

การเปรียบเทียบแบบจำลองการเรียนรู้ของเครื่องและโครงข่ายประสาทเทียมเชิงลึก สำหรับการพยากรณ์ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย

A Comparison of Machine Learning and Deep Learning Models for Daily SET Index Forecasting: An End-to-End Prediction Pipeline

วิวัฒน์ ขัดทาน^{1*}

Wittawat Kadthan^{1*}

¹ คณะเศรษฐศาสตร์และการลงทุน มหาวิทยาลัยกรุงเทพ จังหวัดปทุมธานี 12120

¹School of Economics and Investment, Bangkok University, Pathum Thani, 12120

*Corresponding Author: wittawat.k@bu.ac.th

Received 18 กรกฎาคม 2569; Revised 13 สิงหาคม 2569; Accepted 21 สิงหาคม 2569

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบความสามารถในการพยากรณ์ดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทย (SET Index) รายวัน ระหว่างแบบจำลองทางสถิติดั้งเดิม แบบจำลองการเรียนรู้ของเครื่อง (Machine Learning) และแบบจำลองโครงข่ายประสาทเทียมเชิงลึก (Deep Learning) รวมทั้งสิ้น 8 แบบจำลอง ได้แก่ Naive persistence, ARIMA, Random Forest, XGBoost และแบบจำลอง Long Short-Term Memory (LSTM) กับ Gated Recurrent Unit (GRU) ทั้งในรูปแบบถดถอย (regression) และรูปแบบจำแนกประเภท (classification) ข้อมูลที่ใช้เป็นราคาปิดรายวันของดัชนี SET Index ระหว่างวันที่ 4 มกราคม 2553 ถึง 17 กรกฎาคม 2569 รวม 3,945 วันทำการหลังการสร้างตัวแปรคุณลักษณะเชิงเทคนิค (technical indicators) จำนวน 19 ตัวแปร โดยแบ่งข้อมูลตามลำดับเวลาออกเป็นชุดฝึก (พ.ศ. 2553-2564) ชุดตรวจสอบ (พ.ศ. 2565-2566) และชุดทดสอบ (พ.ศ. 2567-2569) ผลการศึกษาพบว่า Random Forest ให้ความแม่นยำเชิงทิศทางสูงสุดที่ร้อยละ 51.8 รองลงมาคือ XGBoost ที่ร้อยละ 50.8 ขณะที่แบบจำลอง LSTM และ GRU ทั้งสี่รูปแบบให้ความแม่นยำเชิงทิศทางอยู่ในช่วงร้อยละ 48.2-49.8 ซึ่งต่ำกว่าหรือใกล้เคียงกับการสุ่มทายและแบบจำลอง naive persistence (ร้อยละ 49.7) โดยมีเพียง Random Forest และ XGBoost เท่านั้นที่ผ่านเกณฑ์เปรียบเทียบพื้นฐานร้อยละ 50 ในเชิงตัวเลข แม้การทดสอบทวินาม (binomial test) จะชี้ว่าความได้เปรียบดังกล่าวยังไม่มีความสำคัญทางสถิติที่ระดับ 0.05 ก็ตาม ในด้านความคลาดเคลื่อนเชิงระดับราคา (RMSE, MAPE) แบบจำลองส่วนใหญ่ให้ค่าใกล้เคียงกันและใกล้เคียงกับ naive persistence ยกเว้น ARIMA ที่ให้ค่าคลาดเคลื่อนสูงกว่าอย่างชัดเจน ผลลัพธ์ดังกล่าวสอดคล้องกับแนวคิดสมมติฐานตลาดที่มีประสิทธิภาพ (Efficient Market Hypothesis) และชี้ให้เห็นว่าความซับซ้อนของแบบจำลองไม่ได้รับประกันความแม่นยำในการพยากรณ์ที่สูงขึ้นเสมอไป ในเชิงปฏิบัติ ผลการศึกษานี้ชี้แนะว่าแบบจำลองต้นไม้ตัดสินใจแบบรวมกลุ่มที่มีโครงสร้างไม่ซับซ้อนควรเป็นทางเลือกตั้งต้นสำหรับงานพยากรณ์ทิศทางดัชนี SET Index ก่อนพิจารณาลงทุนทรัพยากรเชิงคำนวณกับแบบจำลองโครงข่ายประสาทเทียมเชิงลึกที่ซับซ้อนกว่า งานวิจัยนี้มีส่วนสนับสนุนองค์ความรู้โดยนำเสนอกระบวนการพยากรณ์ดัชนี SET Index แบบครบวงจรที่เปรียบเทียบแบบจำลองหลากหลายตระกูลภายใต้เงื่อนไขข้อมูลและการประเมินผลชุดเดียวกัน ซึ่งเป็นประโยชน์ต่อทั้งแวดวงวิชาการและปฏิบัติงานด้านการลงทุนในตลาดหลักทรัพย์ไทย

คำหลัก: การพยากรณ์อนุกรมเวลา; การเรียนรู้ของเครื่อง; โครงข่ายประสาทเทียมเชิงลึก; ดัชนี SET Index; ความแม่นยำเชิงทิศทาง; สมมติฐานตลาดที่มีประสิทธิภาพ

Abstract

This study aims to compare the forecasting performance of traditional statistical models, machine learning models, and deep learning models for daily prediction of the Stock Exchange of Thailand (SET) Index. Eight models were evaluated: naive persistence, ARIMA, Random Forest, XGBoost, and both regression and classification variants of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. Daily closing price data spanning January 4, 2010 to July 17, 2026 were used, yielding 3,945 trading days after constructing 19 technical indicator features. Data were split chronologically into training (2010-2021), validation (2022-2023), and test (2024-2026) sets. Results show that Random Forest achieved the highest directional accuracy at 51.8%, followed by XGBoost at 50.8%, while all four LSTM/GRU variants achieved directional accuracy between 48.2% and 49.8%, at or below the naive persistence benchmark (49.7%) and coin-flip chance. Only Random Forest and XGBoost numerically exceeded the 50% baseline; a binomial test indicated that this advantage was not statistically significant at the 0.05 level. In terms of price-level error (RMSE, MAPE), most models performed comparably to naive persistence, except ARIMA, which showed markedly higher error. These findings are consistent with the Efficient Market Hypothesis and indicate that model complexity does not guarantee superior forecasting accuracy. In practice, these results suggest that simple tree-ensemble models should be the first choice for SET Index directional forecasting before investing computational resources in more complex deep learning architectures. This study contributes an end-to-end forecasting pipeline that benchmarks diverse model families under identical data and evaluation conditions, offering value to both academic researchers and market practitioners in Thailand.

Keywords: Time Series Forecasting; Machine Learning; Deep Learning; SET Index; Directional Accuracy; Efficient Market Hypothesis

1. บทนำ

ตลาดหลักทรัพย์เป็นกลไกสำคัญของระบบเศรษฐกิจสมัยใหม่ในการระดมทุนและสะท้อนความคาดหวังของนักลงทุนต่อสถานะเศรษฐกิจในอนาคต การพยากรณ์ทิศทางและระดับราคาหลักทรัพย์จึงเป็นประเด็นที่ได้รับความสนใจอย่างต่อเนื่องทั้งในแวดวงวิชาการและภาคปฏิบัติ ในช่วงทศวรรษที่ผ่านมาความก้าวหน้าของปัญญาประดิษฐ์ โดยเฉพาะแบบจำลองการเรียนรู้ของเครื่อง (Machine Learning: ML) และโครงข่ายประสาทเทียมเชิงลึก (Deep Learning: DL) ได้ถูกนำมาประยุกต์ใช้ในการพยากรณ์ราคาหลักทรัพย์อย่างแพร่หลายมากขึ้น งานทบทวนวรรณกรรมอย่างเป็นระบบชี้ให้เห็นว่าจำนวนงานวิจัยที่ประยุกต์ใช้ DL ในการพยากรณ์อนุกรมเวลาทางการเงินเพิ่มขึ้นอย่างก้าวกระโดดในช่วงปี 2548 ถึง 2562 แต่ผลลัพธ์ด้านความแม่นยำกลับมีความหลากหลายและไม่สอดคล้องกันในแต่ละการศึกษา [1] สอดคล้องกับงานทบทวนวรรณกรรมเชิง

ปริมาณที่พบว่าแม้จำนวนงานวิจัยด้าน AI/ML ในสาขาการเงินจะเติบโตอย่างรวดเร็ว แต่ยังคงมีข้อถกเถียงว่าแบบจำลอง DL ให้ผลพยากรณ์ราคาหรือผลตอบแทนที่ดีกว่าแบบจำลอง ML แบบดั้งเดิมจริงหรือไม่ [2] ขณะเดียวกันงานทบทวนวิธีการ ensemble learning สำหรับตลาดหุ้นก็ยืนยันในทิศทางเดียวกันว่าการเลือกใช้คุณลักษณะ (feature) ที่เหมาะสมมีผลต่อความแม่นยำมากพอ ๆ กับการเลือกประเภทแบบจำลอง [3]

สำหรับประเทศไทย ตลาดหลักทรัพย์แห่งประเทศไทย (Stock Exchange of Thailand: SET) เป็นตลาดทุนหลักของภูมิภาคอาเซียนที่มีมูลค่าการซื้อขายและจำนวนนักลงทุนเพิ่มขึ้นอย่างต่อเนื่อง งานศึกษาประสิทธิภาพของตลาดในกลุ่มประเทศอาเซียนรวมถึงไทยพบหลักฐานที่ผสมผสาน (mixed evidence) กล่าวคือในระยะสั้นตลาดอาจมีความไม่มีประสิทธิภาพอยู่บ้าง แต่ในระยะยาวสอดคล้องกับสมมติฐานตลาดที่มีประสิทธิภาพในรูปแบบ

อ่อน (weak-form efficiency) [4] ความพยายามพยากรณ์ดัชนี SET Index ด้วยแบบจำลองปัญญาประดิษฐ์จึงเกิดขึ้นมาอย่างต่อเนื่อง อาทิ การใช้โครงข่ายประสาทเทียมร่วมกับขั้นตอนวิธีเชิงพันธุกรรม (genetic algorithm) เพื่อพยากรณ์แนวโน้มดัชนี SET50 [5] และการใช้แบบจำลอง Gaussian Process Regression เปรียบเทียบกับโครงข่ายประสาทเทียมในการพยากรณ์ดัชนี SET ทั้งช่วงก่อนและหลังวิกฤตโควิด-19 [6] อย่างไรก็ตาม การศึกษาส่วนใหญ่มักทดสอบแบบจำลองเพียงหนึ่งหรือสองตระกูลในแต่ละครั้ง ยังไม่มีการเปรียบเทียบแบบจำลองหลายตระกูลภายใต้กรอบการทดลองเดียวกันอย่างเป็นระบบ

กล่าวโดยสรุป ช่องว่างงานวิจัย (research gap) ที่งานวิจัยนี้มุ่งอุดสามารถสรุปได้เป็นสามประเด็นหลัก ประการแรก การศึกษาที่ผ่านมาเกี่ยวกับการพยากรณ์ดัชนี SET Index ด้วยแบบจำลองปัญญาประดิษฐ์ยังจำกัดขอบเขตอยู่ที่แบบจำลองเพียงหนึ่งหรือสองตระกูลในแต่ละครั้ง เช่น การใช้ ANN ร่วมกับ genetic algorithm [5] หรือ Gaussian Process Regression [6] โดยยังไม่มีเปรียบเทียบแบบจำลองหลายตระกูล (สถิติเชิงเวลา การเรียนรู้ของเครื่อง และโครงข่ายประสาทเทียมเชิงลึก) ภายใต้กรอบข้อมูลและการประเมินผลชุดเดียวกันอย่างเป็นระบบ ประการที่สอง ผลลัพธ์ความแม่นยำในระดับสูงที่รายงานในบางการศึกษา เช่น ความแม่นยำเชิงทิศทางร้อยละ 82.94 ของดัชนี SET50 [7] ยังไม่มีการยืนยันซ้ำด้วยชุดข้อมูลช่วงเวลาอื่นหรือเปรียบเทียบกับแบบจำลองตระกูลอื่นอย่างต้นไม้ตัดสินใจแบบรวมกลุ่ม ทำให้ยังไม่ชัดเจนว่าความสามารถทั่วไปของแบบจำลองหรือคุณลักษณะเฉพาะของช่วงเวลาและระเบียบวิธีที่ศึกษา และประการที่สาม แม้หลักฐานจากการแข่งขันพยากรณ์ระดับโลกอย่าง M4 Competition จะชี้ว่าแบบจำลองซับซ้อนไม่จำเป็นต้องให้ผลดีกว่าวิธีการอย่างง่ายเสมอไป [8], [9] แต่ยังไม่มีการทดสอบข้อสรุปดังกล่าวอย่างเป็นระบบกับดัชนี SET Index โดยตรง งานวิจัยนี้จึงออกแบบมาเพื่ออุดช่องว่างทั้งสามประเด็นนี้พร้อมกันภายใต้กรอบการทดลองเดียว

ในบริบทที่กว้างขึ้น ข้อถกเถียงว่าแบบจำลองที่มีความซับซ้อนสูงอย่าง DL ให้ผลพยากรณ์ที่ดีกว่าวิธีการทางสถิติหรือ ML แบบดั้งเดิมเสมอไปหรือไม่นั้น ได้รับคำตอบบางส่วนจากการแข่งขันพยากรณ์อนุกรมเวลาระดับโลกอย่าง M4 Competition ซึ่งทดสอบแบบจำลองกว่า 60 วิธี

บนอนุกรมเวลากว่าหนึ่งแสนชุด ผลปรากฏว่าแบบจำลอง ML แบบล้วน (pure ML) ทั้งหมดให้ผลแย่กว่าการรวมวิธีการทางสถิติแบบง่ายเพียงสามวิธี [8] และมีเพียงแบบจำลองแบบผสม (hybrid) ระหว่างสถิติกับ ML เท่านั้นที่เอาชนะเกณฑ์เปรียบเทียบได้ แม้จะต้องใช้เวลาในการประมวลผลมากกว่าอย่างมีนัยสำคัญ [9] บทวิเคราะห์เชิงทฤษฎีและปฏิบัติด้านการพยากรณ์ในเวลาต่อมาได้สรุปสอดคล้องกันว่าวิธีการง่าย ๆ มักให้ผลเทียบเท่าหรือดีกว่าแบบจำลองซับซ้อน โดยเฉพาะเมื่อขนาดตัวอย่างมีจำกัดและอนุกรมเวลามีสัญญาณรบกวนสูง [10] ข้อค้นพบเหล่านี้เป็นกรอบสำคัญที่งานวิจัยนี้ใช้ตั้งสมมติฐานและตีความผลการทดลอง

สำหรับกรณีของดัชนี SET Index โดยเฉพาะการศึกษาที่ใช้โครงข่ายประสาทเทียมเชิงลึกสามชั้นซ้อนร่วมกับตัววัดเชิงเทคนิคหลายช่วงเวลาซ้อนหลัง (3, 5, 10 และ 15 วัน) รายงานความแม่นยำเชิงทิศทางของดัชนี SET50 สูงถึงร้อยละ 82.94 ในช่วงข้อมูลปี 2558-2562 [7] ซึ่งนับเป็นตัวเลขอ้างอิงที่สูงมากเมื่อเทียบกับผลการศึกษาในตลาดอื่น อย่างไรก็ตาม ผลลัพธ์ในระดับดังกล่าวยังไม่มีการยืนยันซ้ำด้วยชุดข้อมูลช่วงเวลาอื่นหรือด้วยแบบจำลองตระกูลอื่น เช่น ต้นไม้ตัดสินใจแบบรวมกลุ่ม (tree ensemble) ทำให้ยังไม่ชัดเจนว่าความแม่นยำระดับสูงดังกล่าวเป็นคุณลักษณะเฉพาะของช่วงเวลาและดัชนีที่ศึกษาหรือเป็นความสามารถทั่วไปของแบบจำลอง DL ในตลาดไทย ในทำนองเดียวกัน การศึกษาก่อนหน้าที่ใช้ ANN ร่วมกับ genetic algorithm บนดัชนี SET50 [5] และการศึกษาที่ใช้ Gaussian Process Regression บนดัชนี SET [6] ต่างก็จำกัดขอบเขตอยู่ที่แบบจำลองเพียงหนึ่งถึงสองตระกูลและช่วงเวลาก่อนปี 2563 เป็นส่วนใหญ่

เมื่อพิจารณาความแตกต่างเชิงระเบียบวิธีระหว่างงานวิจัยนี้กับการศึกษาที่รายงานความแม่นยำเชิงทิศทางสูงถึงร้อยละ 82.94 [7] อย่างเป็นระบบ พบความแตกต่างอย่างน้อยสี่ประการ ประการแรก ดัชนีที่ศึกษาต่างกัน คือ SET50 ซึ่งเป็นดัชนีหุ้นขนาดใหญ่ 50 ตัวที่มีสภาพคล่องสูงและเคลื่อนไหวสอดคล้องกันมากกว่า เทียบกับ SET Index ซึ่งเป็นดัชนีรวมตลาดที่ครอบคลุมหุ้นทุกขนาด ประการที่สอง ช่วงเวลาข้อมูลต่างกัน คือ ปี 2558-2562 ซึ่งเป็นช่วงก่อนวิกฤตโควิด-19 ที่ตลาดมีความผันผวนค่อนข้างต่ำ เทียบกับข้อมูลของงานวิจัยนี้ที่ครอบคลุมถึงปี 2569 ซึ่งรวมทั้งช่วง

วิกฤตโควิด-19 และภาวะปรับฐานรุนแรงในปี 2568 ประการที่สาม การศึกษาดังกล่าวมิได้ระบุรายละเอียดวิธีการแบ่งชุดข้อมูลฝึกและทดสอบและมาตรการป้องกัน look-ahead bias ไว้อย่างชัดเจนเท่ากับงานวิจัยนี้ ซึ่งอาจส่งผลต่อความน่าเชื่อถือของตัวเลขความแม่นยำที่รายงาน และประการที่สี่ การศึกษาดังกล่าวทดสอบเฉพาะแบบจำลองโครงข่ายประสาทเทียมตระกูลเดียว มิได้เปรียบเทียบกับแบบจำลองต้นไม้ตัดสินใจแบบรวมกลุ่มดังเช่นงานวิจัยนี้ ความแตกต่างทั้งสี่ประการนี้ทำให้ตัวเลขความแม่นยำร้อยละ 82.94 ยังไม่อาจสรุปเป็นความสามารถทั่วไปของแบบจำลอง DL ในตลาดหุ้นไทยได้โดยตรง

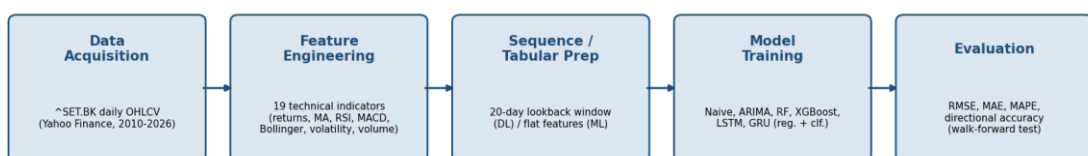
ในเชิงทฤษฎี พฤติกรรมราคาหลักทรัพย์ระยะสั้นที่ใกล้เคียงกับการเดินสุ่ม (random walk) ได้รับการอธิบายไว้ตั้งแต่งานคลาสสิกของ Fama ที่พบว่า การเปลี่ยนแปลงราคาหุ้นรายวันมีลักษณะใกล้เคียงการเดินสุ่มอย่างมีนัยสำคัญ [11] และได้รับการประมวลเป็นกรอบทฤษฎีตลาดที่มีประสิทธิภาพ (Efficient Market Hypothesis: EMH) ในเวลาต่อมา ซึ่งจำแนกประสิทธิภาพของตลาดออกเป็นระดับอ่อน กึ่งเข้ม และเข้ม [12] หากดัชนี SET Index มีลักษณะสอดคล้องกับประสิทธิภาพระดับอ่อนดังที่มีหลักฐานบางส่วนสนับสนุนในภูมิภาคอาเซียน [4] ก็เป็นที่คาดได้ว่าแบบจำลองพยากรณ์ไม่ว่าจะซับซ้อนเพียงใดควรให้ผลใกล้เคียงกับเกณฑ์เปรียบเทียบแบบพื้นฐาน (naive benchmark) ในด้านความคลาดเคลื่อนเชิงระดับราคา แต่คำถามที่ยังไม่มีคำตอบชัดเจนคือแบบจำลองใดสามารถจับสัญญาณเชิงทิศทาง (directional signal) ที่มีประโยชน์เชิงปฏิบัติได้ดีกว่ากัน ภายใต้กรอบข้อมูลและการประเมินผลชุดเดียวกัน

บทความนี้จึงมีวัตถุประสงค์เพื่อนำเสนอกระบวนการพยากรณ์ดัชนี SET Index รายวันแบบครบวงจร (end-to-end pipeline) ที่เปรียบเทียบแบบจำลองแปดชนิดครอบคลุมสามตระกูลหลัก ได้แก่ แบบจำลองพื้นฐานและสถิติเชิงเวลา (naive persistence และ ARIMA) แบบจำลองการเรียนรู้ของเครื่อง (Random Forest และ XGBoost) และแบบจำลองโครงข่ายประสาทเทียมเชิงลึกทั้งรูปแบบถดถอยและจำแนกประเภท (LSTM และ GRU) ภายใต้ชุดข้อมูล ชุดตัวแปรคุณลักษณะ และเกณฑ์การประเมินผลเดียวกัน โดยมีวัตถุประสงค์เฉพาะ 3 ประการ ได้แก่ 1) เพื่อเปรียบเทียบความคลาดเคลื่อนเชิงระดับราคา (RMSE, MAE, MAPE) ของแบบจำลองทั้งแปดชนิดบนชุดข้อมูลทดสอบช่วงปี 2567-2569 2) เพื่อเปรียบเทียบความแม่นยำเชิงทิศทาง (Directional Accuracy) ของแบบจำลองทั้งแปดชนิดเทียบกับเกณฑ์การสุ่มทายที่ร้อยละ 50 และ 3) เพื่อตรวจสอบว่าความซับซ้อนของแบบจำลอง โดยเฉพาะแบบจำลอง DL แบบลำดับเวลา (sequence model) ให้ผลพยากรณ์ที่ดีกว่าแบบจำลองต้นไม้ตัดสินใจแบบรวมกลุ่มที่มีโครงสร้างเรียบง่ายกว่าหรือไม่ โครงสร้างของบทความส่วนที่เหลือประกอบด้วยวิธีดำเนินงานวิจัยในส่วนที่ 2 ผลและการอภิปรายผลการวิจัยในส่วนที่ 3 และบทสรุปในส่วนที่ 4

2. วิธีดำเนินงานวิจัย

งานวิจัยนี้ดำเนินการตามกระบวนการพยากรณ์แบบครบวงจร (end-to-end pipeline) 5 ขั้นตอนหลัก ได้แก่ การจัดเก็บข้อมูล การสร้างตัวแปรคุณลักษณะ การเตรียมข้อมูลเชิงลำดับและเชิงตาราง การฝึกแบบจำลอง และการประเมินผล ดังแสดงในรูปที่ 1 โดยมีรายละเอียดในแต่ละขั้นตอนดังนี้

Forecasting pipeline architecture



รูปที่ 1 แผนภาพระบบและขั้นตอนการพยากรณ์ดัชนี SET Index แบบครบวงจร 5 ขั้นตอน

2.1 ข้อมูลและแหล่งข้อมูล

ข้อมูลราคาปิดรายวันของดัชนีตลาดหลักทรัพย์แห่งประเทศไทย (SET Index สัญลักษณ์ ^SET.BK) รวบรวมผ่าน Yahoo Finance API ด้วยไลบรารี yfinance ครอบคลุมช่วงเวลาตั้งแต่วันที่ 4 มกราคม 2553 ถึงวันที่ 17 กรกฎาคม 2569 รวม 4,012 วันทำการก่อนการตัดข้อมูลที่ไม่สมบูรณ์ออก ผู้วิจัยตรวจสอบความถูกต้องของข้อมูลโดยเปรียบเทียบระดับดัชนีที่ได้ (อยู่ในช่วงประมาณ 1,200-1,800 จุดในช่วงกลางอนุกรม) กับระดับดัชนี SET Index จริงในช่วงเวลาเดียวกัน ซึ่งพบว่าสอดคล้องกัน ยืนยันได้ว่าข้อมูลที่ใช้เป็นข้อมูลดัชนีทางการ มิใช่ตัวแทน (proxy) ที่สร้างขึ้นเอง

งานวิจัยนี้เลือกช่วงเวลาข้อมูลตั้งแต่ปี 2553 ด้วยเหตุผลสามประการ ประการแรก เป็นช่วงเวลายาวเพียงพอ (ประมาณ 16 ปี) ให้ครอบคลุมสภาวะตลาดที่หลากหลาย ทั้งช่วงขาขึ้นต่อเนื่อง ช่วงวิกฤตโควิด-19 และช่วงปรับฐาน ซึ่งช่วยให้แบบจำลองได้รับการฝึกและประเมินผลภายใต้พลวัตของตลาดที่ไม่จำกัดอยู่เพียงสภาวะใดสภาวะหนึ่ง ประการที่สอง ปี 2553 เป็นจุดเริ่มต้นที่ข้อมูลราคาปิดรายวันผ่าน Yahoo Finance API มีความสมบูรณ์และต่อเนื่องสำหรับดัชนี SET Index และประการที่สาม ระยะเวลาดังกล่าวเพียงพอสำหรับการแบ่งชุดข้อมูลออกเป็นชุดฝึก ชุดตรวจสอบ และชุดทดสอบที่มีขนาดเหมาะสมตามหลักการแบ่งข้อมูลเชิงเวลาในงานพยากรณ์อนุกรมเวลาทางการเงิน [13]

2.2 การสร้างตัวแปรคุณลักษณะ

ตัวแปรเป้าหมายของแบบจำลองประกอบด้วย ราคาปิดวันถัดไป ผลตอบแทนวันถัดไป และทิศทางการเปลี่ยนแปลงวันถัดไป (ขึ้น = 1, ลง = 0) โดยผลตอบแทนรายวัน (ณ เวลา (daily log return) คำนวณดังสมการที่ (1) [13]

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (1)$$

โดยที่ P_t คือราคาปิดของดัชนี SET Index ณ วันที่ t และ $P_{(t-1)}$ คือราคาปิดของวันทำการก่อนหน้า ตัวแปรคุณลักษณะเชิงเทคนิคที่สร้างขึ้นทั้งหมด 19 ตัวแปร แบ่ง

ออกเป็น 6 กลุ่ม ได้แก่ กลุ่มผลตอบแทนย้อนหลัง (ret_1d และ ret_lag ที่ 1, 3, 5 และ 10 วัน) กลุ่มอัตราส่วนราคาต่อค่าเฉลี่ยเคลื่อนที่ (ma_ratio ที่หน้าต่าง 5, 10, 20 และ 60 วัน) กลุ่มความผันผวนซึ่งวัดจากส่วนเบี่ยงเบนมาตรฐานของผลตอบแทนรายวัน (vol_5d, vol_10d, vol_20d) กลุ่มดัชนีโมเมนตัม Relative Strength Index ที่หน้าต่าง 14 วัน (rsi_14) กลุ่ม Moving Average Convergence Divergence (macd, macd_signal, macd_hist) และกลุ่มที่เกี่ยวข้องกับ Bollinger Band และปริมาณการซื้อขาย (boll_pctb, vol_chg_1d, vol_ratio_20) หลังจากตัดแถวข้อมูลที่มีค่าว่างอันเกิดจากการคำนวณค่าเฉลี่ยเคลื่อนที่แบบ rolling window เหลือข้อมูลจำนวน 3,945 วันทำการ ตั้งแต่วันที่ 29 มีนาคม 2553 ถึงวันที่ 9 กรกฎาคม 2569 โดยมีสัดส่วนวันที่ราคาปิดปรับตัวขึ้นร้อยละ 53.0 ของจำนวนวันทั้งหมด

ข้อสังเกตประการหนึ่งคือช่วงเวลาข้อมูลดิบที่รวบรวมได้ (ถึงวันที่ 17 กรกฎาคม 2569 ตามที่ระบุในบทความ) สั้นสุดช้ากว่าช่วงเวลาของข้อมูลหลังการสร้างตัวแปรคุณลักษณะที่ใช้ในการฝึกและทดสอบแบบจำลองจริง (ถึงวันที่ 9 กรกฎาคม 2569) อยู่ประมาณ 6 วันทำการ ทั้งนี้เนื่องจากข้อมูลปริมาณการซื้อขายและราคาของวันทำการช่วงปลายอนุกรมบางส่วนยังไม่ถูกปรับปรุงให้สมบูรณ์ (finalize) ในฐานข้อมูลของ Yahoo Finance ณ ขณะที่ดำเนินการดึงข้อมูล ส่งผลให้ค่าตัวแปรคุณลักษณะเชิงเทคนิคบางตัวของแถวข้อมูลในช่วงดังกล่าวมีค่าว่าง (missing) และถูกตัดออกในขั้นตอนคัดกรองข้อมูลที่ไม่สมบูรณ์ตามที่อธิบายไว้ในหัวข้อ 2.1 ความคลาดเคลื่อนของช่วงเวลาดังกล่าวมีขนาดเล็กมากเมื่อเทียบกับความยาวทั้งหมดของอนุกรมข้อมูล (3,945 วัน) จึงไม่ส่งผลกระทบต่อผลการศึกษาเปรียบเทียบแบบจำลองอย่างมีนัยสำคัญต่อผลการเปรียบเทียบแบบจำลอง

การกำหนดขนาดหน้าต่างเวลาของตัวแปรคุณลักษณะกลุ่มผลตอบแทนย้อนหลัง ค่าเฉลี่ยเคลื่อนที่ และความผันผวน (1, 3, 5, 10, 20 และ 60 วัน) อ้างอิงตามธรรมเนียมปฏิบัติทั่วไปในการวิเคราะห์ทางเทคนิคและงานวิจัยด้านการพยากรณ์ตลาดหุ้นที่ใช้ตัวชี้วัดเชิงเทคนิค [14], [15] โดยหน้าต่าง 5 วันสอดคล้องกับจำนวนวันทำการใน 1 สัปดาห์ หน้าต่าง 10 และ 20 วันสอดคล้องกับรอบประมาณ 2 สัปดาห์และ 1 เดือนตามลำดับ และหน้าต่าง 60 วันสอดคล้องกับรอบประมาณ 1 ไตรมาส ซึ่งครอบคลุมมิติ

เวลาตั้งแต่ระยะสั้นถึงระยะกลางที่นักวิเคราะห์ทางเทคนิคใช้พิจารณาแนวโน้มราคาโดยทั่วไป การเลือกใช้ชุดหน้าต่างเวลาที่หลากหลายในลักษณะนี้ยังช่วยให้แบบจำลองทุกตระกูลมีโอกาสเข้าถึงสัญญาณเชิงเวลาในหลายระดับความละเอียดอย่างเท่าเทียมกัน

2.3 การแบ่งชุดข้อมูล

ข้อมูลถูกแบ่งออกเป็นสามชุดตามลำดับเวลาโดยไม่มีการสุ่ม (chronological split) เพื่อป้องกันปัญหาการรั่วไหลของข้อมูลไปข้างหน้า (look-ahead bias) รายละเอียดการแบ่งชุดข้อมูลแสดงในตารางที่ 1

ตารางที่ 1 การแบ่งชุดข้อมูลตามลำดับเวลาสำหรับการฝึก ตรวจสอบ และทดสอบแบบจำลอง

ชุดข้อมูล	ช่วงเวลา	จำนวนวัน (วัน)
ชุดฝึก (Train)	29 มี.ค. 2553 - 30 ธ.ค. 2564	2,860
ชุดตรวจสอบ (Validation)	4 ม.ค. 2565 - 28 ธ.ค. 2566	479
ชุดทดสอบ (Test)	3 ม.ค. 2567 - 9 ก.ค. 2569	606

หมายเหตุ: การแบ่งข้อมูลเป็นแบบเรียงตามเวลา (chronological) ไม่มีการสุ่ม เพื่อป้องกัน look-ahead bias

ชุดฝึกใช้สำหรับปรับพารามิเตอร์ของแบบจำลอง ชุดตรวจสอบใช้สำหรับเลือกค่า hyperparameter และการหยุดฝึกก่อนกำหนด (early stopping) ของแบบจำลองโครงข่ายประสาทเทียม ส่วนชุดทดสอบใช้สำหรับประเมินผลนอกตัวอย่าง (out-of-sample) เพียงครั้งเดียวเมื่อสิ้นสุดกระบวนการปรับแต่งแบบจำลองทั้งหมดแล้ว ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ใช้ในการปรับมาตรฐานตัวแปร (standardization) สำหรับแบบจำลอง LSTM และ GRU คำนวณจากชุดฝึกเท่านั้น เพื่อไม่ให้ข้อมูลจากชุดตรวจสอบและชุดทดสอบปนเปื้อนกระบวนการฝึก

สัดส่วนการแบ่งชุดข้อมูลในตารางที่ 1 (ชุดฝึกร้อยละ 72.5 ชุดตรวจสอบร้อยละ 12.1 และชุดทดสอบร้อยละ 15.4 ของจำนวนวันทั้งหมด) กำหนดขึ้นตามหลักการทั่วไปของการแบ่งข้อมูลอนุกรมเวลาสำหรับงานพยากรณ์ โดยให้น้ำหนักกับชุดฝึกมากที่สุดเพื่อให้แบบจำลองโดยเฉพาะโครงข่ายประสาทเทียมเชิงลึกมีข้อมูลเพียงพอสำหรับการ

เรียนรู้รูปแบบเชิงเวลา ขอบเขตปีของแต่ละชุดยังถูกกำหนดให้สอดคล้องกับเหตุการณ์ทางเศรษฐกิจที่สำคัญ กล่าวคือชุดฝึก (2553-2564) ครอบคลุมทั้งช่วงตลาดขาขึ้นปกติและวิกฤตโควิด-19 ชุดตรวจสอบ (2565-2566) เป็นช่วงเวลาที่ต่อเนื่องจากชุดฝึกและมีความผันผวนรุนแรงเป็นพิเศษ เหมาะสำหรับการเลือก hyperparameter อย่างมีประสิทธิภาพ ส่วนชุดทดสอบ (2567-2569) เป็นช่วงเวลาที่ล่าช้าที่สุดที่สะท้อนสภาวะตลาดปัจจุบันได้ดีที่สุดสำหรับการประเมินผลนอกตัวอย่างอย่างสมจริง

2.4 แบบจำลองที่ใช้เปรียบเทียบ

งานวิจัยนี้เปรียบเทียบแบบจำลองพยากรณ์ทั้งสิ้น 8 แบบจำลอง ครอบคลุม 4 ตระกูล ตระกูลแรกคือแบบจำลองพื้นฐาน (naive persistence) ซึ่งพยากรณ์ราคาปิดวันถัดไปเท่ากับราคาปิดวันปัจจุบัน ตระกูลที่สองคือแบบจำลองสถิติเชิงเวลาแบบ Autoregressive Integrated Moving Average (ARIMA) โดยเลือกอันดับ (p, d, q) ด้วยเกณฑ์ Akaike Information Criterion (AIC) [16] บนชุดฝึก ซึ่งได้ผลลัพธ์เป็น ARIMA(0,1,0) และพยากรณ์แบบ rolling one-step ตระกูลที่สามคือแบบจำลองการเรียนรู้ของเครื่องเชิงต้นไม้แบบรวมกลุ่ม (tree ensemble) ได้แก่ Random Forest ซึ่งใช้หลักการรวมผลของต้นไม้ตัดสินใจจำนวนมากที่ฝึกด้วยกลไก bagging และการสุ่มเลือกตัวแปร [17] และ XGBoost ซึ่งเป็นการปรับปรุงเทคนิค gradient boosting ให้มีประสิทธิภาพเชิงคำนวณสูงขึ้นด้วยขั้นตอนวิธีที่คำนึงถึงข้อมูลการจัดกระจาย (sparsity-aware) [18] ทั้งสองแบบจำลองในตระกูลนี้ถูกออกแบบให้พยากรณ์ผลตอบแทนรายวันแทนระดับราคาโดยตรง สอดคล้องกับแนวทางของงานวิจัยก่อนหน้านี้ที่ใช้ต้นไม้ตัดสินใจร่วมกับตัวชี้วัดเชิงเทคนิคในการพยากรณ์ทิศทางราคาหุ้น [19] Random Forest กำหนดค่า 500 ต้นไม้ ความลึกสูงสุด 8 ชั้น และจำนวนตัวอย่างขั้นต่ำต่อไป 5 ตัวอย่าง ส่วน XGBoost กำหนด 500 รอบการฝึก ความลึกสูงสุด 4 ชั้น และอัตราการเรียนรู้ 0.03

ค่าพารามิเตอร์ของ Random Forest และ XGBoost ข้างต้นกำหนดขึ้นโดยอาศัยค่าที่นิยมใช้ในงานวิจัยด้านการพยากรณ์ตลาดหุ้นที่มีลักษณะข้อมูลใกล้เคียงกัน [19] ร่วมกับการตรวจสอบประสิทธิภาพเบื้องต้นบนชุดตรวจสอบ (validation set) จำนวนต้นไม้ 500 ต้นสำหรับ Random Forest และจำนวนรอบการฝึก 500 รอบสำหรับ

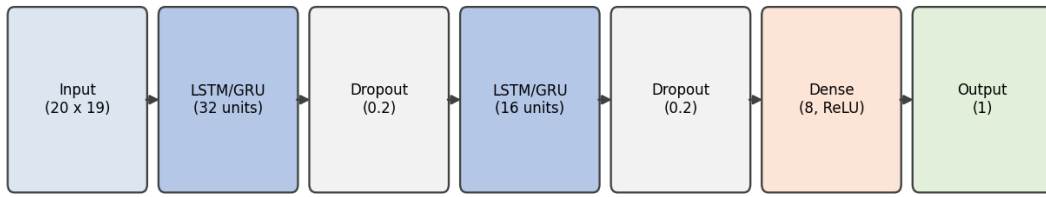
XGBoost เป็นค่าที่มากพอให้แบบจำลองมีเสถียรภาพโดยไม่เพิ่มภาระการคำนวณเกินจำเป็น ขณะที่การจำกัดความลึกสูงสุด (8 ชั้นสำหรับ Random Forest และ 4 ชั้นสำหรับ XGBoost) และจำนวนตัวอย่างขั้นต่ำต่อใบสำหรับ Random Forest มีวัตถุประสงค์เพื่อควบคุมความซับซ้อนของแบบจำลองและป้องกันปัญหา overfitting ซึ่งเป็นความเสี่ยงสำคัญเมื่อฝึกแบบจำลองต้นไม้ตัดสินใจด้วยข้อมูลอนุกรมเวลาทางการเงินที่มีสัญญาณรบกวนสูง ส่วนอัตราการเรียนรู้ที่ค่อนข้างต่ำของ XGBoost (0.03) ช่วยให้การเรียนรู้อย่างค่อยเป็นค่อยไปแบบ gradient boosting ค่อย ๆ ปรับตัวอย่างละเอียดและลดความเสี่ยงของการ overfitting เมื่อใช้ร่วมกับจำนวนรอบการฝึกที่มาก

ตระกูลที่สี่คือแบบจำลองโครงข่ายประสาทเทียมเชิงลึกแบบลำดับเวลา ได้แก่ Long Short-Term Memory (LSTM) ซึ่งใช้กลไกประตูหน่วยความจำ (gated memory cell) เพื่อแก้ปัญหาการไล่ระดับที่หายไปหรือระเบิดในโครงข่ายประสาทเทียมแบบเวียนกลับ [20] และ Gated Recurrent Unit (GRU) ซึ่งเป็นสถาปัตยกรรมที่ลดความซับซ้อนของกลไกประตูจาก LSTM แต่ยังคงความสามารถในการจับรูปแบบเชิงลำดับได้ใกล้เคียงกัน [21] แบบจำลองทั้งสองถูกสร้างขึ้นทั้งในรูปแบบถดถอย (พยากรณ์ผลตอบแทน ใช้ mean squared error เป็นฟังก์ชันการสูญเสีย) และรูปแบบจำแนกประเภท (พยากรณ์ทิศทางโดยตรงผ่านฟังก์ชัน sigmoid ใช้ binary cross-entropy เป็นฟังก์ชันการสูญเสีย) สถาปัตยกรรมประกอบด้วยชั้น LSTM/GRU สองชั้น (32 และ 16 หน่วย ตามลำดับ) คันด้วยชั้น Dropout อัตรา 0.2 ตามด้วยชั้น Dense ขนาด 8 หน่วย ฟังก์ชันกระตุ้น ReLU และชั้นผลลัพธ์ ดังแสดงในรูปที่ 2 โดยใช้หน้าต่างข้อมูลย้อนหลัง (lookback window) 20 วันทำการ และตัวแปรคุณลักษณะทั้งหมดถูกปรับมาตรฐานด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานจากชุดฝึกเท่านั้น

สถาปัตยกรรม LSTM/GRU ข้างต้นออกแบบให้มีขนาดค่อนข้างเล็ก (32 และ 16 หน่วยในสองชั้นซ่อน) โดยเจตนา เนื่องจากขนาดตัวอย่างของชุดฝึก (ประมาณ 2,860 วัน) มีจำกัดเมื่อเทียบกับจำนวนพารามิเตอร์ที่โครงข่ายประสาทเทียมขนาดใหญ่ต้องการเพื่อเรียนรู้อย่างมีประสิทธิภาพ ชั้น Dropout อัตรา 0.2 ทำหน้าที่ลดความเสี่ยงของ overfitting ระหว่างการฝึก ส่วนหน้าต่างข้อมูลย้อนหลัง 20 วันทำการเลือกให้สอดคล้องกับหน้าต่างเวลาที่ยาวที่สุดในกลุ่มตัวแปรคุณลักษณะระยะกลาง (ma_ratio ที่ 20 วัน) เพื่อให้แบบจำลองมีโอกาสเรียนรู้รูปแบบเชิงเวลาในช่วงประมาณ 1 เดือนปฏิทิน ซึ่งเป็นกรอบเวลาที่นิยมใช้ในงานพยากรณ์ราคาหลักทรัพย์ด้วยแบบจำลอง sequence [22], [23] ค่าพารามิเตอร์ทั้งหมดได้รับการตรวจสอบเบื้องต้นบนชุดตรวจสอบก่อนนำไปใช้ฝึกแบบจำลองสุดท้าย

การเลือกจำกัดขอบเขตไว้ที่แบบจำลองแปดชนิดครอบคลุมสี่ตระกูลนี้มีเหตุผลสนับสนุนสองประการ ประการแรก แบบจำลองทั้งแปดชนิดเป็นแบบจำลองที่มีการใช้งานอย่างแพร่หลายและมีหลักฐานเชิงประจักษ์รองรับในบริบทการพยากรณ์ตลาดหุ้นมาอย่างต่อเนื่อง ทำให้ผลการเปรียบเทียบสามารถตีความและเทียบเคียงกับวรรณกรรมที่มีอยู่ได้โดยตรง ประการที่สอง งานวิจัยนี้มุ่งเน้นการเปรียบเทียบเชิงระบบระหว่างตระกูลแบบจำลองหลัก (พื้นฐาน สถิติเชิงเวลา การเรียนรู้ของเครื่อง และโครงข่ายประสาทเทียมเชิงลึกแบบลำดับเวลา) มากกว่าการไล่ตามแบบจำลองที่ทันสมัยที่สุด สถาปัตยกรรมที่ซับซ้อนกว่า เช่น Temporal Fusion Transformer มักต้องการข้อมูลปริมาณมากและการปรับแต่ง hyperparameter อย่างเข้มข้นกว่าจึงจะแสดงศักยภาพได้เต็มที่ [24] ซึ่งอยู่นอกขอบเขตของงานวิจัยนี้ที่ใช้ข้อมูลอนุกรมเวลาเดียวขนาดจำกัด การขยายการเปรียบเทียบไปยังสถาปัตยกรรมดังกล่าวจึงถูกระบุไว้เป็นข้อเสนอแนะสำหรับงานวิจัยในอนาคต

LSTM / GRU sequence-model architecture



รูปที่ 2 สถาปัตยกรรมแบบจำลอง LSTM/GRU ที่ใช้ในงานวิจัย

การเลือกใช้ตัวชี้วัดเชิงเทคนิคที่ผ่านการคัดกรองอย่างเหมาะสมมีบทบาทสำคัญต่อความแม่นยำของแบบจำลองไม่น้อยไปกว่าการเลือกสถาปัตยกรรมแบบจำลอง [14] ซึ่งสอดคล้องกับงานทบทวนวรรณกรรมด้านการคัดเลือกตัวแปรสำหรับการพยากรณ์ตลาดหุ้นที่พบว่าคุณภาพของกระบวนการคัดเลือกตัวแปรส่งผลกระทบต่อความแม่นยำโดยไม่ขึ้นกับขั้นตอนวิธีการเรียนรู้ที่เลือกใช้ [15] งานวิจัยนี้จึงออกแบบตัวแปรคุณลักษณะทั้ง 19 ตัวแปรให้ครอบคลุมมิติผลตอบแทน แนวโน้ม ความผันผวน โมเมนตัม และปริมาณการซื้อขาย เพื่อให้แบบจำลองทุกตระกูลใช้ชุดข้อมูลนำเข้าเดียวกันอย่างเป็นธรรม (fair comparison) สำหรับแบบจำลองที่พยากรณ์ผลตอบแทนราคาที่พยากรณ์คำนวณย้อนกลับด้วยสูตรราคาปิดพยากรณ์เท่ากับราคาปิดวันนี้คูณด้วย exponential ของผลตอบแทนที่พยากรณ์ ทั้งนี้ในการทดลองเบื้องต้นที่ให้แบบจำลองต้นไม้ตัดสินใจพยากรณ์ระดับราคาโดยตรง พบปัญหาการประมาณค่านอกช่วงข้อมูลที่เคยเห็นระหว่างการฝึก (extrapolation) จึงปรับมาพยากรณ์ผลตอบแทนแทน

2.5 ตัวชี้วัดการประเมินผล

การประเมินความแม่นยำของแบบจำลองแบ่งออกเป็นสองมิติ มิติแรกคือความคลาดเคลื่อนเชิงระดับราคา วัดด้วยค่าคลาดเคลื่อนกำลังสองเฉลี่ยรากที่สอง (Root Mean Squared Error: RMSE) ดังสมการที่ (2) [25]

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

และค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) ดังสมการที่ (3)

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

โดยที่ y_i คือราคาปิดจริง $\hat{y}_i$ คือราคาปิดที่พยากรณ์ และ N คือจำนวนวันในชุดทดสอบ มิติที่สองคือความแม่นยำเชิงทิศทาง (Directional Accuracy: DirAcc) ซึ่งวัดสัดส่วนที่แบบจำลองทำนายทิศทางการเปลี่ยนแปลงของราคา (ขึ้นหรือลง) ได้ถูกต้อง ดังสมการที่ (4)

$$DirAcc = \frac{1}{N} \sum_{i=1}^N I(\text{sign}(\Delta y_i) = \text{sign}(\Delta \hat{y}_i)) \times 100\% \quad (4)$$

ตัวชี้วัดความแม่นยำเชิงทิศทางมีความสำคัญเป็นพิเศษสำหรับการประเมินประโยชน์เชิงปฏิบัติของแบบจำลองพยากรณ์ในบริบทการซื้อขายจริง เนื่องจากผลตอบแทนจากกลยุทธ์การซื้อขายมักขึ้นอยู่กับความถูกต้องของทิศทางมากกว่าความแม่นยำของระดับราคา [26] และงานวิจัยก่อนหน้านี้ชี้ให้เห็นว่าตัวชี้วัดทั้งสองมิติอาจให้ข้อสรุปที่แตกต่างกันได้ จึงจำเป็นต้องรายงานควบคู่กัน [27] เกณฑ์เปรียบเทียบพื้นฐานสำหรับความแม่นยำเชิงทิศทางคือร้อยละ 50 ซึ่งเทียบเท่าการสุ่มทาย [28]

เนื่องจากค่าความแม่นยำเชิงทิศทางที่สูงกว่าร้อยละ 50 ในเชิงตัวเลขมิได้เป็นหลักฐานยืนยันความสามารถในการพยากรณ์ที่มีนัยสำคัญทางสถิติเสมอไป โดยเฉพาะเมื่อขนาด

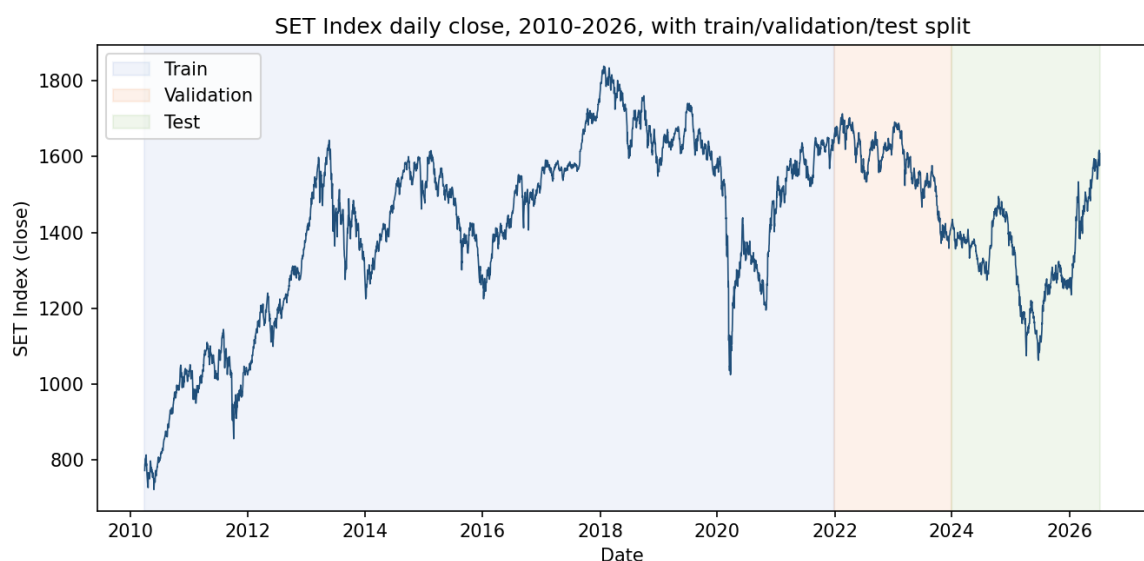
ชุดทดสอบมีจำกัดและผลต่างจากเกณฑ์มีขนาดเล็ก งานวิจัยนี้จึงเพิ่มการทดสอบทวินาม (binomial test) เพื่อทดสอบสมมติฐานว่าสัดส่วนการพยากรณ์ทิศทางที่ถูกต้องของแบบจำลองแต่ละชนิดแตกต่างจากค่าคาดหวังภายใต้การสุ่มทนาย (ร้อยละ 50) อย่างมีนัยสำคัญทางสถิติหรือไม่ โดยกำหนดจำนวนครั้งที่พยากรณ์ทิศทางถูกต้องเป็นตัวแปรสุ่มทวินามที่มีพารามิเตอร์ N เท่ากับจำนวนวันในชุดทดสอบ (606 วัน) และความน่าจะเป็นภายใต้สมมติฐานว่างเท่ากับ 0.5 พร้อมทั้งรายงานช่วงความเชื่อมั่นร้อยละ 95 ของสัดส่วนความแม่นยำเชิงทิศทางด้วยวิธี Wilson score interval แนวทางการทดสอบนัยสำคัญทางสถิติของความแม่นยำเชิงทิศทางในลักษณะนี้สอดคล้องกับหลักการที่เสนอไว้ในงานวิจัยด้านการทดสอบความสามารถในการพยากรณ์เชิงทิศทางของอนุกรมเวลาทางการเงิน [29]

3. ผลและการอภิปรายผลการวิจัย

ส่วนนี้นำเสนอผลเชิงประจักษ์จากกระบวนการพยากรณ์ดัชนี SET Index โดยเรียงลำดับจากลักษณะเชิงพรรณนาของข้อมูล ผลการเปรียบเทียบความคลาดเคลื่อนเชิงระดับราคา ผลการเปรียบเทียบความแม่นยำเชิงทิศทาง และปิดท้ายด้วยการอภิปรายกลไกเชิงทฤษฎีที่อาจอธิบายรูปแบบผลลัพธ์ที่พบ พร้อมทั้งเปรียบเทียบกับวรรณกรรมที่เกี่ยวข้อง

3.1 ลักษณะเชิงพรรณนาของข้อมูล

ก่อนนำเสนอผลการเปรียบเทียบแบบจำลอง การพิจารณาลักษณะเชิงพรรณนาของอนุกรมราคาดัชนี SET Index จะช่วยให้เข้าใจบริบทของการแบ่งชุดข้อมูลได้ชัดเจนยิ่งขึ้น รูปที่ 3 แสดงอนุกรมราคาปิดรายวันของดัชนี SET Index ตลอดช่วง พ.ศ. 2553-2569 พร้อมแรเงาแสดงขอบเขตของชุดฝึก ชุดตรวจสอบ และชุดทดสอบ



รูปที่ 3 อนุกรมราคาปิดรายวันของดัชนี SET Index ปี 2553-2569 พร้อมขอบเขตชุดฝึก ชุดตรวจสอบ และชุดทดสอบ

จากรูปที่ 3 จะเห็นว่าดัชนีเคลื่อนไหวอยู่ในช่วงประมาณ 700-1,830 จุดตลอดทั้งอนุกรม โดยชุดฝึก (พ.ศ. 2553-2564) ครอบคลุมทั้งช่วงขาขึ้นต่อเนื่องระหว่างปี 2555-2561 และภาวะปรับฐานรุนแรงในช่วงต้นปี 2563 อันเนื่องมาจากสถานการณ์โควิด-19 ชุดตรวจสอบ (พ.ศ. 2565-2566) อยู่ในช่วงดัชนีทรงตัวถึงปรับตัวลดลงเล็กน้อย ขณะที่ชุดทดสอบ (พ.ศ. 2567-2569) ครอบคลุมช่วงดัชนีปรับฐานลงไปและระดับต่ำกว่า 1,100 จุดในช่วงกลางปี 2568 ก่อน

จะฟื้นตัวกลับขึ้นไปเหนือ 1,600 จุดในช่วงกลางปี 2569 ลักษณะที่ชุดทดสอบครอบคลุมทั้งช่วงขาลงและขาขึ้นในสัดส่วนที่ใกล้เคียงกันมีข้อดีในเชิงระเบียบวิธี เนื่องจากช่วยลดโอกาสที่ผลการประเมินแบบจำลองจะถูกครอบงำโดยแนวโน้มทิศทางเดียวของตลาด

3.2 ผลการเปรียบเทียบความคลาดเคลื่อนเชิงระดับราคา

ตารางที่ 2 นำเสนอผลการเปรียบเทียบตัวชี้วัดความคลาดเคลื่อนเชิงระดับราคา (RMSE, MAE, MAPE) และ

ความแม่นยำเชิงทิศทาง (DirAcc) ของแบบจำลองทั้ง 8 แบบจำลองบนชุดข้อมูลทดสอบ ซึ่งครอบคลุมวันที่ 3 มกราคม 2567 ถึง 9 กรกฎาคม 2569 รวม 606 วันทำการ

ตารางที่ 2 ผลการเปรียบเทียบแบบจำลองพยากรณ์ทั้ง 8 แบบจำลอง บนชุดข้อมูลทดสอบ (3 ม.ค. 2567 - 9 ก.ค. 2569, 606 วัน) เรียงลำดับตามความแม่นยำเชิงทิศทางจากมากไปน้อย

แบบจำลอง	RMSE (จุด)	MAE (จุด)	MAPE (%)	Directional Accuracy (%)
Random Forest	13.71	10.04	0.76	51.8
XGBoost	13.99	10.26	0.77	50.8
LSTM (classification)	16.02	12.78	0.96	49.8
Naive persistence	13.52	9.94	0.75	49.7
GRU (regression)	13.52	9.95	0.75	49.7
LSTM (regression)	13.53	9.95	0.75	49.5
ARIMA(0,1,0)	19.28	14.45	1.09	49.0
GRU (classification)	16.55	12.97	0.97	48.2

หมายเหตุ: เกณฑ์การสุ่มหาย (chance) สำหรับ Directional Accuracy เท่ากับร้อยละ 50 ที่มา: model_comparison_master.csv

จากตารางที่ 2 ในมิติความคลาดเคลื่อนเชิงระดับราคา แบบจำลองส่วนใหญ่ให้ค่า RMSE ใกล้เคียงกันมากในช่วง 13.5-14.0 จุด และค่า MAPE ในช่วงร้อยละ 0.75-0.77 ซึ่งใกล้เคียงกับแบบจำลอง naive persistence ที่ให้ค่า RMSE เท่ากับ 13.52 จุด และ MAPE เท่ากับร้อยละ 0.75 มีเพียงแบบจำลอง ARIMA(0,1,0) ที่ให้ค่าความคลาดเคลื่อนสูงกว่าอย่างชัดเจน (RMSE เท่ากับ 19.28 จุด MAPE เท่ากับร้อยละ 1.09) แม้ ARIMA(0,1,0) ที่ AIC เลือกด้วยตนเองจะมีโครงสร้างเทียบเท่าการเดินสุ่มบวกค่าคงที่ (random walk with drift) ก็ตาม ผลลัพธ์นี้บ่งชี้ว่าองค์ประกอบของการเปลี่ยนแปลงราคาดัชนี SET Index ในช่วงเวลา 1 วันส่วนใหญ่มีลักษณะสุ่มเดินแบบที่แบบจำลองซับซ้อนไม่สามารถพยากรณ์ระดับราคาได้แม่นยำไปกว่าการใช้ราคาล่าสุดเป็นตัวพยากรณ์

เมื่อพิจารณาสาเหตุที่ ARIMA(0,1,0) ให้ค่าความคลาดเคลื่อนสูงกว่า naive persistence อย่างชัดเจนทั้งที่ทั้งสองแบบจำลองมีโครงสร้างใกล้เคียงกันในเชิงทฤษฎี คำอธิบายที่เชื่อมโยงกับคุณลักษณะของข้อมูล SET Index มีดังนี้ ความแตกต่างเชิงโครงสร้างระหว่างสองแบบจำลองอยู่

ที่ค่าคงที่การเลื่อน (drift term) ซึ่ง ARIMA(0,1,0) ประมาณค่าจากผลตอบแทนเฉลี่ยของชุดฝึกและนำไปบวกเพิ่มในการพยากรณ์ทุกวัน ขณะที่ naive persistence ไม่มีค่าคงที่ดังกล่าว (เทียบเท่ากับกำหนด drift เท่ากับศูนย์) เนื่องจากดัชนี SET Index ในช่วงชุดทดสอบ (2567-2569) มีลักษณะผันผวนสลับทิศทางค่อนข้างบ่อย รวมถึงช่วงปรับฐานรุนแรงในกลางปี 2568 ที่ตามด้วยการฟื้นตัวอย่างรวดเร็ว ค่าคงที่การเลื่อนที่ประมาณจากชุดฝึกในอดีตจึงไม่สอดคล้องกับทิศทางการเปลี่ยนแปลงราคาจริงในชุดทดสอบเสมอไป และเมื่อค่าคงที่ดังกล่าวถูกสะสมซ้ำในการพยากรณ์แบบวันต่อวัน (rolling one-step) ความคลาดเคลื่อนเชิงระบบ (systematic bias) จึงสะสมมากกว่าแบบจำลอง naive persistence ที่อ้างอิงราคาล่าสุดโดยตรงโดยไม่มีองค์ประกอบค่าคงที่ปนเปื้อน ข้อสังเกตนี้ชี้ให้เห็นว่าสำหรับอนุกรมราคาหลักทรัพย์ที่มีค่าเฉลี่ยผลตอบแทนใกล้ศูนย์และมีความผันผวนสูงเช่น SET Index การเพิ่มพารามิเตอร์แม้เพียงตัวเดียวเข้าไปในแบบจำลองพื้นฐานก็อาจส่งผลเสียต่อความแม่นยำเชิงระดับราคาได้ หากพารามิเตอร์นั้นประมาณค่าจากช่วงเวลาที่มิพลวัตแตกต่างจากช่วงทดสอบ

3.3 ผลการเปรียบเทียบความแม่นยำเชิงทิศทาง

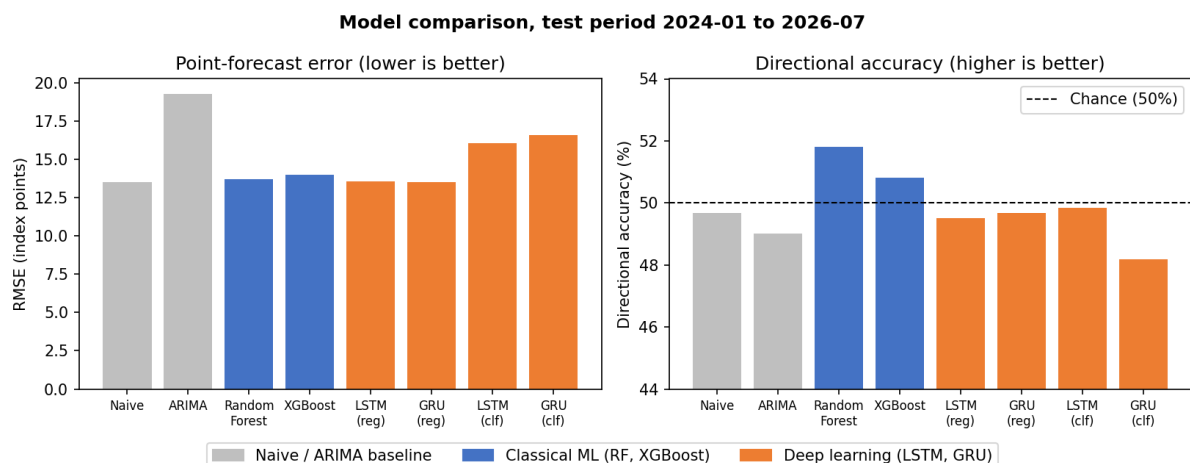
ในมิติความแม่นยำเชิงทิศทางซึ่งเป็นตัวชี้วัดหลักของงานวิจัยนี้ ผลลัพธ์แสดงรูปแบบที่ชัดเจนกว่ามิติความคลาดเคลื่อนเชิงระดับราคา โดย Random Forest ให้ความแม่นยำเชิงทิศทางสูงสุดที่ร้อยละ 51.8 รองลงมาคือ XGBoost ที่ร้อยละ 50.8 ซึ่งตัวเลขของทั้งสองแบบจำลองสูงกว่าเกณฑ์การสุ่มทายที่ร้อยละ 50 และสูงกว่า naive persistence ที่ร้อยละ 49.7 ในเชิงตัวเลข อย่างไรก็ตาม ผลการทดสอบทวินาม (binomial test) เทียบกับเกณฑ์ร้อยละ 50 ดังแสดงในตารางที่ 3 ชี้ว่าความแตกต่างดังกล่าวยังไม่ถึง

ระดับนัยสำคัญทางสถิติที่ 0.05 สำหรับแบบจำลองใดเลยในทั้งแปดแบบจำลอง (p อยู่ในช่วง 0.39-0.97 และช่วงความเชื่อมั่นร้อยละ 95 ของทุกแบบจำลองครอบคลุมค่าร้อยละ 50) ในทางตรงกันข้าม แบบจำลอง LSTM และ GRU ทั้ง 4 รูปแบบ ทั้งแบบถดถอยและแบบจำแนกประเภท ให้ความแม่นยำเชิงทิศทางอยู่ในช่วงร้อยละ 48.2-49.8 ซึ่งต่ำกว่าหรือใกล้เคียงกับ naive persistence และต่ำกว่าเกณฑ์การสุ่มทายในกรณีของ GRU แบบจำแนกประเภท (ร้อยละ 48.2) รูปที่ 4 แสดงกราฟแท่งเปรียบเทียบทั้งสองมิติของแบบจำลองทั้ง 8 แบบจำลองในภาพเดียวกัน

ตารางที่ 3 ผลการทดสอบทวินาม (binomial test) เทียบกับเกณฑ์การสุ่มทายร้อยละ 50 สำหรับความแม่นยำเชิงทิศทางของแบบจำลองทั้ง 8 แบบจำลอง บนชุดข้อมูลทดสอบ (606 วัน)

แบบจำลอง	Directional Accuracy (%)	ถูกต้อง/N	95% CI (%)	p-value (two-sided)
Random Forest	51.8	314/606	47.8 - 55.8	0.394
XGBoost	50.8	308/606	46.9 - 54.8	0.715
LSTM (classification)	49.8	302/606	45.9 - 53.8	0.968
Naive persistence	49.7	301/606	45.7 - 53.6	0.903
GRU (regression)	49.7	301/606	45.7 - 53.6	0.903
LSTM (regression)	49.5	300/606	45.5 - 53.5	0.839
ARIMA(0,1,0)	49.0	297/606	45.1 - 53.0	0.655
GRU (classification)	48.2	292/606	44.2 - 52.2	0.394

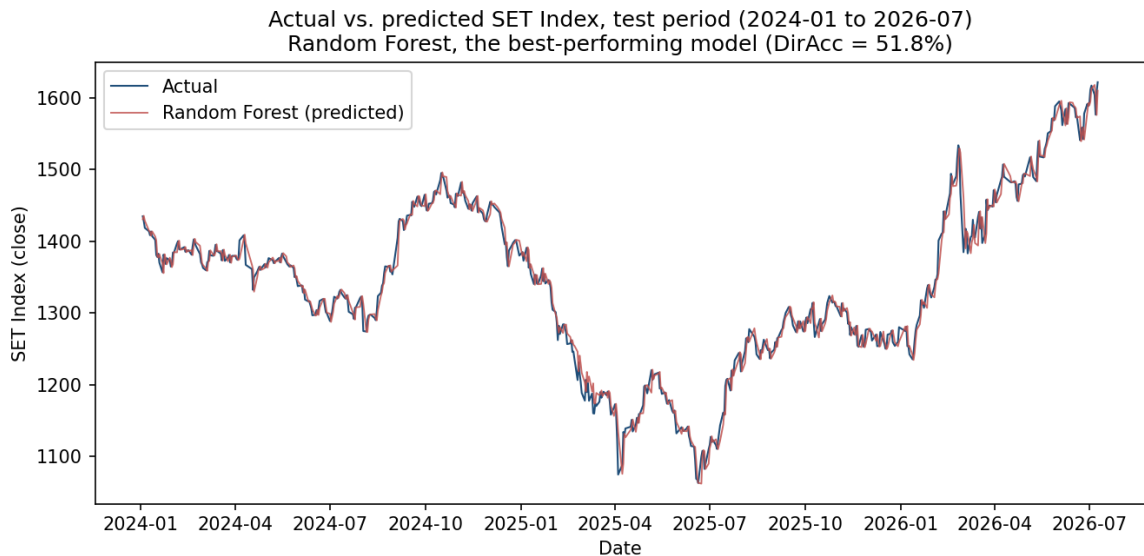
หมายเหตุ: ช่วงความเชื่อมั่นร้อยละ 95 คำนวณด้วยวิธี Wilson score interval จำนวนครั้งที่พยากรณ์ถูกต้อง (ถูกต้อง/N) คำนวณย้อนกลับจากค่าความแม่นยำเชิงทิศทางที่รายงานในตารางที่ 2 [29]



รูปที่ 4 เปรียบเทียบ RMSE (ซ้าย) และ Directional Accuracy (ขวา) ของแบบจำลองทั้ง 8 แบบจำลอง บนชุดข้อมูลทดสอบ

รูปที่ 4 แสดงให้เห็นรูปแบบที่สอดคล้องกับตารางที่ 2 อย่างชัดเจน คือแบบจำลองต้นไม้ตัดสินใจ (Random Forest และ XGBoost) เป็นกลุ่มเดียวที่อยู่เหนือเส้นเกณฑ์การสุ่มทายร้อยละ 50 อย่างสม่ำเสมอในแผงขวา ขณะที่แบบจำลองโครงข่ายประสาทเทียมเชิงลึกทั้งหมดอยู่ต่ำกว่าหรือใกล้เคียงเส้นดังกล่าว ข้อสังเกตที่น่าสนใจคือผลลัพธ์นี้คงที่ไม่่ว่าจะฝึกแบบจำลอง LSTM/GRU ให้พยากรณ์ผลตอบแทนด้วยฟังก์ชันการสูญเสียแบบ mean squared

error หรือพยากรณ์ทิศทางโดยตรงด้วยฟังก์ชัน binary cross-entropy ก็ตาม ซึ่งช่วยลดความเป็นไปได้ที่ผลลัพธ์ต่ำของแบบจำลองแบบถดถอยจะเกิดจากความไม่สอดคล้องกันระหว่างฟังก์ชันการสูญเสียที่ใช้ฝึกกับตัวชี้วัดที่ใช้ประเมินผล เพื่อแสดงให้เห็นคุณภาพการพยากรณ์ของแบบจำลองที่ดีที่สุดในการเลือกรูปธรรม รูปที่ 5 เปรียบเทียบราคาจริงกับราคาที่พยากรณ์โดย Random Forest ตลอดชุดข้อมูลทดสอบ



รูปที่ 5 ราคาจริงเทียบกับราคาที่พยากรณ์โดย Random Forest บนชุดข้อมูลทดสอบ (DirAcc = 51.8%)

จากรูปที่ 5 เส้นราคาที่พยากรณ์โดย Random Forest เกือบติดกับเส้นราคาจริงอย่างใกล้ชิดตลอดทั้งช่วงทดสอบ รวมถึงในช่วงที่ดัชนีปรับฐานรุนแรงในช่วงกลางปี 2568 ซึ่งสอดคล้องกับค่า RMSE ที่ต่ำ อย่างไรก็ตาม ความใกล้ชิดของเส้นทั้งสองส่วนหนึ่งเป็นผลจากธรรมชาติของราคาดัชนีที่เปลี่ยนแปลงทีละน้อยในแต่ละวัน (autocorrelation สูงของระดับราคา) มิใช่เครื่องหมายความสามารถในการพยากรณ์ทิศทางที่แม่นยำแต่เพียงอย่างเดียว ตัวชี้วัดความแม่นยำเชิงทิศทางในตารางที่ 2 จึงยังคงเป็นเกณฑ์ตัดสินที่สำคัญกว่าในการประเมินประโยชน์เชิงปฏิบัติของแบบจำลอง

3.4 การอภิปรายผลและกลไกเชิงทฤษฎี

ผลการศึกษาที่ Random Forest และ XGBoost ให้ความแม่นยำเชิงทิศทางสูงกว่าแบบจำลอง LSTM และ GRU อย่างสม่ำเสมอ ชวนให้ตั้งคำถามเชิงกลไกว่าเหตุใดแบบจำลองที่มีความสามารถในการแทนค่า

(representational capacity) สูงกว่าอย่างโครงข่ายประสาทเทียมเชิงลำดับ จึงไม่สามารถให้ผลพยากรณ์ที่ดีกว่าแบบจำลองต้นไม้ตัดสินใจที่มีโครงสร้างเรียบง่ายกว่าได้ คำอธิบายที่เป็นไปได้ประการแรกเกี่ยวข้องกับความสามารถของแบบจำลองต้นไม้ในการจับรูปแบบไม่เชิงเส้นแบบมีเงื่อนไข (conditional nonlinearity) ผ่านกลไกการแบ่ง (split) ตัวแปร ซึ่งเหมาะสมกับลักษณะของตัวชี้วัดเชิงเทคนิคอย่าง RSI และ MACD ที่มักให้สัญญาณชัดเจนเฉพาะเมื่อค่าอยู่ในบริเวณสุดขั้ว (extreme region) มากกว่าจะมีความสัมพันธ์เชิงเส้นต่อเนื่องกับทิศทางราคา แนวคิดนี้สอดคล้องกับข้อค้นพบของงานวิจัยที่ใช้ Random Forest ร่วมกับตัวชี้วัดเชิงเทคนิคจำพวก RSI และ stochastic oscillator ในการพยากรณ์ทิศทางราคาหุ้น ซึ่งรายงานว่แบบจำลองต้นไม้สามารถให้ค่าความผิดพลาดนอกกลุ่มตัวอย่าง (out-of-bag error) ในระดับที่แข่งขันได้ [19] และกับงานวิจัยที่พบว่า การวิเคราะห์ความสำคัญของตัวแปร (variable importance) จากแบบจำลองต้นไม้ยังคงมี

ประโยชน์แม้ในกรอบการทำงานแบบผสมกับโครงข่ายประสาทเทียม [30]

คำอธิบายประการที่สองเกี่ยวข้องกับขนาดตัวอย่างของข้อมูล การฝึกแบบจำลองในการศึกษาที่ใช้ข้อมูลอนุกรมเวลาเดียวจำนวนประมาณ 2,860 วัน ซึ่งมีขนาดค่อนข้างเล็กเมื่อเทียบกับปริมาณข้อมูลที่โครงข่ายประสาทเทียมเชิงลึกโดยทั่วไปต้องการเพื่อเรียนรู้รูปแบบเชิงเวลาที่ซับซ้อนได้อย่างมีประสิทธิภาพ แบบจำลอง LSTM และ GRU จึงมีแนวโน้มเรียนรู้ไปสู่ค่าเฉลี่ยแบบไม่มีเงื่อนไข (unconditional mean ใกล้ศูนย์) แทนที่จะพบสัญญาณที่มีประโยชน์เพิ่มเติมเหนือกว่าตัวแปร lag และ window ที่สร้างไว้ล่วงหน้า ข้อสังเกตนี้สอดคล้องกับหลักฐานจากการแข่งขันพยากรณ์ M4 ที่พบว่าแบบจำลอง ML แบบล้วนให้ผลแย่กว่าการรวมวิธีการทางสถิติแบบง่าย โดยเฉพาะภายใต้เงื่อนไขข้อมูลจำกัดและมีสัญญาณรบกวนสูง [8], [9] เช่นเดียวกับบทวิเคราะห์ด้านการพยากรณ์ในเวลาต่อมาที่สรุปในทิศทางเดียวกัน [10]

คำอธิบายประการที่สามเกี่ยวข้องกับความสัมพันธ์กับสมมติฐานตลาดที่มีประสิทธิภาพ เนื่องจากค่า RMSE และ MAPE ของแบบจำลองเกือบทั้งหมดใกล้เคียงกับ naive persistence อย่างมาก ผลลัพธ์นี้จึงสอดคล้องกับกรอบทฤษฎีของ Fama (ค.ศ. 1965, 1970) ที่ระบุว่าการเปลี่ยนแปลงราคาหลักทรัพย์ระยะสั้นมีองค์ประกอบสุ่มเด่นแบบเป็นส่วนใหญ่ [12], [11] และกับหลักฐานเชิงประจักษ์ในภูมิภาคอาเซียนที่พบว่าตลาดหุ้นในกลุ่มนี้รวมถึงไทยมีแนวโน้มสอดคล้องกับประสิทธิภาพระดับอ่อนในระยะยาว แม้จะมีความไม่มีประสิทธิภาพหลงเหลืออยู่บ้างในระยะสั้น [4] ความไม่มีประสิทธิภาพระยะสั้นในระดับเล็กน้อยนี้เองที่อาจเป็นช่องว่างซึ่งแบบจำลองต้นไม่ตัดสินใจสามารถจับสัญญาณเชิงทิศทางได้ดีกว่าการสุ่มทายเล็กน้อย ในขณะที่ยังไม่มากพอที่แบบจำลอง sequence ที่ซับซ้อนกว่าจะสกัดสัญญาณเพิ่มเติมออกมาได้อย่างมีนัยสำคัญ

ในแง่ของนัยเชิงปฏิบัติต่อการลงทุน คำถามที่สำคัญไม่แพ้กันคือความแม่นยำเชิงทิศทางร้อยละ 51.8 ของ Random Forest เพียงพอสำหรับการสร้างผลตอบแทนที่เป็นบวกจากกลยุทธ์การซื้อขายจริงหรือไม่ ผลการทดสอบทวินามในตารางที่ 3 ที่ชี้ว่าความได้เปรียบดังกล่าวยังไม่มีนัยสำคัญทางสถิติ บ่งชี้ในเบื้องต้นว่าคำตอบมีแนวโน้มเป็น

ลบมากกว่าบวก เนื่องจากผลตอบแทนสุทธิจากกลยุทธ์การซื้อขายด้วยสัญญาณทิศทางยังต้องหักลบด้วยต้นทุนธุรกรรม (transaction cost) ค่าคอมมิชชัน และส่วนต่างราคาเสนอซื้อเสนอขาย (bid-ask spread) ซึ่งในทางปฏิบัติมักมีขนาดใกล้เคียงหรือมากกว่าความได้เปรียบเชิงทิศทางร้อยละ 1.8 ที่พบในงานวิจัยนี้ ดังนั้นแม้ Random Forest จะให้ผลพยากรณ์ทิศทางที่ดีกว่าการสุ่มทายในเชิงตัวเลข แต่ยังไม่มีหลักฐานเพียงพอในงานวิจัยนี้จะสรุปว่าความได้เปรียบดังกล่าวสามารถแปลงเป็นผลตอบแทนสุทธิที่เป็นบวกได้อย่างสม่ำเสมอในทางปฏิบัติ การประเมินความคุ้มค่าเชิงเศรษฐกิจที่แท้จริงจำเป็นต้องอาศัยการจำลองกลยุทธ์การซื้อขายย้อนหลัง (backtesting) ที่คำนึงถึงต้นทุนธุรกรรมอย่างครบถ้วน ซึ่งอยู่นอกขอบเขตของงานวิจัยนี้

เมื่อเปรียบเทียบกับผลการศึกษาก่อนหน้าที่ใช้แบบจำลองโครงข่ายประสาทเทียมเชิงลึกกับตลาดหุ้นไทย ผลการศึกษานี้มีความแตกต่างอย่างมีนัยสำคัญจากงานที่รายงานความแม่นยำเชิงทิศทางของดัชนี SET50 สูงถึงร้อยละ 82.94 ด้วยโครงข่ายประสาทเทียมสามชั้นซ่อนบนข้อมูลช่วงปี 2558-2562 [7] ความแตกต่างนี้อาจเกิดจากหลายปัจจัยประกอบกัน ได้แก่ความแตกต่างของดัชนีที่ศึกษา (SET50 ซึ่งเป็นดัชนีหุ้นขนาดใหญ่ 50 ตัวเทียบกับ SET Index ซึ่งเป็นดัชนีรวมตลาด) ความแตกต่างของช่วงเวลาข้อมูล (ก่อนวิกฤตโควิด-19 เทียบกับข้อมูลที่ครอบคลุมถึงปี 2569) และความแตกต่างของวิธีการแบ่งชุดข้อมูลและการป้องกัน look-ahead bias ผลลัพธ์ที่แตกต่างกันมากระหว่างสองการศึกษาชี้ให้เห็นว่าความแม่นยำเชิงทิศทางของแบบจำลอง DL ในตลาดหุ้นไทยอาจมีความอ่อนไหวสูงต่อการเลือกช่วงเวลาและระเบียบวิธีการทดลอง จึงยังไม่ควรสรุปโดยรวมว่าแบบจำลอง DL ให้ผลดีหรือแย่กว่า ML แบบดั้งเดิมในทุกบริบทของตลาดไทย ผลลัพธ์นี้กลับสอดคล้องมากกว่ากับข้อค้นพบในตลาดอื่นที่ระบุว่า LSTM และ GRU มักให้ผลลัพธ์ใกล้เคียงกันโดยไม่มีแบบจำลองใดเหนือกว่าอย่างชัดเจน [22], [23] ซึ่งบ่งชี้ว่าความแตกต่างระหว่างสถาปัตยกรรม sequence model ด้วยกันเองมีน้อยกว่าความแตกต่างระหว่างตระกูลแบบจำลอง (tree-based เทียบกับ sequence-based) อย่างมีนัยสำคัญ

ในด้านความมั่นคงของผลลัพธ์ (robustness) งานวิจัยนี้ได้ตรวจสอบว่าผลลัพธ์ที่ต่ำของแบบจำลอง LSTM/GRU มิได้เป็นผลจากการเลือกฟังก์ชันการสูญเสียที่

ไม่เหมาะสม โดยฝึกแบบจำลองทั้งในรูปแบบถดถอยด้วยฟังก์ชัน mean squared error และรูปแบบจำแนกประเภทด้วยฟังก์ชัน binary cross-entropy ซึ่งให้ผลความแม่นยำเชิงทิศทางในทิศทางเดียวกัน (ร้อยละ 48.2-49.8 ทั้งสี่รูปแบบ) นอกจากนี้ชุดข้อมูลทดสอบที่ใช้ยังครอบคลุมทั้งช่วงตลาดขาลงและขาขึ้นในสัดส่วนใกล้เคียงกันดังแสดงในรูปที่ 3 จึงลดโอกาสที่ผลลัพธ์จะถูกขับเคลื่อนโดยแนวโน้มทิศทางเดียวของตลาดในช่วงทดสอบ อย่างไรก็ตาม ผลลัพธ์ที่ได้ยังจำกัดอยู่ที่สถาปัตยกรรม LSTM/GRU ขนาดค่อนข้างเล็ก (32 และ 16 หน่วย) และหน้าต่างข้อมูลย้อนหลังคงที่ 20 วัน จึงยังไม่อาจสรุปโดยรวมได้ว่าโครงข่ายประสาทเทียมเชิงลึกทุกสถาปัตยกรรมจะให้ผลลัพธ์ในทิศทางเดียวกันนี้เสมอไป

4. สรุปผลการวิจัย

งานวิจัยนี้เปรียบเทียบความสามารถในการพยากรณ์ดัชนี SET Index รายวันของแบบจำลอง 8 ชนิด ครอบคลุมแบบจำลองพื้นฐาน สถิติเชิงเวลา การเรียนรู้ของเครื่อง และโครงข่ายประสาทเทียมเชิงลึก ผลการศึกษาตามวัตถุประสงค์ทั้งสามข้อสรุปได้ดังนี้ ประการแรก ในมิติความคลาดเคลื่อนเชิงระดับราคา แบบจำลองส่วนใหญ่ให้ค่า RMSE ในช่วง 13.5-14.0 จุด และ MAPE ในช่วงร้อยละ 0.75-0.77 ใกล้เคียงกับ naive persistence ยกเว้น ARIMA(0,1,0) ที่ให้ค่าคลาดเคลื่อนสูงกว่าอย่างชัดเจน (RMSE เท่ากับ 19.28 จุด) ประการที่สอง ในมิติความแม่นยำเชิงทิศทาง Random Forest ให้ผลดีที่สุดที่ร้อยละ 51.8 รองลงมาคือ XGBoost ที่ร้อยละ 50.8 ขณะที่ LSTM และ GRU ทั้ง 4 รูปแบบให้ผลอยู่ในช่วงร้อยละ 48.2-49.8 ซึ่งต่ำกว่าหรือใกล้เคียงกับเกณฑ์การสุ่มทายและ naive persistence ประการที่สาม ผลการตรวจสอบความมั่นคงยืนยันว่าความซับซ้อนของแบบจำลอง โดยเฉพาะแบบจำลองโครงข่ายประสาทเทียมเชิงลำดับ มิได้รับประกันความแม่นยำในการพยากรณ์ทิศทางที่สูงขึ้น ไม่ว่าจะฝึกด้วยฟังก์ชันการสูญเสียรูปแบบใดก็ตาม

ในด้านข้อเสนอแนะเชิงปฏิบัติ ผลการศึกษาชี้ให้เห็นว่าสำหรับงานพยากรณ์ทิศทางดัชนี SET Index รายวัน

นักวิจัยและผู้ปฏิบัติงานด้านการลงทุนควรพิจารณาแบบจำลองต้นไม้ตัดสินใจแบบรวมกลุ่มอย่าง Random Forest หรือ XGBoost เป็นทางเลือกเบื้องต้นก่อนการลงทุนทรัพยากรเชิงคำนวณและเวลาไปกับการพัฒนาแบบจำลองโครงข่ายประสาทเทียมเชิงลึกที่ซับซ้อนกว่า เนื่องจากหลักฐานเชิงประจักษ์จากทั้งงานวิจัยนี้และการแข่งขันพยากรณ์ระดับนานาชาติชี้ตรงกันว่าความซับซ้อนของแบบจำลองมิได้แปรผันตรงกับความแม่นยำเสมอไป โดยเฉพาะภายใต้เงื่อนไขข้อมูลอนุกรมเวลาเดียวที่มีขนาดจำกัด [8], [10] ทั้งนี้ ข้อเสนอแนะดังกล่าวอาศัยหลักฐานเชิงประจักษ์ที่มีขนาดผลลัพธ์ค่อนข้างเล็กและยังไม่ถึงระดับนัยสำคัญทางสถิติตามผลการทดสอบทวินามในตารางที่ 3 จึงควรตีความว่าเป็นข้อเสนอแนะเชิงประสิทธิภาพการคำนวณ (computational efficiency) มากกว่าการรับประกันผลตอบแทนเชิงบวกจากการลงทุน

สำหรับหน่วยงานกำกับดูแลและนักวิชาการด้านตลาดทุน ผลการศึกษาที่ตัวชี้วัดความคลาดเคลื่อนเชิงระดับราคาของแทบทุกแบบจำลองใกล้เคียงกับ naive persistence อย่างมาก เป็นหลักฐานเชิงประจักษ์เพิ่มเติมที่สนับสนุนสมมติฐานตลาดที่มีประสิทธิภาพระดับอ่อนสำหรับตลาดหลักทรัพย์ไทย ซึ่งมีนัยเชิงนโยบายต่อการกำกับดูแลผลิตภัณฑ์การลงทุนที่อ้างความสามารถในการพยากรณ์ราคาแบบเจาะจง (active forecasting-based products) ว่าควรได้รับการตรวจสอบหลักฐานเชิงประจักษ์อย่างเข้มงวด เนื่องจากความได้เปรียบเชิงทิศทางที่พบในงานวิจัยนี้แม้จะมีนัยสำคัญในเชิงเปรียบเทียบ แต่มีขนาดผลลัพธ์ (effect size) ค่อนข้างเล็ก (ร้อยละ 51.8 เทียบกับร้อยละ 50)

สำหรับผู้พัฒนาระบบพยากรณ์ในภาคการเงิน ผลการศึกษาชี้ให้เห็นความสำคัญของการรายงานผลควบคู่กันทั้งมิติความคลาดเคลื่อนเชิงระดับราคาและมิติความแม่นยำเชิงทิศทาง เนื่องจากแบบจำลองที่มีค่า RMSE ต่ำใกล้เคียงกันอาจมีความแม่นยำเชิงทิศทางที่แตกต่างกันอย่างมีนัยสำคัญดังปรากฏในตารางที่ 2 การประเมินแบบจำลอง

พยากรณ์ราคาหลักทรัพย์จึงควรกำหนดตัวชี้วัดให้สอดคล้องกับวัตถุประสงค์การใช้งานจริงตั้งแต่ขั้นตอนออกแบบงานวิจัย มิใช่เลือกใช้ตัวชี้วัดใดตัวชี้วัดหนึ่งเพียงลำพัง

ข้อจำกัดของงานวิจัย

งานวิจัยนี้มีข้อจำกัดที่ควรระบุไว้สองประการ ประการแรก การศึกษานี้จำกัดขอบเขตอยู่ที่การพยากรณ์แบบ 1 วันล่วงหน้า (one-step-ahead) และยังไม่ได้ทดสอบขอบเขตเวลาพยากรณ์ที่ยาวขึ้น เช่น 5 วันหรือ 20 วันล่วงหน้า ซึ่งรูปแบบผลลัพธ์เชิงเปรียบเทียบระหว่างแบบจำลองอาจแตกต่างกันไปจากที่พบในการศึกษานี้ ประการที่สอง สถาปัตยกรรมโครงข่ายประสาทเทียมเชิงลึกที่ใช้มีขนาดค่อนข้างเล็กและมีได้ผ่านกระบวนการปรับแต่ง hyperparameter อย่างเข้มข้นด้วยวิธีการเชิงเมทาฮิวริสติก เทียบเท่ากับแบบจำลองต้นไม้ตัดสินใจ จึงยังเปิดช่องให้มีการตั้งคำถามว่าผลลัพธ์จะเปลี่ยนแปลงหรือไม่หากมีการปรับแต่งสถาปัตยกรรมอย่างละเอียดมากขึ้น

ข้อเสนอแนะสำหรับงานวิจัยในอนาคต

ทิศทางการวิจัยในอนาคตควรครอบคลุมสามประเด็นหลัก ได้แก่ การขยายขอบเขตเวลาพยากรณ์ให้ครอบคลุมหลายช่วงเวลาล่วงหน้าเพื่อทดสอบความมั่นคงของรูปแบบผลลัพธ์ การทดลองใช้ตัวแปรคุณลักษณะเพิ่มเติมนอกเหนือจากตัวชี้วัดเชิงเทคนิค เช่น ข้อมูลอารมณ์ตลาด (market sentiment) จากข่าวและสื่อสังคมออนไลน์ หรือดัชนีตลาดต่างประเทศที่เกี่ยวข้อง และการขยายการเปรียบเทียบไปยังแบบจำลองสถาปัตยกรรมใหม่ เช่น Transformer [24] หรือแบบจำลองผสมระหว่างต้นไม้ตัดสินใจกับโครงข่ายประสาทเทียม

โดยสรุป งานวิจัยนี้ได้แสดงให้เห็นว่าการเปรียบเทียบแบบจำลองพยากรณ์อย่างเป็นระบบภายใต้กรอบข้อมูลและการประเมินผลเดียวกันเป็นขั้นตอนสำคัญที่ช่วยให้ทั้งแวดวงวิชาการและผู้ปฏิบัติงานสามารถเลือกใช้เครื่องมือพยากรณ์ตลาดหุ้นไทยได้อย่างมีประสิทธิภาพบนพื้นฐานของหลักฐานเชิงประจักษ์ มากกว่าความเชื่อที่ว่าแบบจำลองที่ซับซ้อนกว่าย่อมให้ผลดีกว่าเสมอไป

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณมหาวิทยาลัยกรุงเทพและคณะเศรษฐศาสตร์และการลงทุนที่เอื้ออำนวยความสะดวกด้านทรัพยากรและสภาพแวดล้อมในการทำวิจัยครั้งนี้ ตลอดจนขอขอบพระคุณบรรณาธิการและผู้ทรงคุณวุฒิที่ได้กรุณาสละเวลาพิจารณาและให้ข้อเสนอแนะอันเป็นประโยชน์ต่อการปรับปรุงบทความให้มีความสมบูรณ์ยิ่งขึ้น

เอกสารอ้างอิง

- [1] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005-2019," *Applied Soft Computing*, vol. 90, pp. 106181, May. 2020, doi: 10.1016/j.asoc.2020.106181.
- [2] S. Ahmed, M. M. Alshater, A. El Ammari, and H. Hammami, "Artificial intelligence and machine learning in finance: A bibliometric review," *Research in International Business and Finance*, vol. 61, pp. 101646, Oct. 2022, doi: 10.1016/j.ribaf.2022.101646.
- [3] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *Journal of Big Data*, vol. 7, pp. 20, 2020, doi: 10.1186/s40537-020-00299-5.
- [4] F. Guidi and R. Gupta, "Market efficiency in the ASEAN region: Evidence from multivariate and cointegration tests," *Applied Financial Economics*, vol. 23, no. 4, pp. 265-274, 2013, doi: 10.1080/09603107.2012.718064.
- [5] M. Inthachot, V. Boonjing, and S. Intakosum, "Artificial neural network and genetic algorithm hybrid intelligence for predicting Thai stock price index trend," *Computational Intelligence and Neuroscience*, vol. 2016, Art. no. 3045254, 2016, doi: 10.1155/2016/3045254.

- [6] K. Suphawan, R. Kardkasem, and K. Chaisee, "A Gaussian process regression model for forecasting stock exchange of Thailand," *Trends in Sciences*, vol. 19, no. 6, pp. 3045, Mar. 2022, doi: 10.48048/tis.2022.3045.
- [7] M. Inthachot, V. Boonjing, and S. Intakosum, "Predicting Thai stock index trend using deep neural network based on technical indicators," *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 2, pp. 428-435, Mar. 2025, doi: 10.53894/ijirss.v8i2.5191.
- [8] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 Competition: Results, findings, conclusion and way forward," *International Journal of Forecasting*, vol. 34, no. 4, pp. 802-808, Oct. 2018, doi: 10.1016/j.ijforecast.2018.06.001.
- [9] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 Competition: 100,000 time series and 61 forecasting methods," *International Journal of Forecasting*, vol. 36, no. 1, pp. 54-74, Jan. 2020, doi: 10.1016/j.ijforecast.2019.04.014.
- [10] F. Petropoulos, S. Makridakis, and N. Stylianou, "Forecasting: theory and practice," *International Journal of Forecasting*, vol. 38, no. 3, pp. 705-871, 2022, doi: 10.1016/j.ijforecast.2021.11.001.
- [11] E. F. Fama, "The behavior of stock market prices," *The Journal of Business*, vol. 38, no. 1, pp. 34-105, 1965, doi: 10.1086/294743.
- [12] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance*, vol. 25, no. 2, pp. 383-417, May. 1970, doi: 10.1111/j.1540-6261.1970.tb00518.x.
- [13] R. S. Tsay, *Analysis of Financial Time Series*, 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, 2010.
- [14] G. Ji, J. Yu, K. Hu, J. Xie, and X. Ji, "An adaptive feature selection schema using improved technical indicators for predicting stock price movements," *Expert Systems with Applications*, vol. 200, p. 116941, Aug. 2022, doi: 10.1016/j.eswa.2022.116941.
- [15] H. H. Htun, M. Biehl, and N. Petkov, "Survey of feature selection and extraction techniques for stock market prediction," *Financial Innovation*, vol. 9, p. 26, Jan. 2023, doi: 10.1186/s40854-022-00441-7.
- [16] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723, Dec. 1974, doi: 10.1109/TAC.1974.1100705.
- [17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.
- [18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceeding of 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.
- [19] L. Khaidem, S. Saha, and S. R. Dey, "Predicting the direction of stock market prices using random forest," arXiv:1605.00003, 2016, doi: 10.48550/arXiv.1605.00003.
- [20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [21] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceeding of 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724-1734, doi: 10.3115/v1/D14-1179.
- [22] Y. Gao, R. Wang, and E. Zhou, "Stock prediction based on optimized LSTM and GRU models,"

- Scientific Programming*, vol. 2021, Art. no. 4055281, Sep. 2021, doi: 10.1155/2021/4055281.
- [23] N. K. Khairunisa and P. Hendikawati, "Long short-term memory and gated recurrent unit modeling for stock price forecasting," *Jurnal Matematika, Statistika dan Komputasi*, vol. 21, no. 1, pp. 321-333, Sep. 2024, doi: 10.20956/j.v21i1.35930.
- [24] B. Lim, S. Ö. Arik, N. Loeff, and T. Pfister, "Temporal Fusion Transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748-1764, Oct. 2021, doi: 10.1016/j.ijforecast.2021.03.012.
- [25] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679-688, Oct. 2006, doi: 10.1016/j.ijforecast.2006.03.001.
- [26] Y. Kara, M. A. Boyacioglu, and O. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5311-5319, May. 2011, doi: 10.1016/j.eswa.2010.10.027.
- [27] M. Costantini, J. Crespo Cuaresma, and J. Hlouskova, "Forecasting errors, directional accuracy and profitability of currency trading: The case of EUR/USD exchange rate," *Journal of Forecasting*, vol. 35, no. 7, pp. 652-668, Mar. 2016, doi: 10.1002/for.2398.
- [28] Ö. İcan and T. B. Çelik, "Stock market prediction performance of neural networks: A literature review," *International Journal of Economics and Finance*, vol. 9, no. 11, pp. 100-108, Oct. 2017, doi: 10.5539/ijef.v9n11p100.
- [29] M. H. Pesaran and A. Timmermann, "A simple nonparametric test of predictive performance," *Journal of Business & Economic Statistics*, vol. 10, no. 4, pp. 461-465, Oct. 1992, doi: 10.1080/07350015.1992.10509922.
- [30] H. J. Park, Y. Kim, and H. Y. Kim, "Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework," *Applied Soft Computing*, vol. 114, pp. 108106, Jan. 2022, doi: 10.1016/j.asoc.2021.108106.