'y': y_test, 'y_pred': y_pred_series})
mae  = mean_absolute_error(compare['y'], compare['y_pred'])
mse  = mean_squared_error(compare['y'], compare['y_pred'])
rmse = np.sqrt(mse)
mape = np.mean(np.abs((compare['y'] - compare['y_pred']) / compare['y'])) * 100

print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")

df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)

lags = [1, 7, 14, 30, 120, 180, 365]
for col in df.columns:
    for lag in lags:
        df[f'{col}_lag_{lag}'] = df[col].shift(lag)

df = df.dropna()

X = df.drop(columns=['Price','Gold Price'])
y = df['Price']
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
train_size = int(len(df) * 0.8)
X_train, X_test = X.iloc[:train_size], X.iloc[train_size:]
y_train, y_test = y.iloc[:train_size], y.iloc[train_size:]

model = XGBRegressor(objective='reg:squarederror', n_estimators=100,
random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
y_pred_series = pd.Series(y_pred, index=y_test.index, name='y_pred')

compare = pd.DataFrame({'y': y_test, 'y_pred': y_pred_series})
mae  = mean_absolute_error(compare['y'], compare['y_pred'])
mse  = mean_squared_error(compare['y'], compare['y_pred'])
rmse = np.sqrt(mse)
mape = np.mean(np.abs((compare['y'] - compare['y_pred']) / compare['y'])) * 100

print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")

df['Date'] = pd.to_datetime(df['Date'])
df = df.sort_values('Date')
df.set_index('Date', inplace=True)
lags = [1, 7, 14, 30]
for col in df.columns:
    for lag in lags:
        df[f'{col}_lag_{lag}'] = df[col].shift(lag)

df = df.dropna()

X = df['Open']
y = df['Price']

train_size = int(len(df) * 0.8)
X_train, X_test = X.iloc[:train_size], X.iloc[train_size:]
y_train, y_test = y.iloc[:train_size], y.iloc[train_size:]
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
model = XGBRegressor(objective='reg:squarederror', n_estimators=100,
random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
y_pred_series = pd.Series(y_pred, index=y_test.index, name='y_pred')

compare = pd.DataFrame({'y': y_test, 'y_pred': y_pred_series})
mae  = mean_absolute_error(compare['y'], compare['y_pred'])
mse  = mean_squared_error(compare['y'], compare['y_pred'])
rmse = np.sqrt(mse)
mape = np.mean(np.abs((compare['y'] - compare['y_pred']) / compare['y'])) * 100

print(f"MAE : {mae:,.4f}")
print(f"RMSE: {rmse:,.4f}")
print(f"MAPE: {mape:.2f}%")
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นาย วรัญชัย ไชยปัญญา
วัน เดือน ปีเกิด	25 มีนาคม พ.ศ. 2538
ที่อยู่ปัจจุบัน	555/271 หมู่ 15 ตำบลบางแก้ว อำเภอบางพลี จังหวัดสมุทรปราการ 10540
ประวัติการศึกษา	(2560) วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมขนส่งทางราง สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง เกรดเฉลี่ย 2.59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้