

การพยากรณ์ราคาหุ้นโดยใช้การวิเคราะห์อนุกรมเวลา SARIMA Stock Price Forecasting using SARIMA Time Series

ซัลมาน สาเล้ง, มุฮัยมิน เจนะนะ, สมศรี จารุผดุง, วรารัตน์ จักรหวัด, ศิริรัตน์ วณิชโยบล*

Salman Saleng, Muhaimin Chena, Somsri Jarupadung, Wararat Jakawat, Sirirut Vanichayobon*

สาขาวิชาวิทยาการคอมพิวเตอร์ สาขาวิทยาศาสตร์การคำนวณ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์
Computer Science, Division of Computational Science, Faculty of Science, Prince of Songkla University

บทคัดย่อ

บทความนี้นำเสนอการพยากรณ์ราคาหุ้นโดยใช้การวิเคราะห์อนุกรมเวลาแบบ SARIMA ซึ่งเหมาะสมสำหรับข้อมูลที่มีลักษณะตามฤดูกาล การศึกษานี้ใช้ข้อมูลราคาปิดรายวันของหุ้นบริษัทซีพี ออลล์ จำกัด (มหาชน) ในช่วงปี 2557 ถึง 2567 โดยขั้นตอนการวิเคราะห์ประกอบด้วยการเก็บรวบรวมข้อมูล การจัดรูปแบบและทำความสะอาดข้อมูล การวิเคราะห์ข้อมูลก่อนสร้างแบบจำลอง และการสร้างและทดสอบแบบจำลอง SARIMA โดยใช้ภาษา Python บน Google Colab ผลการวิจัยแสดงให้เห็นว่าแบบจำลอง SARIMA(0,1,1)(0,0,2,12) สามารถพยากรณ์แนวโน้มราคาหุ้นได้อย่างมีประสิทธิภาพ โดยมีค่า RMSE เท่ากับ 6.226 บาท ซึ่งแสดงถึงความคลาดเคลื่อนเฉลี่ยระหว่างราคาที่พยากรณ์และราคาจริงอยู่ในระดับต่ำ สะท้อนความแม่นยำของแบบจำลองในการจับแนวโน้มราคาหุ้น ผลการวิจัยนี้ช่วยในการสนับสนุนการตัดสินใจของนักลงทุนระยะยาวได้อย่างมีประสิทธิภาพ

คำสำคัญ: อนุกรมเวลา SARIMA การพยากรณ์ราคาหุ้น RMSE

Abstract

This article presents stock price forecasting using SARIMA time series analysis, which is suitable for seasonal data. The study utilizes daily closing prices of CP All Public Company Limited (CPALL) stock from 2014 to 2024. The analysis process includes data collection, data formatting and cleaning, exploratory data analysis, and the creation and testing of the SARIMA model using Python on Google Colab. The research findings indicate that the SARIMA(0,1,1)(0,0,2,12) model can effectively forecast stock price trends with an RMSE of 6.226 baht, which represents a low average deviation between forecasted and actual prices. This reflects the accuracy of the model in capturing stock price trends. These research findings can effectively support long-term investment decision-making.

Keywords: Time series, SARIMA, Stock forecasting, RMSE

* Corresponding author : sirirut.v@psu.ac.th

1. บทนำ

ในภาวะเศรษฐกิจปัจจุบัน ตลาดหลักทรัพย์ถือเป็นกลไกสำคัญในการระดมทุน สร้างผลตอบแทนทางการเงิน และขับเคลื่อนเสถียรภาพทางเศรษฐกิจของบุคคลและองค์กร โดยราคาหุ้นสะท้อนพลวัตทางเศรษฐกิจและความเชื่อมั่นของนักลงทุนในแต่ละช่วงเวลา อย่างไรก็ตาม ราคาหุ้นมีลักษณะผันผวนสูง อันเป็นผลมาจากปัจจัยแวดล้อมที่ซับซ้อนและหลากหลาย ทั้งภาวะเศรษฐกิจมหภาค นโยบายทางการเงิน เหตุการณ์ทางภูมิรัฐศาสตร์ ความเชื่อมั่นของนักลงทุน รวมถึงพฤติกรรมการบริโภคที่เปลี่ยนแปลงตามฤดูกาล ปัจจัยเหล่านี้ก่อให้เกิดความเสี่ยงและข้อจำกัดในกระบวนการตัดสินใจลงทุน การมีเครื่องมือพยากรณ์ราคาหุ้นที่มีประสิทธิภาพจึงมีความสำคัญอย่างยิ่งในการสนับสนุนการตัดสินใจลงทุนอย่างมีเหตุผล ลดความเสี่ยง และเพิ่มโอกาสในการได้รับผลตอบแทนที่เหมาะสม

การวิเคราะห์ข้อมูลอนุกรมเวลาได้รับการยอมรับในฐานะเครื่องมือเชิงปริมาณที่มีประสิทธิภาพในการศึกษารูปแบบการเปลี่ยนแปลงของข้อมูลตามลำดับช่วงเวลา เพื่อพยากรณ์แนวโน้มที่อาจเกิดขึ้นในช่วงเวลาถัดไปได้อย่างเป็นระบบ (Tibshirani, 2023) ในบรรดาแบบจำลองอนุกรมเวลาที่ได้รับการพัฒนา แบบจำลอง Auto-Regressive Integrated Moving Average (ARIMA) นับเป็นแบบจำลองพื้นฐานที่ได้รับการยอมรับอย่างกว้างขวางในการพยากรณ์ข้อมูลที่มีแนวโน้มทั่วไป (Box et al., 2015; Saadeddin, 2024) อย่างไรก็ตาม แบบจำลอง ARIMA มีข้อจำกัดในการจัดการกับรูปแบบความผันผวนตามฤดูกาล จึงนำไปสู่การพัฒนาแบบจำลอง Seasonal Autoregressive Integrated Moving Average (SARIMA) ซึ่งบูรณาการองค์ประกอบฤดูกาลเข้าสู่กระบวนการพยากรณ์ (Artley, 2022) แบบจำลอง SARIMA มีพื้นฐานทางสถิติที่แข็งแกร่ง ให้ความสมดุลระหว่างความแม่นยำและความซับซ้อนในการคำนวณ และเหมาะสมอย่างยิ่งสำหรับข้อมูลตัวแปรเดียวที่มีรูปแบบตามฤดูกาลชัดเจน

งานวิจัยนี้มุ่งประยุกต์ใช้แบบจำลอง SARIMA ในการพยากรณ์ราคาปิดรายเดือนของหุ้นบริษัท ซีพี ออลล์ จำกัด (มหาชน) หรือ CPALL ซึ่งเป็นหนึ่งในหุ้นที่ได้รับความนิยมสูงในตลาดหลักทรัพย์แห่งประเทศไทย เนื่องจากบริษัทมีบทบาทสำคัญในธุรกิจค้าปลีกผ่านเครือข่าย 7-Eleven ที่กระจายอยู่ทั่วประเทศ ส่งผลให้ราคาหุ้นมีความผันผวนตามพฤติกรรมผู้บริโภคที่เปลี่ยนแปลงตามฤดูกาลอย่างชัดเจน โดยวัตถุประสงค์ของหุ้น CPALL สะท้อนรูปแบบการจับจ่ายตามเทศกาลและผลประกอบการรายไตรมาสที่เกิดซ้ำอย่างสม่ำเสมอ ลักษณะดังกล่าวทำให้หุ้น CPALL เป็นกรณีศึกษาที่เหมาะสมอย่างยิ่งสำหรับการประยุกต์ใช้แบบจำลอง SARIMA การศึกษานี้ใช้ข้อมูลราคาปิดรายวันในช่วงปี พ.ศ. 2557 ถึง พ.ศ. 2567 และดำเนินการวิเคราะห์ด้วยภาษา Python บน Google Colab ร่วมกับ Microsoft Power BI เพื่อสร้างกราฟเชิงโต้ตอบ ผลการวิจัยนี้คาดว่าจะประโยชน์ต่อนักลงทุนระยะยาวในการวางแผนการลงทุนอย่างมีข้อมูลสนับสนุน รวมทั้งเป็นแนวทางสำหรับการประยุกต์ใช้แบบจำลองอนุกรมเวลากับข้อมูลทางการเงินอื่น ๆ ที่มีลักษณะตามฤดูกาลต่อไป

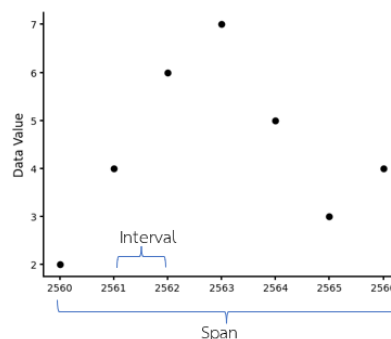
2. วัตถุประสงค์การวิจัย

1. เพื่อสร้างและทดสอบประสิทธิภาพของแบบจำลอง SARIMA ในการพยากรณ์ราคาปิดหุ้นรายเดือนของบริษัท ซีพี ออลล์ จำกัด (มหาชน)
2. เพื่อประเมินความแม่นยำของแบบจำลองในการทำนายแนวโน้มราคาหุ้น โดยเปรียบเทียบกับพยากรณ์กับข้อมูลจริงเพื่อวางแผนและสนับสนุนการตัดสินใจลงทุนระยะยาว

3. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ส่วนนี้จะกล่าวถึงแนวคิดของข้อมูลอนุกรมเวลา แบบจำลอง SARIMA และงานวิจัยที่เกี่ยวข้อง
ข้อมูลอนุกรมเวลา

ข้อมูลอนุกรมเวลา (Time Series Data) (Tibshirani, 2023) หมายถึงข้อมูลเกี่ยวกับเหตุการณ์เดียวกันภายในช่วงเวลาที่กำหนด โดยข้อมูลจะถูกรวบรวมและจัดเก็บ ซึ่งข้อมูลมีลักษณะที่ถูกจัดลำดับตามช่วงเวลาต่อเนื่อง เช่น ชั่วโมง วัน สัปดาห์ เดือน หรือปี ทำให้สามารถแสดงถึงการเปลี่ยนแปลงของปรากฏการณ์ที่เกิดขึ้นตามเวลาที่เปลี่ยนแปลงไป ตัวอย่างของข้อมูลอนุกรมเวลา ได้แก่ ข้อมูลราคาหุ้นที่บันทึกเป็นรายวัน จำนวนผู้เข้าชมเว็บไซต์รายชั่วโมง ระดับ PM2.5 รายวัน อุณหภูมิอากาศรายเดือน หรือระดับน้ำในเขื่อนรายวัน เป็นต้น ข้อมูลอนุกรมเวลาประกอบด้วยองค์ประกอบหลัก 2 อย่าง ได้แก่ ระยะเวลา (Span) คือ ช่วงเวลารวมที่ใช้ในการเก็บข้อมูล เช่น 5 ปี หรือ 1 เดือน เป็นต้น และช่วงเวลา (Interval) คือ ช่วงห่างระหว่างการบันทึกข้อมูลแต่ละครั้ง เช่น ทุกวัน ทุกชั่วโมง หรือทุกปี เป็นต้น ภาพที่ 1 แสดงข้อมูลอนุกรมเวลา ซึ่งประกอบด้วยองค์ประกอบหลัก 2 ส่วน ได้แก่ ระยะเวลาและช่วงเวลา โดยครอบคลุมระยะเวลาดังแต่ปี พ.ศ. 2560 ถึงปี พ.ศ. 2566 และมีการบันทึกข้อมูลในลักษณะรายปี



ภาพที่ 1 องค์ประกอบหลัก 2 อย่างของข้อมูลอนุกรมเวลา

ลักษณะสำคัญของข้อมูลอนุกรมเวลา คือ ความสัมพันธ์ระหว่างข้อมูลในอดีตและอนาคต ซึ่งสามารถใช้ในการวิเคราะห์เพื่อศึกษาพฤติกรรมและลักษณะของข้อมูล และค้นหารูปแบบหรือแนวโน้มที่เกิดขึ้นในอดีตเพื่อทำการพยากรณ์สิ่งที่จะเกิดขึ้นในอนาคต เช่น การพยากรณ์ราคาปิดหุ้น การพยากรณ์ยอดขาย การพยากรณ์

อุณหภูมิ หรือการพยากรณ์ระดับมลพิษ เป็นต้น การวิเคราะห์ข้อมูลอนุกรมเวลาถือเป็นเครื่องมือเชิงปริมาณที่มีประสิทธิภาพสูงในการศึกษาและทำความเข้าใจรูปแบบการเปลี่ยนแปลงของข้อมูลตามลำดับช่วงเวลา ทั้งนี้ เพื่อประโยชน์ในการสนับสนุนกระบวนการตัดสินใจเชิงกลยุทธ์ทางธุรกิจ การวางแผนนโยบาย ตลอดจนการพยากรณ์แนวโน้มในอนาคตอย่างมีระบบและน่าเชื่อถือ

ข้อมูลอนุกรมเวลาที่เป็นข้อมูลคงที่ (Stationary Data) (Saadeddin, 2024) หมายถึงข้อมูลที่ไม่มีแนวโน้มหรือฤดูกาลที่ชัดเจน กล่าวคือข้อมูลค่าทางสถิติ เช่น ค่าเฉลี่ย (Mean) ความแปรปรวน (Variance) หรือความแปรผันร่วม (Covariance) คงที่เมื่อเวลาผ่านไป ข้อมูลที่เป็นข้อมูลคงที่เหมาะต่อการสร้างแบบจำลองและการวิเคราะห์ เนื่องจากไม่มีลักษณะที่เปลี่ยนแปลงที่อาจทำให้การพยากรณ์ผิดพลาด ข้อมูลอนุกรมเวลาที่เป็นข้อมูลคงที่จำเป็นในการพยากรณ์ด้วยวิธี Box-Jenkins (กระบวนการพัฒนาแบบจำลองอนุกรมเวลา ARIMA หรือ SARIMA) (Box et al., 2015) หากข้อมูลมีลักษณะไม่คงที่ที่จะต้องดำเนินการแปลงข้อมูลด้วยวิธีหาผลต่าง (Differencing) ก่อนนำไปสร้างแบบจำลอง

ข้อมูลอนุกรมเวลาที่เป็นข้อมูลไม่คงที่ (Non-Stationary Data) หมายถึงข้อมูลที่มีค่าทางสถิติเปลี่ยนแปลงไปตามเวลา ซึ่งอาจเกิดจากการมีแนวโน้มเพิ่มขึ้นหรือลดลงเมื่อเวลาผ่านไป เกิดจากความผันผวนตามฤดูกาล โดยมีรูปแบบที่เกิดซ้ำตามช่วงเวลาที่กำหนด เช่น การเพิ่มขึ้นของยอดขายในช่วงเทศกาลหรือความแปรปรวนที่ไม่คงที่ในแต่ละช่วงเวลา ข้อมูลประเภทนี้ไม่สามารถคาดการณ์ได้ง่าย ๆ เนื่องจากลักษณะของข้อมูลเปลี่ยนแปลงตลอดเวลา ไม่สามารถนำมาใช้ในการสร้างแบบจำลอง ARIMA หรือ SARIMA ได้โดยตรง จำเป็นต้องทำให้ข้อมูลมีความคงที่ก่อน

Augmented Dickey-Fuller (ADF Test) (Subhash, 2023) เป็นการตรวจสอบลักษณะข้อมูลว่าชุดข้อมูลคงที่หรือไม่ ซึ่งมีสมมติฐานหลัก (H_0) คือข้อมูลมีลักษณะ Unit Root (ไม่คงที่) โดยพิจารณาจากค่าสถิติทดสอบเทียบกับค่าวิกฤต หรือพิจารณาค่า p-value (Dickey & Fuller, 1979)

- ถ้าค่า p-value ≤ 0.05 ปฏิเสธสมมติฐานหลัก แสดงว่าข้อมูลมีแนวโน้มว่าจะเป็นข้อมูลคงที่
- ถ้าค่า p-value > 0.05 ยอมรับสมมติฐานหลัก แสดงว่าข้อมูลมีแนวโน้มว่าจะเป็นข้อมูลไม่คงที่

แบบจำลอง SARIMA

แบบจำลอง SARIMA ใช้ในการพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Forecasting) เพื่อจัดการกับข้อมูลที่มีความผันผวนตามฤดูกาล (Seasonality) เช่น อุณหภูมิ ปริมาณฝน ราคาหุ้น หรือยอดขายที่มีความผันผวนตามเดือนหรือฤดูกาล แบบจำลอง SARIMA หรือ Seasonal ARIMA (Artley, 2022) เป็นการพัฒนาต่อยอดจากแบบจำลอง ARIMA (Saadeddin, 2024) เพื่อให้สามารถจัดการกับข้อมูลที่มีฤดูกาลได้

แบบจำลอง ARIMA(p, d, q) ประกอบ 3 ส่วน ดังนี้ 1) Auto-Regressive: AR หรือค่า p หมายถึงจำนวนช่วงเวลาในอดีตที่ใช้ในการพยากรณ์ค่าในอนาคต 2) Integrated: I หรือค่า d หมายถึงการทำให้ข้อมูลมีความคงที่ ด้วยความแตกต่างของข้อมูล เพื่อทำให้ข้อมูลคงที่ และ 3) Moving Average: MA หรือค่า q หมายถึงจำนวนข้อผิดพลาดที่ใช้ในการพยากรณ์ ข้อดีของ ARIMA คือความสามารถในการจัดการกับข้อมูลที่ซับซ้อนได้ดี แม้จะไม่มีฤดูกาล แต่ข้อเสียคือไม่เหมาะกับข้อมูลที่มีความผันผวนตามฤดูกาล

แบบจำลอง SARIMA(p,d,q)(P,D,Q,s) ประกอบ 2 ส่วนหลัก คือ ส่วนที่ไม่ใช่ตามฤดูกาล (Non-Seasonal Components) ประกอบด้วยพารามิเตอร์ p, d, และ q ซึ่งเหมือนกับแบบจำลอง ARIMA โดยใช้ในการจัดการกับข้อมูลทั่วไป และส่วนที่ตามฤดูกาล (Seasonal Components) ประกอบด้วยพารามิเตอร์ P, D, Q และ s สำหรับการจัดการฤดูกาล P คือการถดถอยอัตโนมัติ D คือความแตกต่างแบบฤดูกาล Q คือการเฉลี่ยเคลื่อนที่แบบฤดูกาล และ s คือ ระยะเวลาของฤดูกาล เช่น 12 เดือน (P,D,Q,s) หมายถึง พารามิเตอร์ P, D, Q ถูกคำนวณโดยพิจารณาการเปลี่ยนแปลงที่เกิดขึ้นใน s ช่วงเวลาที่ผ่านไป ตัวอย่างเช่น (P,D,Q,s) = (1,1,0, 12) หมายถึง พารามิเตอร์ P, D, Q ถูกคำนวณโดยพิจารณาการเปลี่ยนแปลงที่เกิดขึ้นใน 12 ช่วงเวลาที่ผ่านไปซึ่งถ้าใช้ชุดข้อมูลรายเดือน หมายความว่าข้อมูลปัจจุบันได้รับอิทธิพลจากข้อมูลในรอบ 12 เดือนก่อนหน้า ตารางที่ 1 แสดงการเปรียบเทียบความแตกต่างระหว่างแบบจำลอง ARIMA และ SARIMA

ตารางที่ 1 เปรียบเทียบแบบจำลอง ARIMA และ SARIMA

ลักษณะ	ARIMA	SARIMA
พารามิเตอร์	p, d, q	p, d, q, P, D, Q, s
การจัดการฤดูกาล	ไม่รองรับ	จัดการฤดูกาลได้ดี
ความซับซ้อน	ง่ายกว่า SARIMA และเหมาะกับข้อมูล	ซับซ้อนกว่า ARIMA แต่เหมาะกับข้อมูล
และความเหมาะสม	ทั่วไปที่มีแนวโน้มแต่ไม่มีลักษณะฤดูกาล	ที่มีแนวโน้มและฤดูกาล
ความยืดหยุ่น	ยืดหยุ่นน้อยกว่าและใช้ได้กับข้อมูลที่ไม่มีฤดูกาล	ยืดหยุ่นมากกว่าและรองรับข้อมูลที่มีฤดูกาลที่ชัดเจน
ตัวอย่างการใช้งาน	การพยากรณ์ยอดขายที่มีแนวโน้มทั่วไปแต่ไม่มีฤดูกาลที่ชัดเจน	การพยากรณ์ยอดขายในเทศกาลหรือข้อมูลอนุกรมเวลาที่เปลี่ยนแปลงตามฤดูกาล

แบบจำลอง SARIMA ถูกออกแบบมาเพื่อการวิเคราะห์และพยากรณ์ข้อมูลอนุกรมเวลาประเภทตัวแปรเดียวที่มีความเป็นฤดูกาล ไม่ว่าจะเป็นราคาหุ้น ยอดขายรายเดือน หรือข้อมูลเศรษฐกิจอื่น ๆ เช่น ยอดขายที่เพิ่มขึ้นตามเทศกาลหรือจำนวนนักท่องเที่ยวในช่วงฤดูกาลท่องเที่ยว นอกจากนี้ยังสามารถจัดการข้อมูลที่ไม่อยู่ในสภาวะนิ่งได้อีกด้วย อย่างไรก็ตามแบบจำลอง SARIMA มีข้อจำกัดคือ 1) การเลือกพารามิเตอร์ที่เหมาะสมต้องใช้เวลาและความรู้ในการวิเคราะห์ 2) ไม่สามารถจัดการกับข้อมูลที่สูญหายได้โดยตรง 3) หากข้อมูลมีปัจจัยหลายตัวที่มีผลต่อกันอาจต้องพิจารณาใช้แบบจำลองอื่นที่สามารถจัดการกับตัวแปรหลายตัวได้ เช่น SARIMAX 4) ไม่สามารถจับความเปลี่ยนแปลงของข้อมูลที่มีแนวโน้มซับซ้อนหรือมีการเปลี่ยนแปลงโครงสร้างอย่างฉับพลัน และ 5) เมื่อระยะเวลาพยากรณ์นานขึ้นความแม่นยำของการพยากรณ์อาจลดลง

งานวิจัยที่เกี่ยวข้อง

Selvakumar et al. (2024) นำเสนอการประยุกต์ใช้แบบจำลอง SARIMAX ในการพยากรณ์ยอดขาย โดยอาศัยการวิเคราะห์ข้อมูลอนุกรมเวลา ซึ่งบูรณาการข้อมูลยอดขายในอดีตเข้ากับปัจจัยภายนอกที่เกี่ยวข้อง อาทิ ฤดูกาล แคมเปญโฆษณา และสถานะทางเศรษฐกิจ กระบวนการพัฒนาแบบจำลองประกอบด้วย 5 ขั้นตอน สำคัญ ได้แก่ การเก็บรวบรวมข้อมูล การเตรียมข้อมูล การคัดเลือกคุณลักษณะที่เหมาะสม การทดสอบแบบจำลอง และการแสดงผลลัพธ์ ผลการศึกษาพบว่าแบบจำลองที่พัฒนาขึ้นสามารถเพิ่มความแม่นยำในการพยากรณ์ยอดขายได้อย่างมีนัยสำคัญ ส่งผลให้ผู้บริหารสามารถวางแผนการผลิตและการกระจายสินค้าได้อย่างมีประสิทธิภาพยิ่งขึ้น รวมถึงเอื้อต่อการลดต้นทุนการดำเนินงานและเสริมสร้างศักยภาพในการทำกำไรขององค์กร

Sirisha et al. (2022) ศึกษาเปรียบเทียบประสิทธิภาพของแบบจำลอง ARIMA, SARIMA และ LSTM ในการพยากรณ์กำไร โดยใช้ข้อมูลการขายย้อนหลัง 46 ปี พบว่าแบบจำลอง LSTM มีความแม่นยำสูงสุดที่ร้อยละ 97.01 รองลงมาคือแบบจำลอง SARIMA ที่ร้อยละ 94.38 และแบบจำลอง ARIMA ที่ร้อยละ 93.84 ตามลำดับ เมื่อพิจารณาลักษณะเฉพาะของแต่ละแบบจำลอง พบว่าแบบจำลอง ARIMA มีความเหมาะสมสำหรับข้อมูลที่มีความซับซ้อนต่ำและสะดวกในการใช้งาน แต่มีข้อจำกัดในการจัดการกับรูปแบบความผันผวนตามฤดูกาล ในขณะที่แบบจำลอง SARIMA มีความเหมาะสมสำหรับข้อมูลที่มีลักษณะตามฤดูกาล แต่มีความซับซ้อนในกระบวนการคัดเลือกพารามิเตอร์ที่เหมาะสม สำหรับแบบจำลอง LSTM นั้น แม้จะมีความสามารถในการจัดการข้อมูลที่มีความซับซ้อนสูงได้อย่างมีประสิทธิภาพ แต่ต้องใช้ทรัพยากรการประมวลผลและระยะเวลาในการฝึกสอนแบบจำลองสูงกว่าแบบจำลองอื่น ตารางที่ 2 แสดงผลการเปรียบเทียบแบบจำลอง ARIMA, SARIMA และ LSTM

ตารางที่ 2 เปรียบเทียบแบบจำลอง ARIMA, SARIMA และ LSTM

แบบจำลอง	จุดเด่น	ข้อจำกัด	การใช้งานที่เหมาะสม
ARIMA	ใช้งานง่าย เหมาะกับข้อมูลที่ซับซ้อนต่ำ	ไม่รองรับฤดูกาลได้ดี	ข้อมูลเชิงเวลาแบบเรียบง่าย ไม่มีฤดูกาล
SARIMA	จัดการข้อมูลที่มีฤดูกาลได้ดี	การเลือกพารามิเตอร์ซับซ้อน	ข้อมูลที่มีความผันผวนตามฤดูกาล เช่น ยอดขายรายเดือน
LSTM	รองรับข้อมูลซับซ้อนสูง ประสิทธิภาพสูงสุด	ต้องใช้ทรัพยากรและเวลาฝึกสูง	ข้อมูลขนาดใหญ่และซับซ้อน มีความสัมพันธ์เชิงลึก

Szostek et al. (2024) เปรียบเทียบประสิทธิภาพของแบบจำลอง ARIMA, SARIMA และ Support Vector Regression (SVR) ในการพยากรณ์การผลิตพลังงานรายเดือน โดยใช้ชุดข้อมูลจากฟาร์มกังหันลม 2 แห่งในประเทศโปแลนด์ ได้แก่ Gizałki และ Łęki Dukielskie ด้วยวิธีการค้นหาแบบกริด (Grid Search) ในการปรับพารามิเตอร์ของแบบจำลอง และประเมินผลด้วยเกณฑ์วัด MAPE, MBE, MAE และ R² Score

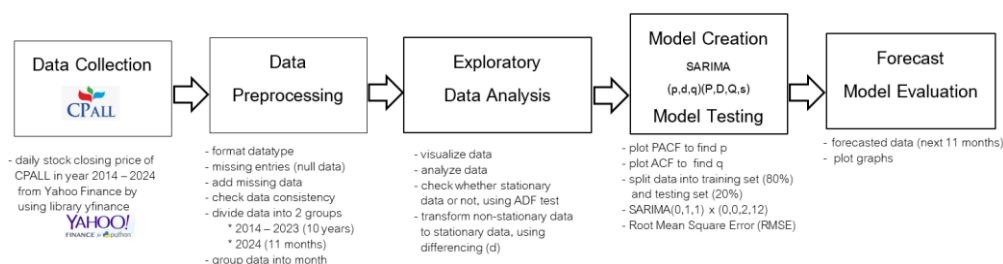
ผลการศึกษาพบว่าแบบจำลอง ARIMA มีความแม่นยำสูงสุดสำหรับข้อมูลจาก Gizatki ในขณะที่แบบจำลอง SARIMA มีความเหมาะสมสูงสุดสำหรับข้อมูลจาก Łęki Dukielskie ซึ่งสะท้อนให้เห็นถึงอิทธิพลของปัจจัยตามฤดูกาลที่มีผลต่อประสิทธิภาพของแบบจำลอง ส่วนแบบจำลอง SVR มีความแม่นยำต่ำสุดในทั้งสองชุดข้อมูล ผลการวิจัยดังกล่าวยืนยันว่าแบบจำลอง ARIMA และ SARIMA เป็นเครื่องมือพยากรณ์ที่มีประสิทธิภาพ อย่างไรก็ตาม ความเหมาะสมของแบบจำลองขึ้นอยู่กับลักษณะเฉพาะของข้อมูลและรูปแบบความผันผวนตามฤดูกาลของแต่ละชุดข้อมูล

Handayani et al. (2020) นำแบบจำลอง SARIMA (21,2,0)(1,0,0,12) มาประยุกต์ใช้ในการพยากรณ์ยอดขายปลาตุก โดยได้ค่า RMSE เท่ากับ 88.72 และนำไปใช้บูรณาการกับระบบออนไลน์ SITRAN เพื่อสนับสนุนการดำเนินธุรกิจ อย่างไรก็ตาม แบบจำลองดังกล่าวมีความซับซ้อนสูงและมีความเสี่ยงต่อปัญหาการปรับแบบจำลองให้เหมาะสมกับข้อมูลฝึกสอนมากเกินไป (Overfitting)

4. วิธีดำเนินการวิจัย

ขั้นตอนการทำงานของพยากรณ์ราคาหุ้นโดยใช้การวิเคราะห์อนุกรมเวลา SARIMA ประกอบด้วย 5 ขั้นตอน (ภาพที่ 2) ดังนี้

1. เก็บรวบรวมข้อมูล (Data Collection)
2. จัดรูปแบบและทำความสะอาดข้อมูล (Data Preprocessing)
3. วิเคราะห์ชุดข้อมูลก่อนการสร้างแบบจำลอง (Exploratory Data Analysis: EDA)
4. สร้างและทดสอบแบบจำลอง SARIMA (Model Creation and Model Testing)
5. พยากรณ์และประเมินแบบจำลอง (Forecast and Model Evaluation)



ภาพที่ 2 ขั้นตอนการทำงานของพยากรณ์ราคาหุ้นโดยใช้การวิเคราะห์อนุกรมเวลา SARIMA

ขั้นตอนที่ 1: เก็บรวบรวมข้อมูล

งานวิจัยนี้ใช้ชุดข้อมูลรายวันของหุ้นบริษัท ซีพี ออลล์ จำกัด (มหาชน) เป็นกรณีศึกษา ครอบคลุมช่วงระยะเวลาดังแต่วันที่ 1 มกราคม พ.ศ. 2557 ถึงวันที่ 30 พฤศจิกายน พ.ศ. 2567

ขั้นตอนที่ 2: จัดรูปแบบและทำความสะอาดข้อมูล

จากชุดข้อมูลให้ทำการเลือกเฉพาะข้อมูลราคาปิด (Close) มาใช้ในการวิเคราะห์ จากนั้นทำการตรวจสอบและกำจัดค่าที่ผิดพลาด รายการ Null รายการซ้ำกัน ความไม่สอดคล้องกันของรายการข้อมูล เติมค่าข้อมูลสำหรับวันที่ขาดหายไป เช่น วันเสาร์ วันอาทิตย์ หรือวันหยุดที่ตลาดไม่เปิดทำการ หลังจากนั้นทำการแปลงข้อมูลราคาปิดรายวันให้เป็นรายเดือนโดยคำนวณค่าเฉลี่ยของราคาปิดในแต่ละเดือน

งานวิจัยนี้เน้นการพยากรณ์เพื่อวางแผนการลงทุนระยะยาว จึงเลือกใช้ข้อมูลรายเดือนเพราะช่วยลดสัญญาณรบกวน (Noise) ที่มีในข้อมูลรายวันและทำให้เห็นแนวโน้มของราคาหุ้นชัดเจนขึ้น ซึ่งเหมาะสมกับแบบจำลอง SARIMA ที่ใช้ฤดูกาล 12 เดือน ($s=12$) นอกจากนี้ยังช่วยลดภาวะการคำนวณและลดปัญหาการแตกตัวของโมเดล ทำให้ SARIMA สามารถค้นหาพารามิเตอร์ได้อย่างมีประสิทธิภาพมากขึ้น อย่างไรก็ตาม หากต้องการลงทุนระยะสั้นหรือติดตามความผันผวนของตลาด ควรพิจารณาใช้ข้อมูลรายวันแทน

จากนั้นทำการแบ่งข้อมูลออกเป็น 2 ส่วนคือ

- ข้อมูลวันที่ 1 มกราคม 2557 ถึงวันที่ 31 ธันวาคม 2566 เพื่อใช้สร้างแบบจำลอง (120 เดือน)
- ข้อมูลวันที่ 1 มกราคม 2567 ถึงวันที่ 30 พฤศจิกายน 2567 เพื่อใช้เปรียบเทียบผลการพยากรณ์ (11 เดือน)

ขั้นตอนที่ 3: วิเคราะห์ชุดข้อมูลก่อนการสร้างแบบจำลอง

1. สร้างกราฟแสดงข้อมูลราคาปิดรายวันและรายเดือนเพื่อดูแนวโน้มและฤดูกาล
2. แยกองค์ประกอบของข้อมูลออกเป็น 4 ส่วน คือ ข้อมูลที่สังเกตได้ แนวโน้ม ฤดูกาล และส่วนที่ไม่สามารถอธิบายได้ เพื่อช่วยให้เข้าใจลักษณะของข้อมูลในแต่ละส่วน เช่น แนวโน้มหลักของข้อมูลที่เกิดขึ้นวัฏจักรฤดูกาลที่มีรูปแบบซ้ำ ๆ และส่วนที่ไม่สามารถอธิบายได้
3. ทดสอบข้อมูลคงที่ด้วยการทดสอบ Augmented Dickey-Fuller เพื่อทดสอบว่าชุดข้อมูลคงที่หรือไม่ โดยใช้ค่า p-value ในการพิจารณา
 - ถ้าค่า $p\text{-value} \leq 0.05$ แสดงว่าข้อมูลมีแนวโน้มว่าจะเป็นข้อมูลคงที่
 - ถ้าค่า $p\text{-value} > 0.05$ แสดงว่าข้อมูลมีแนวโน้มว่าจะเป็นข้อมูลไม่คงที่ นั่นคือข้อมูลอาจมีแนวโน้มหรือฤดูกาล

4. แสดงกราฟ ACF (Auto Correlation) และกราฟ PACF (Partial Auto Correlation) เพื่อช่วยในการกำหนดค่าพารามิเตอร์ p และ q

ขั้นตอนที่ 4: สร้างและทดสอบแบบจำลอง SARIMA ดังนี้

1. ค้นหาพารามิเตอร์ที่เหมาะสมสำหรับแบบจำลอง
2. แบ่งข้อมูลเป็นชุดการฝึกและทดสอบ
3. คำนวณค่า Root Mean Square Error (RMSE)
4. ใช้แบบจำลองที่ปรับแต่งแล้วในการพยากรณ์ค่าราคาปิดรายเดือนล่วงหน้า 11 เดือน

ขั้นตอนที่ 5: สรุปผลการประเมินแบบจำลอง ดังนี้

1. แสดงกราฟของข้อมูลจริงและค่าที่พยากรณ์ เพื่อแสดงผลการเปรียบเทียบระหว่างชุดข้อมูลสำหรับการฝึก (Training Data) ชุดข้อมูลสำหรับทดสอบ (Test Data) ชุดข้อมูลจริง (Actual Data) และผลลัพธ์การพยากรณ์ (Forecasted Data) ในกราฟเดียวกัน เพื่อให้เห็นภาพรวมของข้อมูลแต่ละชุด และเปรียบเทียบระหว่างผลลัพธ์ที่พยากรณ์กับข้อมูลที่เกิดขึ้นจริง โดยมีการเพิ่มช่วงความเชื่อมั่นเพื่อแสดงถึงความไม่แน่นอนของผลการพยากรณ์ด้วย

2. การแสดงกราฟค่าความคลาดเคลื่อน (Residuals) ของแบบจำลอง ซึ่งเป็นเครื่องมือที่สำคัญในการวิเคราะห์ความแม่นยำของการพยากรณ์ของแบบจำลอง

5. การพัฒนาและอภิปรายผลการวิจัย

เครื่องมือที่ใช้ในการพัฒนาคือ Google Colab (Google Colab, 2024) โดยใช้ภาษา Python และใช้ Microsoft Power BI (Microsoft Power BI, 2024) เพื่อสร้างกราฟรายวันที่สามารถโต้ตอบได้ การพัฒนาเป็นไปตามลำดับขั้นตอนดังต่อไปนี้

1) ติดตั้งไลบรารี pmdarima เพื่อเรียกใช้ auto_arima สำหรับการเลือกพารามิเตอร์ที่เหมาะสมในแบบจำลอง SARIMA ทำให้กระบวนการสร้างแบบจำลองมีความง่ายและรวดเร็ว

2) เก็บรวบรวมข้อมูล โดยดึงข้อมูลราคาปิดรายวันของหุ้น CPALL จาก Yahoo Finance โดยใช้ไลบรารี yfinance แบ่งออกเป็น 2 ชุด ได้แก่ ชุดที่ 1 คือข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2557 ถึงวันที่ 31 ธันวาคม พ.ศ. 2566 สำหรับใช้ในการสร้างแบบจำลอง โดยจัดเก็บข้อมูลในรูปแบบ DataFrame ชื่อ df ดังแสดงในภาพที่ 3 และภาพที่ 4(ก) และชุดที่ 2 คือข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2567 ถึงวันที่ 30 พฤศจิกายน พ.ศ. 2567 สำหรับใช้เปรียบเทียบผลการพยากรณ์

```
# API Import data
import yfinance as yf
import matplotlib.pyplot as plt
import pandas as pd

stockname = 'CPALL.BK'

# Fetch data using yfinance
df = yf.download(stockname, start='2014-01-01', end='2023-12-31') ## yyyy-mm-dd
df= df[['Close']]
```

ภาพที่ 3 คำสั่งการดึงข้อมูลรายวันราคาปิดหุ้น CPALL

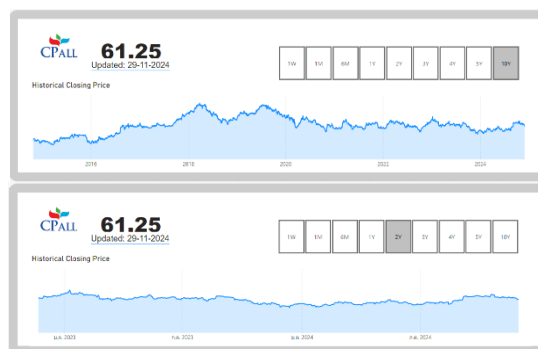
Close	Close
Date	Date
2014-01-04 00:00:00+00:00	40.00
2014-01-05 00:00:00+00:00	40.00
2014-01-06 00:00:00+00:00	40.25
2014-01-07 00:00:00+00:00	41.00
2014-01-08 00:00:00+00:00	40.50
...	...
2023-12-24 00:00:00+00:00	55.50
2023-12-25 00:00:00+00:00	55.25
2023-12-26 00:00:00+00:00	56.00
2023-12-27 00:00:00+00:00	55.50
2023-12-28 00:00:00+00:00	56.00
2014-01-31 00:00:00+00:00	39.946429
2014-02-28 00:00:00+00:00	38.937500
2014-03-31 00:00:00+00:00	41.903226
2014-04-30 00:00:00+00:00	43.950000
2014-05-31 00:00:00+00:00	43.524194
...	...
2023-08-31 00:00:00+00:00	62.701613
2023-09-30 00:00:00+00:00	62.958333
2023-10-31 00:00:00+00:00	58.024194
2023-11-30 00:00:00+00:00	55.641667
2023-12-31 00:00:00+00:00	53.937500

(ก) ข้อมูลรายวัน

(ข) ข้อมูลรายเดือน

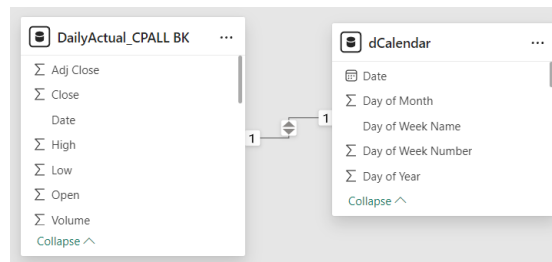
ภาพที่ 4 ข้อมูลรายวันและรายเดือนของราคาปิดหุ้น CPALL

3) จัดรูปแบบชุดข้อมูล ทำความสะอาดข้อมูล และสร้างข้อมูลรายเดือนจากค่าเฉลี่ยของแต่ละเดือน
ภาพที่ 4(ข) แสดงตัวอย่างข้อมูลรายเดือน



ภาพที่ 5 กราฟแสดงข้อมูลราคาปิดรายวันระยะเวลา 10 ปี และ 2 ปี

4) วิเคราะห์ชุดข้อมูลก่อนการสร้างแบบจำลองโดยการสร้างกราฟที่สามารถโต้ตอบได้จากข้อมูลราคาปิดรายวัน โดยใช้ Microsoft Power BI (ภาพที่ 5) ความสัมพันธ์ระหว่างตาราง dCalendar และตารางราคาหุ้น แสดงในภาพที่ 6 และภาพที่ 7 แสดง DAX (Data Analysis Expressions) หรือคำสั่งในการสร้างตาราง dCalendar และคำสั่งการควบคุมปุ่มเพื่อแสดงข้อมูลตามระยะเวลาที่กดเลือกปุ่มในกราฟ เช่น หนึ่งอาทิตย์ หนึ่งเดือน หนึ่งปี สองปี สามปี สี่ปี ห้าปี หรือสิบปี



ภาพที่ 6 Data Model แสดงความสัมพันธ์ระหว่างตาราง dCalendar และตารางราคาหุ้น

(ก) DAX ในการสร้างตาราง dCalendar

```
dCalendar =
ADDCOLUMNS (
CALENDAR(MIN('DailyActual_CPALL BK'[Date]), MAX('DailyActual_CPALL BK'[Date])),
"Year", YEAR ([Date] ),
"Year Thai", YEAR ([Date] ) + 543,
"Month Number", MONTH ([Date] ),
"Month Name short", FORMAT ([Date], "MMM", "en-US" ),
"Month Name", FORMAT ([Date], "MMMM", "en-US" ),
"Month Name Thai short", FORMAT ([Date], "MMM", "th-TH" ),
"Month Name Thai", FORMAT ([Date], "MMMM", "th-TH" ),
"Quarter", QUARTER ([Date] ),
"Day of Week Number", WEEKDAY ([Date] ),
"Day of Week Name", FORMAT ([Date], "dddd" ),
"Day of Month", DAY ([Date] ),
"Week of Year", WEEKNUM ([Date], 2 ),
"Year Month", FORMAT ([Date], "YYYY-MM" ),
"Year Quarter", YEAR ([Date] ) & "-Q" & FORMAT ([Date], "Q" ),
"Year Week", YEAR ([Date] ) & "-W" & WEEKNUM ([Date], 2 ),

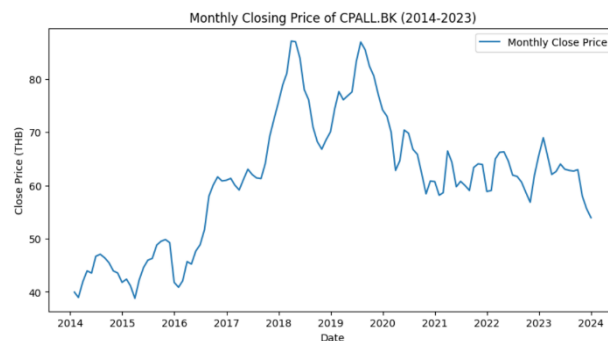
```

(ข) DAX ควบคุมปุ่มเพื่อแสดงข้อมูลตามระยะเวลา
ที่เลือกในกราฟ

```
Maximum Date =
IF(
ISBLANK(MAX('DailyActual_CPALL BK'[Date])),
TODAY(), -- Default to today if no dates are found
MAX('DailyActual_CPALL BK'[Date])
)

Minimum Date =
VAR _MaxDate = [Maximum Date]
VAR _SelectedPeriod = SELECTEDVALUE(Period[Period], "1M") -- Default to "1M" if nothing is :
VAR _MinDate =
SWITCH(
TRUE(),
_SelectedPeriod = "1M", _MaxDate - 7,
_SelectedPeriod = "1M", EDATE(_MaxDate, -1),
_SelectedPeriod = "6M", EDATE(_MaxDate, -6),
_SelectedPeriod = "1Y", EDATE(_MaxDate, -12),
_SelectedPeriod = "2Y", EDATE(_MaxDate, -24),
_SelectedPeriod = "3Y", EDATE(_MaxDate, -36),
_SelectedPeriod = "4Y", EDATE(_MaxDate, -48),
_SelectedPeriod = "5Y", EDATE(_MaxDate, -60),
_SelectedPeriod = "10Y", EDATE(_MaxDate, -120),
_MaxDate -- Default to the maximum date if no period matches
)
RETURN _MinDate
```

ภาพที่ 7 ตัวอย่าง DAX ในการสร้างตารางและการควบคุมปุ่ม



ภาพที่ 8 กราฟแสดงข้อมูลราคาปิดรายเดือนเพื่อดูแนวโน้มและฤดูกาลของราคาหุ้น CPALL

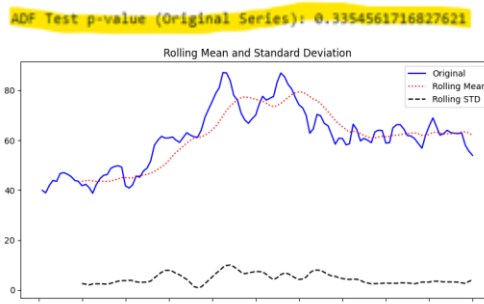
กราฟข้อมูลราคาปิดรายเดือนเพื่อดูภาพรวมของแนวโน้มและฤดูกาลแสดงในภาพที่ 8 และภาพที่ 9 แสดงการแยกองค์ประกอบของข้อมูลออกเป็น 4 ส่วน คือ ข้อมูลที่สังเกตได้ (Observed) แนวโน้ม (Trend) ฤดูกาล (Seasonal) และส่วนที่ไม่สามารถอธิบายได้ (Residual) จากผลการศึกษาพบว่าข้อมูลมีลักษณะไม่คงที่มี

แนวโน้ม วัฏจักร และฤดูกาลที่มีรูปแบบซ้ำ ๆ และมีส่วนที่ไม่สามารถอธิบายได้ จะเห็นว่าหุ่น CPALL มี วัฏจักร ฤดูกาลที่ชัดเจนและเกิดซ้ำทุก 12 เดือน โดย Seasonal Component ซึ่งเป็นส่วนประกอบของอนุกรมเวลาที่สะท้อนการเปลี่ยนแปลงตามฤดูกาล มีลักษณะเป็นคลื่นสม่ำเสมอ สอดคล้องกับธรรมชาติของธุรกิจค้าปลีก (7-Eleven) และพฤติกรรมผู้บริโภคตามเทศกาล จุดสูงสุดมักเกิดในช่วงต้นถึงกลางปี (ไตรมาส 1-2) ขณะที่จุดต่ำสุดมักเกิดในช่วงปลายปี (ไตรมาส 4) รูปแบบนี้สะท้อนปัจจัยเชิงธุรกิจ เช่น การจับจ่ายตามเทศกาลและผลประกอบการรายไตรมาส จึงสามารถใช้เป็นแนวทางในการวางกลยุทธ์การลงทุนและประเมินความเสี่ยง อย่างไรก็ตาม แม้จะยืนยันวัฏจักรนี้ได้ แต่ยังพบ Residuals ที่แบบจำลองไม่สามารถอธิบายได้ ซึ่งอาจเกิดจากปัจจัยภายนอกหรือเหตุการณ์เศรษฐกิจที่ไม่คาดคิด



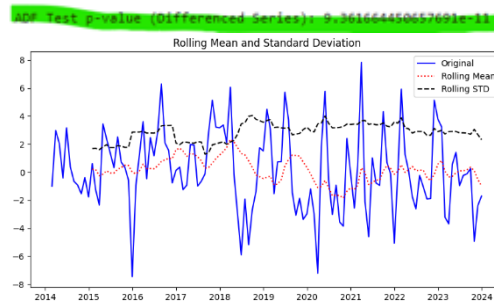
ภาพที่ 9 กราฟแสดงข้อมูล แนวโน้ม ฤดูกาลและส่วนที่ไม่สามารถอธิบายได้

ภาพที่ 10 แสดงผลการทดสอบ ADF เพื่อตรวจสอบความคงที่ของข้อมูล จากผลการทดสอบพบว่า p-value มีค่าเท่ากับ 0.335456 ซึ่งมากกว่า 0.05 บ่งชี้ว่าข้อมูลดังกล่าวไม่เป็นข้อมูลคงที่ นอกจากนี้กราฟที่แสดงข้อมูลยังยืนยันว่าข้อมูลไม่คงที่ เนื่องจากค่าเฉลี่ยและค่าความแปรปรวนของข้อมูลมีการเปลี่ยนแปลงไปตามเวลา ดังนั้นจำเป็นต้องทำการปรับข้อมูลโดยใช้วิธี Differencing ในลำดับที่หนึ่ง ($d = 1$) เพื่อให้ข้อมูลมีลักษณะเป็นข้อมูลคงที่ก่อนนำไปวิเคราะห์เพิ่มเติม

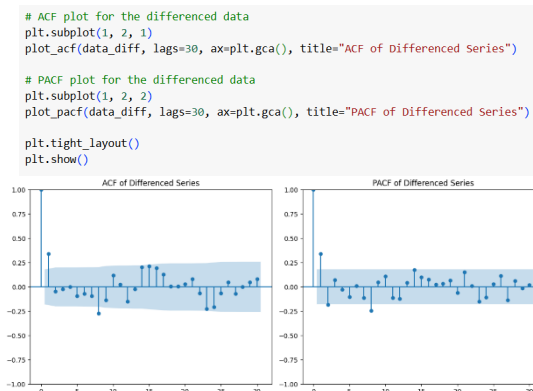


ภาพที่ 10 ค่า p-value และกราฟก่อนการทำ Differencing

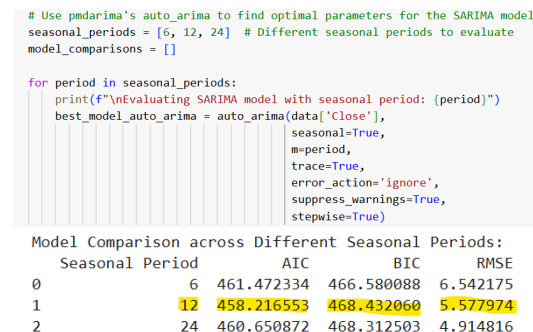
จากนั้นได้ดำเนินการทดสอบ ADF ซ้ำอีกครั้ง โดยค่า p-value ที่ได้ใหม่เท่ากับ $9.361664e-11$ ซึ่งมีค่าน้อยกว่า 0.05 แสดงให้เห็นว่าข้อมูลได้รับการปรับให้อยู่ในรูปแบบของข้อมูลคงที่แล้ว ทั้งนี้ ภาพที่ 11 แสดงค่า p-value หลังการดำเนินการ Differencing และแสดงกราฟของข้อมูลคงที่ โดยสังเกตได้จากกราฟที่ได้มีค่าเฉลี่ยและค่าความแปรปรวนค่อนข้างคงที่ สำหรับกราฟในภาพที่ 12 ชี้ให้เห็นว่าค่าพารามิเตอร์ q (จาก ACF) มีค่าเท่ากับ 1 และค่าพารามิเตอร์ p (จาก PACF) มีค่าเท่ากับ 1



ภาพที่ 11 ค่า p-value และกราฟหลังการทำ Differencing 1 ครั้ง



ภาพที่ 12 ACF และ PACF



ภาพที่ 13 ค่าพารามิเตอร์ $s = 12$ คือค่าที่เหมาะสมที่สุด

```
best_period = comparison_df.loc[comparison_df['AIC'].idxmin(), 'Seasonal Period']
print(f"\nfitting SARIMA Model for Best Seasonal Period: {best_period}")

best_model_auto_arima = auto_arima(data['Close'],
                                    seasonal=True,
                                    m=int(best_period), #best_period = 12
                                    trace=True,
                                    error_action='ignore',
                                    suppress_warnings=True,
                                    stepwise=True) # Use stepwise to reduce search time

sarima_result_finetuned = best_model_auto_arima.fit(data['Close'])
```

Best model: ARIMA(0,1,1)(0,0,2)[12]

ภาพที่ 14 การสร้างแบบจำลอง SARIMA

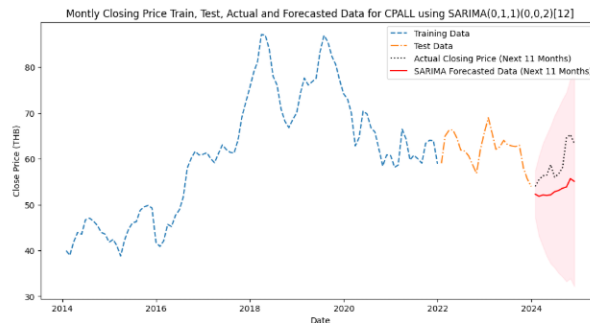
ก่อนการสร้างแบบจำลอง ข้อมูลตั้งแต่วันที่ 1 มกราคม พ.ศ. 2557 ถึงวันที่ 31 ธันวาคม พ.ศ. 2566 ได้ถูกแบ่งออกเป็น 2 ชุด ได้แก่ ชุดข้อมูลสำหรับการฝึก (Training Set) จำนวน 80% ของข้อมูลทั้งหมด (ข้อมูลตั้งแต่แถวที่ 0 ถึงแถวที่มีค่าดัชนีเท่ากับ train_size) และชุดข้อมูลสำหรับการทดสอบ (Test Set) จำนวน 20% ของข้อมูลที่เหลือ (ข้อมูลตั้งแต่แถวที่ train_size ไปจนถึงสิ้นสุดชุดข้อมูล) เพื่อใช้สำหรับการฝึกและทดสอบแบบจำลอง ในกระบวนการนี้ได้มีการเรียกใช้ฟังก์ชัน auto_arima แบบ Stepwise เพื่อค้นหาค่าพารามิเตอร์ที่เหมาะสมที่สุดสำหรับแบบจำลอง SARIMA ได้แก่ ค่าพารามิเตอร์ p, d, q, P, D, Q และ s โดยได้ทำการทดลองกับค่าช่วงฤดูกาลจำนวน 3 ค่า ได้แก่ s = 6, s = 12 และ s = 24 ผลการทดลองพบว่าค่าพารามิเตอร์ที่เหมาะสมที่สุดคือ (0,1,1)(0,0,2,12) ดังแสดงในภาพที่ 13 และ 14 จากนั้นทำการพยากรณ์และประเมินความแม่นยำของแบบจำลองพยากรณ์ราคาหุ้นด้วยค่า RMSE โดยเปรียบเทียบค่าพยากรณ์กับข้อมูลจริง ค่า RMSE คือรากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ค่าต่ำหมายถึงแบบจำลองแม่นยำกว่า แต่ไวต่อค่าผิดปกติเนื่องจากการยกกำลังสอง แสดงดังสมการต่อไปนี้

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

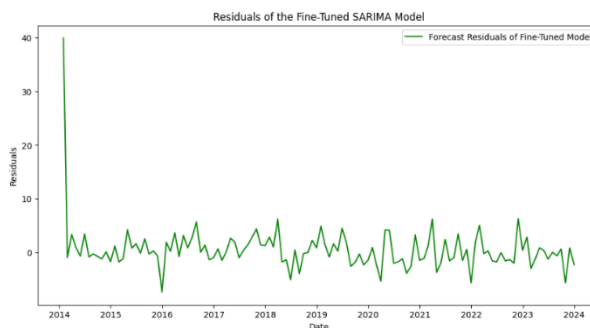
ภาพที่ 15 เป็นคำสั่งการพยากรณ์ข้อมูลในอนาคตในช่วงเวลา 11 เดือนโดยใช้แบบจำลอง SARIMA(0,1,1)(0,0,2,12) และภาพที่ 16 แสดงกราฟจำนวน 4 กราฟบนแกนเดียวกัน คือชุดข้อมูลสำหรับการฝึก (Training Data) เส้นประสีฟ้า ชุดข้อมูลสำหรับทดสอบ (Test Data) เส้นประจุดสีเหลือง ชุดข้อมูลจริง (Actual Closing Price (Next 11 Months)) เส้นจุดสีดำ และผลลัพธ์การพยากรณ์ (Forecasted Data (Next 11 Months)) ด้วยแบบจำลอง SARIMA โดยเป็นการพยากรณ์ค่าราคาปิดของหุ้น CPALL รายเดือนล่วงหน้า 11 เดือน (เดือนมกราคมถึงเดือนพฤศจิกายน 2567) เส้นทึบสีแดง จะเห็นว่าผลการเปรียบเทียบระหว่างชุดข้อมูลจริงและผลลัพธ์การพยากรณ์มีแนวโน้มไปในทางเดียวกันแม้ว่าจะได้ค่าไม่เท่ากัน แบบจำลอง SARIMA(0,1,1)(0,0,2,12) มีค่า RMSE = 6.226

```
forecast_steps = 11
forecast_finertuned = sarima_result_finertuned.predict(n_periods=forecast_steps)
forecast_index_finertuned = pd.date_range(start=data.index[-1] +
pd.offsets.MonthEnd(1), periods=forecast_steps, freq='M')
forecast_series_finertuned = pd.Series(forecast_finertuned, index=forecast_index_finertuned)
```

ภาพที่ 15 การพยากรณ์ข้อมูลในอนาคต 11 เดือนด้วยแบบจำลอง SARIMA

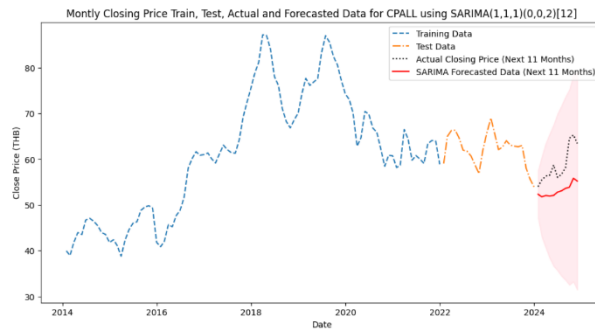


ภาพที่ 16 กราฟของข้อมูลจริงและค่าที่พยากรณ์จากแบบจำลอง SARIMA(0,1,1)(0,0,2,12)



ภาพที่ 17 กราฟความคลาดเคลื่อนของแบบจำลองที่ปรับแต่งแล้ว

ภาพที่ 17 แสดงกราฟความคลาดเคลื่อนของแบบจำลอง SARIMA ที่ปรับแต่ง โดยแกน X แสดงเวลา (Date) และแกน Y แสดงค่าความคลาดเคลื่อน (Residuals) ระหว่างค่าจริงกับค่าพยากรณ์ของแบบจำลอง มีลักษณะการกระจายของค่าความคลาดเคลื่อนที่ค่อนข้างสุ่มรอบเส้นค่าศูนย์และไม่มีรูปแบบชัดเจนหลังปี 2014 ซึ่งแสดงถึงประสิทธิภาพของแบบจำลองที่เหมาะสมกับข้อมูล อย่างไรก็ตามจากกราฟ PACF และกราฟ ACF ในภาพที่ 12 พบว่าค่าพารามิเตอร์ ค่า $p=1$ ค่า $q=1$ ซึ่งทำให้ได้แบบจำลอง SARIMA(1,1,1)(0,0,2,12) ผู้วิจัยได้ดำเนินการทดสอบแบบจำลอง และผลลัพธ์ที่ได้แสดงในภาพที่ 18 พบว่าแบบจำลองดังกล่าวมีความคลาดเคลื่อนรากที่สองของค่าเฉลี่ยกำลังสองเท่ากับ 6.247 ซึ่งสูงกว่าค่า RMSE ที่ได้จากแบบจำลอง SARIMA(0,1,1)(0,0,2,12) ทั้งนี้ ตารางที่ 3 แสดงผลการเปรียบเทียบค่า RMSE ของแบบจำลองทั้ง 2 แบบ



ภาพที่ 18 กราฟของข้อมูลจริงและค่าที่พยากรณ์จากแบบจำลอง SARIMA(1,1,1)(0,0,2,12)

ตารางที่ 3 เปรียบเทียบค่า RMSE ของแบบจำลอง

แบบจำลอง	พารามิเตอร์	RMSE
SARIMA(0,1,1)(0,0,2,12)	(p=0, d=1, q=1) (P=0, D=0, Q=2, s=12)	6.226
SARIMA(1,1,1)(0,0,2,12)	(p=1, d=1, q=1) (P=0, D=0, Q=2, s=12)	6.247

แบบจำลอง SARIMA มีข้อจำกัดที่สำคัญหลายประการ ได้แก่ ความไวต่อความไม่คงที่ของข้อมูล ความยากในการจัดการกับข้อมูลสูญหายและค่าผิดปกติซึ่งพบบ่อยในข้อมูลการเงิน และข้อจำกัดในการจับความสัมพันธ์เชิงเส้นและรองรับฤดูกาลเพียงรูปแบบเดียว จึงไม่สามารถจับรูปแบบความผันผวนแบบไม่เป็นเชิงเส้นหรือข้อมูลที่มีหลายฤดูกาลได้ดี นอกจากนี้ การกำหนดพารามิเตอร์ (p, d, q, P, D, Q) มีความซับซ้อนและเสี่ยงต่อการปรับแบบจำลองให้เหมาะกับข้อมูลมากเกินไป ส่งผลให้ความแม่นยำของการพยากรณ์ลดลงเมื่อช่วงเวลาที่คาดการณ์ยาวขึ้น

แบบจำลอง SARIMA ยังถูกออกแบบมาสำหรับการวิเคราะห์ข้อมูลอนุกรมเวลาที่มีตัวแปรเดียว (Univariate Time Series) และแสดงรูปแบบความผันแปรตามฤดูกาล กล่าวคือแบบจำลอง SARIMA สามารถจัดการและพยากรณ์ข้อมูลที่ขึ้นอยู่กับตัวแปรเพียงตัวเดียว เช่น ราคาหุ้นรายวันหรือยอดขายรายเดือน ถ้าข้อมูลมีความซับซ้อนมากขึ้นและปัจจัยหลายตัวที่ส่งผลต่อกัน เช่น การพยากรณ์ยอดขายที่ได้รับผลกระทบจากฤดูกาล โฆษณา หรือภาวะเศรษฐกิจ แบบจำลอง SARIMA จะไม่สามารถจัดการได้อย่างมีประสิทธิภาพ เนื่องจากแบบจำลองไม่รองรับข้อมูลหลายตัวแปร (Multivariate Time Series) และเมื่อระยะเวลาการพยากรณ์ยาวขึ้น ความแม่นยำของแบบจำลอง อาจลดลงเนื่องจากขอบเขตการคาดการณ์แนวโน้มและฤดูกาลอาจไม่ครอบคลุมข้อมูลในระยะยาว เพื่อขจัดข้อจำกัดของแบบจำลอง SARIMA ในการจัดการข้อมูลที่มีหลายปัจจัยส่งผลต่อกัน

ทางเลือกเพื่อแก้ข้อจำกัดเหล่านี้ ได้แก่ การใช้แบบจำลองที่รองรับข้อมูลหลายมิติ เช่น แบบจำลอง SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) ซึ่งเป็นการพัฒนาแบบจำลอง SARIMA ให้สามารถรวมตัวแปรภายนอก เช่น ตัวชี้วัดเศรษฐกิจหรือเหตุการณ์สำคัญ เข้าในแบบจำลอง หรือการใช้แบบจำลอง LSTM (Long Short-Term Memory) ซึ่งเป็นแบบจำลองเชิงลึก (Deep

Learning) ที่มีศักยภาพในการจับลำดับเวลาและความสัมพันธ์เชิงซับซ้อนระหว่างตัวแปรหลายตัวได้อย่างมีประสิทธิภาพ นอกจากนี้แบบจำลอง LSTM ยังเหมาะสำหรับการจัดการข้อมูลที่ไม่เป็นเชิงเส้นหรือมีโครงสร้างซับซ้อน ทำให้สามารถพยากรณ์ข้อมูลที่มีหลายปัจจัยหรือความซับซ้อนได้ดีขึ้น

6. สรุปผลการวิจัยและข้อเสนอแนะ

บทความนี้นำเสนอการวิเคราะห์และพยากรณ์ราคาหุ้นของบริษัท ซีพี ออลล์ จำกัด (มหาชน) (CPALL) โดยใช้เทคนิคการวิเคราะห์อนุกรมเวลาแบบ SARIMA ซึ่งเป็นเทคนิคที่มีประสิทธิภาพในการทำนายข้อมูลที่มีความผันแปรตามฤดูกาล การศึกษานี้มีขอบเขตการใช้ข้อมูลราคาปิดรายวันของหุ้น CPALL ระหว่างวันที่ 1 มกราคม 2557 ถึง 30 พฤศจิกายน 2567 โดยแบ่งข้อมูลออกเป็นสองส่วนหลัก คือ ช่วงปี 2557-2566 ใช้สำหรับสร้างแบบจำลองพยากรณ์ (80% เป็นข้อมูลฝึกและ 20% เป็นข้อมูลทดสอบ) และ ช่วงปี 2567 ใช้สำหรับเปรียบเทียบข้อมูลทำนาย ขั้นตอนการดำเนินงานประกอบด้วย 1) การเก็บรวบรวมและทำความสะอาดข้อมูลราคาปิดรายวันของหุ้น CPALL 2) การวิเคราะห์ข้อมูลเบื้องต้นเพื่อเข้าใจถึงแนวโน้มและฤดูกาลที่อาจส่งผลต่อการพยากรณ์การสร้างแบบจำลอง 3) การสร้างแบบจำลอง โดยมีการทดลองใช้พารามิเตอร์ต่าง ๆ เพื่อหาค่าที่เหมาะสมที่สุดสำหรับแบบจำลอง ผลการศึกษพบว่า แบบจำลอง SARIMA(0,1,1)(0,0,2,12) มีประสิทธิภาพสูงสุดในการจับภาพแนวโน้มและรูปแบบของราคาหุ้น CPALL งานวิจัยนี้ใช้ภาษา Python และ Google Colab ในการพัฒนาแบบจำลอง รวมถึงใช้ Microsoft Power BI ในการสร้างกราฟรายวันเชิงโต้ตอบเพื่อแสดงผลการวิเคราะห์ในรูปแบบที่ชัดเจนและเข้าใจง่าย การศึกษาในครั้งนี้ไม่เพียงแต่มุ่งเน้นไปที่การพยากรณ์ราคาหุ้นเท่านั้น แต่ยังเป็นการพัฒนาองค์ความรู้ด้านการวิเคราะห์อนุกรมเวลาในบริบทของการเงินและการลงทุน การใช้แบบจำลอง SARIMA สามารถนำไปประยุกต์ใช้กับการวิเคราะห์ข้อมูลอื่นที่มีลักษณะตามฤดูกาล เช่น ยอดขายหรือข้อมูลเศรษฐกิจ ซึ่งสามารถสนับสนุนการตัดสินใจทางธุรกิจได้อย่างมีประสิทธิภาพ งานวิจัยที่จะทำต่อไปคือใช้แบบจำลองอื่น ๆ เช่น แบบจำลอง SARIMAX ที่รวมตัวแปรภายนอก เช่น ภาวะเศรษฐกิจ หรือแบบจำลอง LSTM ในการจัดการข้อมูลหลายมิติและซับซ้อน

7. กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนสนับสนุนค่าใช้จ่ายเพื่อไปเสนอผลงานวิชาการสำหรับนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ประจำปีงบประมาณ 2568

8. เอกสารอ้างอิง

Artley, B. (2022, April 26). *Time series forecasting with ARIMA, SARIMA and SARIMAX*. Towards Data Science. <https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6>

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431.
- Google Colab. (2024). *Google Colaboratory*. <https://colab.research.google.com>
- Handayani, N. R. W. T., Maslim, M., & Mudjihartono, P. (2020). Forecasting of catfish sales by time series using the SARIMA method. *Jurnal Buana Informatika*, 11(2), 84–93.
- Microsoft Power BI. (2024). *Microsoft Power BI*. <https://powerbi.microsoft.com>
- Saadeddin, Z. (2024). *ARIMA for time series forecasting: Complete guide*. DataCamp. <https://www.datacamp.com/tutorial/arima>
- Selvakumar, S., Renugadevi, G., Vinishah, N., & Yashwanth, R. (2024). Sales forecasting based on time series analysis. In *Proceedings of the 2024 International Conference on Science and Technology Engineering and Management (ICSTEM 2024)* (pp. 658–664).
- Sirisha, U. M., Belavagi, M. C., & Attigeri, G. (2022). Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison. *IEEE Access*, 10, 124715–124727.
- Subhash. (2023, February 27). *Augmented Dickey-Fuller test in time-series analysis*. AskPython. <https://www.askpython.com/python/examples/augmented-dickey-fuller-test-time-series>
- Szostek, K., Mazur, D., Draľus, G., & Kuszniar, J. (2024). Analysis of the effectiveness of ARIMA, SARIMA, and SVR models in time series forecasting: A case study of wind farm energy production. *Energies*, 17(19), 4803. <https://doi.org/10.3390/en17194803>
- Tibshirani, R. (2023). *Lecture 1: Characteristics and examples of time series data*. <https://www.stat.berkeley.edu/~ryantibs/timeseries-f23/lectures/characteristics.pdf>