

ผลกระทบของข้อมูลโควิด-19 ต่อการคาดการณ์ผลตอบแทนของหุ้นสามัญ

โดยใช้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง

The Impact of COVID-19 data in Stock Return Estimation by Machine Learning

วิศรุต แก้วมหา¹ และวิรัช ปัญญาฉัตรพร²Witsarut Kaewmaha¹ and Varis Punyachatporn²

วิทยาลัยการจัดการ

มหาวิทยาลัยมหิดล ประเทศไทย^{1,2}

College of Management

Mahidol University, Thailand^{1,2}E-mail: witsarut.kmh@gmail.com¹, varis116@hotmail.com²

วันที่รับบทความ 22 ธันวาคม 2565

วันที่แก้ไขบทความ 25 ธันวาคม 2565

วันที่ตอบรับบทความ 26 ธันวาคม 2565

บทคัดย่อ

งานวิจัยนี้พยากรณ์ผลตอบแทนของหุ้นสามัญในตลาดหลักทรัพย์แห่งประเทศไทยโดยใช้ระบบ Machine Learning ประกอบกับข้อมูลและปัจจัยต่าง ๆ ในอดีตที่มีผลต่อราคา รวมทั้งข้อมูลเหตุการณ์สำคัญอย่าง ข้อมูลเกี่ยวกับ โควิด-19 โดยคาดการณ์ผลตอบแทน 1 วัน 1 เดือน และ 3 เดือน โดยใช้อัลกอริทึมที่แตกต่างกันของ Machine Learning 3 อัลกอริทึมคือ Artificial neural network (ANN) , Long Short-Term Memory (LSTM) และ Random Forest (RF) เพื่อทดสอบว่าอัลกอริทึมใด สามารถสร้างแบบจำลองในการพยากรณ์ผลตอบแทนของหุ้นโดยให้ค่าการพยากรณ์ที่แม่นยำที่สุด ผลการศึกษาแสดงให้เห็นถึงความสามารถในการพยากรณ์ผลตอบแทนของหุ้นโดยใช้ข้อมูลของบริษัท ได้แก่ ข้อมูลในงบการเงิน ข้อมูลอัตราส่วนทางการเงิน ข้อมูลปัจจัยทางเทคนิค ข้อมูลเศรษฐกิจมหภาค อัตราแลกเปลี่ยนเงินตรา ดัชนีหุ้น ดัชนีทองคำ และข้อมูลในอดีตของการซื้อขายหุ้น ประกอบกับการใช้ระบบ Machine Learning ส่งผลให้สามารถคาดการณ์ผลตอบแทนของหุ้นได้แม่นยำที่สุดเพื่อนำไปกำหนดกลยุทธ์ในการลงทุนที่เหมาะสม โดยจากงานวิจัยนี้แสดงให้เห็นว่าการคาดการณ์ผลตอบแทนของหุ้นในระยะสั้นนั้นการใช้ระบบ Machine Learning นั้นยังไม่เหมาะสมในการนำมาพยากรณ์ แต่หากเป็นการคาดการณ์ผลตอบแทนของหุ้นในระยะยาวแบบจำลอง Random Forest สามารถนำมาใช้ในการพยากรณ์ได้แม่นยำที่สุด

คำสำคัญ: โควิด-19 โครงข่ายประสาทเทียม ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง การคาดการณ์ผลตอบแทนหุ้นสามัญ

Abstract

This paper is to predicting stock return of equity securities in the Stock Exchange of Thailand by using machine learning with various factors including COVID-19 data. By predicting returns in 1-day, 1-month and 3-month returns and using 3 different machine learning algorithms are Artificial neural network (ANN) , Long Short-Term Memory (LSTM) and Random Forest (RF) to prove which algorithm can forecast a stock's return with the most accurate forecast. The results of the study demonstrated the ability to forecast stock returns using company data such as financial statements, financial ratio data, technical indicators, macroeconomic data. Macroeconomics data, Exchange rate, Stock Index, Gold index and Historical Data. By using machine learning, the result is able to predict the returns of stocks as accurately as possible in order to determine the right investment strategy. This research has been shown that predicting stock returns in the short-term using Machine Learning is not suitable for forecasting. But when forecasting stock returns over the long term, the Random Forest model can be used to make the most accurate predictions.

Keywords: COVID-19, Neural Network, Machine Learning, Stock Return Prediction

บทนำ

นักลงทุนในตราสารทางการเงินต่าง ๆ ตัดสินใจลงทุนบนข้อมูลข่าวสารที่ได้รับ ในอดีตการรับรู้ข่าวสารของนักลงทุนมาจากหนังสือพิมพ์ โทรทัศน์ การติดต่อสื่อสารผ่านทางโทรศัพท์ ความรวดเร็วและครอบคลุมของข่าวสารในด้านต่าง ๆ มีข้อจำกัด ซึ่งแตกต่างในปัจจุบัน การติดต่อสื่อสารที่สะดวกรวดเร็ว ทำให้การพัฒนาของการพยากรณ์ผลตอบแทนของตราสารทางการเงินนั้นพัฒนาอย่างรวดเร็วตามการพัฒนาของเทคโนโลยีสารสนเทศ ดังนั้นการศึกษากการพยากรณ์ผลตอบแทนของตราสารทุนก็มีการพัฒนาวิธีการที่หลากหลาย ซึ่งวิธีการ Machine Learning เป็นการทำให้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเองโดยใช้ข้อมูลในอดีต ในการศึกษาในครั้งนี้ได้รวมข้อมูลเกี่ยวข้องกับโควิด-19 เข้ามาศึกษาด้วย นอกจากนี้ยังเปรียบเทียบแบบจำลองสำหรับการคาดการณ์ผลตอบแทนในอนาคต เพื่อหาอัลกอริทึมที่เหมาะสมกับการพยากรณ์ผลตอบแทนของตราสารทุนได้แม่นยำที่สุดในช่วงเวลาต่าง ๆ ที่กำหนด 1 วัน 1 เดือน และ 3 เดือน โดยวัดความแม่นยำที่ใช้การการประเมินผลของโมเดล มุ่งมองเฉพาะค่าความใกล้เคียงระหว่างค่าคาดหวังกับค่าที่พยากรณ์ได้ เช่น MAE, RMSE และนอกจากตัววัดดังกล่าวแล้วตัววัดที่เรียกว่า Hit Rate ซึ่งใช้วัดความแม่นยำของทิศทางขึ้น-ลงของข้อมูล ได้ถูกนำมาใช้ร่วมด้วย โดยงานวิจัยนี้จะแสดงให้เห็นว่าโมเดล RF นั้นแม่นยำกว่าโมเดลอื่น ๆ อย่างไรในแต่ละช่วงเวลาของการพยากรณ์เมื่อเปรียบเทียบกับโมเดล ANN และ LSTM โดยเมื่อได้

โมเดลที่มีความแม่นยำในการพยากรณ์มากที่สุดแล้ว สามารถนำมาใช้ในการพยากรณ์ผลตอบแทนของตราสารทุนเพื่อเป็นแนวทางในการคัดเลือกหลักทรัพย์ว่าหลักทรัพย์ใดมีความน่าลงทุนและกำหนดกลยุทธ์ ว่าควรซื้อ-ขายหลักทรัพย์เมื่อใดหรือแม้กระทั่งนำมาใช้ในการวางแผนการลงทุนในอนุพันธ์ต่าง ๆ ไม่ว่าจะเป็นสัญญาซื้อขายล่วงหน้า (Future) หรือสัญญาสิทธิ (Option) เพื่อดูทิศทางว่าควรซื้อหรือควรขายควรกำหนดกลยุทธ์แบบใดในการลงทุนที่เหมาะสมเพื่อให้ได้ผลตอบแทนที่เหมาะสมกับความคาดหวังของนักลงทุน

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง (Theories and Literature Review)

1. ทฤษฎี (Theories)

Universal Approximation Theorem (Baker & Patil, 1998) จุดเริ่มต้นของทฤษฎีเกิดจากการที่ The Universal Approximation Theorem บอกว่า Neural Networks สามารถประมาณฟังก์ชันใดก็ได้ เพราะความหมายของทฤษฎีคือเรื่องใดก็ตามที่คำนวณเป็นฟังก์ชันได้ก็สามารถคำนวณโดยโครงข่ายประสาทเทียมได้ ซึ่งฟังก์ชันคือความสัมพันธ์ระหว่าง Input (x) และ output (y) โดยที่แต่ละค่าของ Input จะสัมพันธ์ได้เพียงแค่ค่าหนึ่งของ Output เท่านั้น โดยหากเราใช้โครงข่ายประสาทเทียมเพื่อให้ AI แยกภาพระหว่างสุนัขกับแมว นั่นฟังก์ชันที่ใช้ในการคำนวณก็จะเป็นฟังก์ชันแบบ Linear เนื่องจากการคำนวณไม่ว่าจะมีกี่ตัวแปรก็จะใช้การถ่วงน้ำหนักธรรมดาเพื่อให้ได้ผลลัพธ์ แต่ในความเป็นจริงไม่ได้มีแค่ฟังก์ชันที่เป็น Linear แต่ยังมีฟังก์ชัน Non-Linear ที่ซับซ้อนกว่าที่สามารถนำมาคำนวณโดยโครงข่ายประสาทเทียมได้ ซึ่งทฤษฎี Universal Approximation เป็นจริงได้นั้นเกิดจากการที่โครงข่ายประสาทเทียมมี Activation function เพื่อช่วยให้โครงข่ายประสาทเทียมสามารถประมาณฟังก์ชันที่เป็น Non-linear ได้ และมีหลากหลายแบบให้เลือกใช้ บาง Network ก็มีมากกว่าหนึ่งแบบแล้วแต่ดีไซน์ ซึ่งไม่มีข้อจำกัดใด ๆ (Nilson, 2019)

2. งานศึกษาในอดีต (Empirical Studies)

งานศึกษาเกี่ยวกับแบบจำลองการพยากรณ์ผลตอบแทนของหลักทรัพย์ด้วยวิธีการใช้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง (Machine Learning) มีการศึกษาอย่างแพร่หลายในหลายตลาดหลักทรัพย์ที่สำคัญ เช่น ในตลาดหลักทรัพย์อินเดีย Patel, Shah, Thakkar, and Kotecha (2015) ได้ทำการพยากรณ์มูลค่าในอนาคตของดัชนีตลาดหุ้นสองตัว ได้แก่ CNX Nifty และ S&P Bombay Stock Exchange (BSE) โดยบทความนี้เสนอวิธีการใช้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเอง (Machine Learning) เพื่อเปรียบเทียบประสิทธิภาพความแม่นยำของการพยากรณ์วิธีแบบขั้นตอนเดียว Artificial Neural Network (ANN), Random Forest (RF) และ Support Vector Regression (SVR) ผลการศึกษาพบว่าโมเดลที่รวมกันแล้วให้ความแม่นยำมากที่สุดคือ SVR-ANN และผลลัพธ์เมื่อมีการคาดการณ์ล่วงหน้าจำนวนวันมากขึ้น ค่าความผิดพลาดก็จะเพิ่มขึ้นด้วย (Basak, Kar, Saha, Khaidem, & Dey, 2019; Kim, 2003)

Jiemwiriyakul et al. (2019) การนำเสนอการคาดการณ์ผลตอบแทนในตราสารทุนในตลาดหลักทรัพย์แห่งประเทศไทยโดยใช้ Artificial neural network ประกอบกับข้อมูลและปัจจัยต่าง ๆ ในอดีตที่มีผลต่อราคา ในตราสารทุนและนำผลตอบแทนที่คาดการณ์ได้มาใช้ในการจัดรูปแบบและปรับพอร์ตโฟลิโอ

เพื่อ สร้างโอกาสการลงทุนระยะสั้น โดยทฤษฎีที่ใช้ในการจัดพอร์ตโฟลิโอ คือ การเลือกลงทุนที่จุดบนเส้น Efficient Frontier และเลือกน้ำหนักสำหรับการลงทุนโดยใช้ Maximum Sharpe Ratio เพื่อให้ผลตอบแทนของพอร์ตโฟลิโอต่อหนึ่งหน่วยความเสี่ยงสูงสุด ซึ่งแสดงให้เห็นถึงการคาดการณ์ผลตอบแทนตราสารทุนที่มีความแม่นยำจากการใช้ Machine Learning โดยเลือกเทคนิคแบบจำลองโครงข่ายประสาทเทียม (Neural Network) ซึ่งสามารถนำผลที่ได้ไปประยุกต์ใช้ประกอบการตัดสินใจในการปรับพอร์ตโฟลิโอให้กับนักลงทุนสถาบันหรือนักลงทุนอื่นๆได้

Han (2019) ได้ทำการเปรียบเทียบความแม่นยำในการทำนายผลตอบแทนของหุ้น (NASDAQ) ระหว่างอัลกอริทึม Random Forest (RF) กับ Long Short-Term Memory(LSTM) ซึ่งผลที่ได้จากการวิจัยนี้คืออัลกอริทึม LSTM ให้ผลในการพยากรณ์ผลตอบแทนของตราสารทุนได้ดีกว่า ต่อมาได้มีการวิจัยโดยการเพิ่มประสิทธิภาพพอร์ตโฟลิโอด้วยการทำนายผลตอบแทนของตราสารทุนโดยใช้ Deep Learning และ Machine Learning ซึ่งได้มีการนำอัลกอริทึมต่าง ๆ มารวมเข้ากับการทำ Mean-Variance (MV) model ซึ่งให้ผลว่า การทำนายผลตอบแทนของตราสารทุนจากอัลกอริทึม Random Forest (RF) โดยใช้ MSE, MAE และHit Rate เป็นตัววัดความแม่นยำของการพยากรณ์ ได้แม่นยำที่สุด และมีการนำผลการพยากรณ์จากการใช้อัลกอริทึม Random Forest (RF) จะทำให้การปรับพอร์ตโฟลิโอมีประสิทธิภาพมากที่สุดจากบรรดาอัลกอริทึมทั้งหมด เนื่องจากเมื่อใช้ร่วมกันแล้วทำให้ได้ผลตอบแทนรวมจากการลงทุน (Total Return) มากกว่าการใช้อัลกอริทึมอื่นทั้งหมด (Ma, Han, & Wang, 2021)

แบบจำลอง (Model)

การศึกษานี้ใช้ เทคนิคคอมพิวเตอร์เรียนรู้ด้วยตนเอง (Machine Learning) คือการทำให้ระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเองโดยใช้ข้อมูลแบ่งออกเป็น 2 ประเภทหลัก ๆ คือ

1. การเรียนรู้โดยมีผู้ช่วยสอน (Supervised Learning) เป็นการให้ระบบคอมพิวเตอร์เรียนรู้โดยนำใส่ข้อมูลตัวแปรต้น (Input) ได้แก่ตัวแปรทางเศรษฐกิจต่าง ๆ ที่ส่งผลต่อราคาและผลตอบแทนของหลักทรัพย์ตามทฤษฎี และผลลัพธ์ตัวแปรตาม (Output) ในการศึกษา คือ ผลตอบแทนของตราสารทุน จากนั้นให้ระบบคอมพิวเตอร์เรียนรู้แบบจำลองที่เชื่อมโยงระหว่าง Input และ Output เมื่อเรียนรู้เสร็จ ระบบจะพยายามทำนายผลตอบแทนของตราสารทุน ซึ่งหากผลลัพธ์ที่ทำนายได้นั้นผิด ระบบจะพยายามแก้ไขแบบจำลองที่ใช้ทำนายไปเรื่อย ๆ ตามข้อมูลที่เรป้อนเข้าเพื่อให้เกิดข้อผิดพลาดน้อยที่สุด

2. การเรียนรู้โดยไม่มีผู้ช่วยสอน (Unsupervised Learning) เป็นการให้ระบบคอมพิวเตอร์เรียนรู้ด้วยการจำแนกข้อมูล วิธีนี้เราจะใส่เพียงข้อมูลนำเข้า (Input) จากนั้นระบบคอมพิวเตอร์จะทำการจำแนกข้อมูล (Clustering) โดยวิธีนี้จะเน้นการใช้งานในรูปแบบการวิเคราะห์ข้อมูล (Analysis)

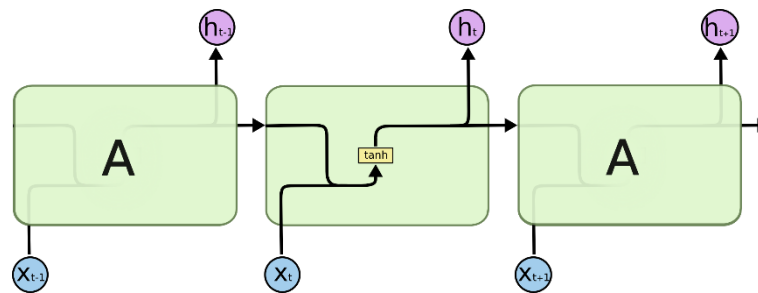
งานวิจัยนี้จะเลือกใช้ Random Forest (RF), Artificial neural network(ANN) และLong Short-Term Memory (LSTM) ใน การคาดการณ์ผลตอบแทนของตราสารทุนแต่ละตัว

แบบจำลอง Random Forest (RF) เป็นโมเดล Classification ที่มีการพัฒนามาจากต้นไม้การตัดสินใจ (Decision Tree) และ Random Forest นอกจากนี้ยังสามารถทำ Regression หรือการทำนายผลลัพธ์ในเชิงตัวเลขได้อีกด้วย ซึ่งมีหลักการคือการเรียนรู้โมเดลที่มีเป้าหมายเหมือนกันหลายๆ ครั้ง (หลาย Instance) บนข้อมูลชุดเดียวกัน โดยแต่ละครั้งของการเรียนรู้จะเลือกส่วนของข้อมูลที่เรียนรู้ไม่เหมือนกัน และคำนวณผล prediction ด้วยการ Vote Output ที่ถูกเลือกโดย Decision Tree มากที่สุด (กรณี Classification) หรือหาค่า Mean จาก Output ของแต่ละ Decision Tree (กรณี Regression) เป็นค่าเฉลี่ยว่า Input เข้าไปแบบนี้จะให้ผล Output เป็นแบบใด (Breiman, 2001) ในงานวิจัยนี้กำหนดจำนวนการสร้าง Decision Tree โดยกำหนดจำนวน คือ 305 ต้น เพื่อสุ่มตัวอย่างข้อมูล โดยการสุ่มข้อมูลตัวอย่าง (Bootstrapping หรือการสร้างต้นไม้หลาย ๆ ต้นไม่ให้ซ้ำกัน) จาก Data set ที่เป็นข้อมูลตัวแปรต้น (Input data) ให้ได้ข้อมูลออกมา 305 ชุดตามจำนวนชุดข้อมูลของตัวแปรทั้งหมดที่ไม่เหมือนกัน ตามจำนวน Decision Tree ใน Random Forest โดยที่ต้อง Minimize ค่าของ Cost Function ให้น้อยที่สุด (Regression - MSE) (Loh, 2011)

โครงข่ายประสาทเทียม (Artificial Neural Network, “ANN”) เป็นระบบคอมพิวเตอร์เรียนรู้ได้ด้วยตนเองโดยใช้ข้อมูล (Machine Learning) เป็นแนวคิดซึ่งจำลองมาจากรูปแบบการประมวลผลของสมองมนุษย์ ซึ่งแบบจำลองนี้เริ่มต้นจากป้อนข้อมูลตัวแปรต้น (Input Node) หรือปัจจัยต่าง ๆ ที่มีผลต่อผลตอบแทนของตราสารทุนแต่ละตัว และผลลัพธ์ตัวแปรตาม (Output Node) คือผลตอบแทน 1 วัน 1 เดือน และ 3 เดือนของตราสารเพื่อหารูปแบบสร้างเป็นแบบจำลองไว้ใช้ในการพยากรณ์หรือคาดการณ์ผลตอบแทนเมื่อป้อนตัวแปรต้นใหม่ ๆ เข้าไปแบบจำลองก็จะสามารถคาดการณ์ตัวแปรตามได้ใกล้เคียงค่าจริงที่เกิดขึ้นได้ โดยวัดจากฟังก์ชันเปรียบเทียบค่าจริงกับค่าพยากรณ์ หรือฟังก์ชันวัดความคลาดเคลื่อน (Cost Function)

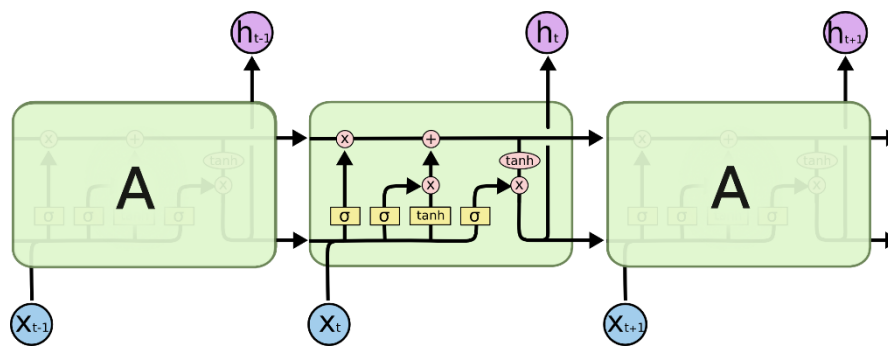
แบบจำลอง Long Short-Term Memory (LSTM) Long short-term memory (LSTM) เป็นโครงข่ายประเภท RNN ซึ่งเหมาะนำมาใช้งานกับข้อมูลที่มีลักษณะเป็นลำดับ (Sequence) หรือข้อมูลที่มีความต่อเนื่อง เช่น ข้อมูลอนุกรมเวลา (Time Series) ข้อมูลเสียง ข้อมูลประเภทข้อความ ข้อมูลภาพและวิดีโอ เป็นต้น (Srivastava, Koutník, Steunebrink, & Schmidhuber, 2017) ซึ่ง LSTM ถูกพัฒนาขึ้นมาให้มีความเสถียรและมีประสิทธิภาพมากขึ้น โดยมีหลักการทำงานคือ สามารถเก็บ ‘สถานะ’ หรือข้อมูลของแต่ละโหนดเอาไว้เพื่อที่เวลาย้อนกลับไปดูจะได้ทราบถึงที่มาของข้อมูลค่าดังกล่าวว่าเดิมเป็นค่าอะไร และจุดเด่นของแบบจำลอง LSTM คือฟังก์ชันพิเศษที่มีหน้าที่เสมือนประตู (Gate) ที่คอยควบคุมข้อมูลที่จะเข้าในแต่ละโหนดซึ่งประกอบด้วย Forget gate layer, Input gate layer และ Output gate layer (Jozefowicz, Zaremba, & Sutskever, 2015)

โดยการนำผลลัพธ์ที่ได้จากการคำนวณจากโหนดก่อนหน้า กลับมาใช้เป็นข้อมูลที่ใช้ในการคำนวณของโหนดถัดไป



รูปที่ 1 แสดงโครงสร้างการทำงานของ RNN

ที่มา: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>



รูปที่ 2 แสดงโครงสร้างการทำงานของ LSTM

ที่มา: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

ระเบียบวิธีวิจัย (Research Methodology)

ข้อมูลและตัวแปรที่ใช้ในงานวิจัย

ขอบเขตข้อมูลที่ใช้ในการศึกษานี้ คัดเลือกจากข้อมูลจากแหล่งทุติยภูมิประกอบด้วย ธุรสารทุนที่เป็นองค์ประกอบของดัชนี SET100 จากตลาดหลักทรัพย์แห่งประเทศไทย (Stock Exchange of Thailand) ณ เดือนมกราคม 2009 จำนวน 61 หุ้น ใน SET100 ในช่วงปี 2009 – 2021 ยกเว้นกลุ่มอุตสาหกรรมกลุ่มธนาคาร (Banking Sector) ซึ่งมีข้อมูลปัจจัยและโครงสร้างงบการเงินที่แตกต่างจากธุรกิจในกลุ่มอื่น ๆ และ ตัวแปรต้นทั้งหมด 34 ตัวแปร โดยแบ่งกลุ่มเป็น 9 ประเภท ประกอบด้วย ข้อมูลธุรสารทุนในตลาด งบการเงิน ข้อมูลเศรษฐกิจศาสตร์มหภาค ข้อมูลการซื้อขายหุ้นทางเทคนิค อัตราส่วนทางการเงิน อัตราแลกเปลี่ยนเงินตรา ดัชนีหุ้น ดัชนีทองคำ และตัวเลขผู้ติดเชื้อไวรัสโคโรนา 2019 และเก็บข้อมูลแบบรายวัน

การพยากรณ์ผลตอบแทน

การประมวลข้อมูลเพื่อพยากรณ์ผลตอบแทนรายวัน รายเดือน และรายสามเดือน โดยใช้ Machine Learning 3 แบบ คือ ANN, RF และ LSTM ด้วยโปรแกรมภาษา Python โดยทำงานร่วมกับ Scikit-learn ซึ่งเป็น Machine Learning Library และใช้งานร่วมกับ Keras ซึ่งเป็น High-level Neuron Network API สำหรับแบบจำลองทั้งหมดจำนวน 61 แบบจำลองตามจำนวนของตราสารทุน และเปรียบเทียบความแม่นยำของผลตอบแทนที่พยากรณ์ได้ของแต่ละแบบจำลอง โดยมีขั้นตอนดังนี้

1. นำข้อมูลช่วง 380 วันแรกเพื่อให้แบบจำลองเรียนรู้และทำการทดสอบแบบจำลองในอีก 100 วันถัดไป เช่น ปัจจุบันอยู่ที่สัปดาห์ ณ วันที่ 5 มกราคม 2016 จะใช้ข้อมูลย้อนหลัง 380 วัน สำหรับให้แบบจำลองเรียนรู้ และจะใช้ข้อมูลย้อนหลัง 100 วันถัดไป

2. ทดสอบแบบจำลองโดยใช้ค่า Cost Function คือ MAE ซึ่งกล่าวในบทก่อนหน้านี้ในการทดสอบว่าแบบจำลองทำงานได้มีประสิทธิภาพในการคาดการณ์ผลตอบแทนได้แม่นยำหรือไม่

3. ปรับแบบจำลองตั้งแต่ขั้นตอนการออกแบบให้เหมาะสม เช่น เพิ่มเส้นโครงข่ายประสาทเทียม, ปรับ Activation Function ให้เหมาะสมกับข้อมูล, เพิ่มจำนวนโหนดในแต่ละ Layer ของโครงข่ายประสาทเทียม เป็นต้น และทำตามขั้นตอนข้างต้นวนไปทั้งหมด 127 รอบโดยการยับยั้งข้อมูลไปข้างหน้ารอบละ 20 วัน ซึ่งยังคงโครงสร้างการใช้ข้อมูล 480 สัปดาห์ก่อนหน้านี้ในการให้แบบจำลองเรียนรู้และทำการทดสอบแบบจำลอง

4. นำผลการพยากรณ์ที่ได้ไปทดสอบความแม่นยำของแต่ละแบบจำลอง โดยใช้ RMSE, MAE และ MSE ในการเปรียบเทียบกันว่าแบบจำลองได้มีค่าความแม่นยำสูงที่สุดและวัดผลกับ Benchmark ซึ่งคำนวณจากค่าเฉลี่ยของผลตอบแทนก่อนหน้านี้(ทั้งระยะสั้น 5 วัน และระยะยาว 20 วัน(นับตาม Day-trade) และเปรียบเทียบการพยากรณ์ด้วยว่าช่วงการพยากรณ์ใด (1 วัน, 1 เดือน และ 3 เดือน)

5. ผลการวิจัย (Empirical Results)

งานวิจัยนี้นำเสนอการคาดการณ์ผลตอบแทนตราสารทุนโดยใช้ Machine Learning 3 อัลกอริทึม (ANN, RF และ LSTM) ที่เรียนรู้จากข้อมูลและปัจจัยที่มีผลกระทบต่อราคาตราสารทุน ผนวกกับการนำผลตอบแทนที่คาดการณ์วัดผลหาค่าความผิดพลาดเพื่อเปรียบเทียบว่าแบบจำลองได้ให้ค่าความผิดพลาดน้อยที่สุดหรือก็คือมีความแม่นยำมากที่สุด

โดยผลจากการวิจัยพบว่าในการพยากรณ์แบบรายวันในทุกแบบจำลองนั้นให้ผลที่แม่นยำน้อยกว่าค่าเฉลี่ยของผลตอบแทนช่วงเวลาก่อนหน้านี้ทั้งระยะสั้น 5 วันและระยะยาว 20 วันตามลำดับ โดยแบบจำลอง ANN มีค่า RMSE ที่ 2.57% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 0.17%, มากกว่าค่าเฉลี่ยระยะยาว 0.15%, มากกว่าแบบจำลอง RF 0.07% แต่น้อยกว่าแบบจำลอง LSTM 0.23% และมีค่า Hitrate ที่ 54.14% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 4.54%, มากกว่าค่าเฉลี่ยระยะยาว 4.04%, มากกว่าแบบจำลอง RF 0.93% แต่น้อยกว่า

แบบจำลอง LSTM 1.3% แบบจำลอง LSTM มีค่า RMSE ที่ 2.8% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 0.4%, มากกว่าค่าเฉลี่ยระยะยาว 0.38%, มากกว่าแบบจำลอง ANN 0.23% และกว่าแบบจำลอง RF 0.3% และมีค่า Hitrate ที่ 55.44% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 5.84%, มากกว่าค่าเฉลี่ยระยะยาว 5.34%, มากกว่าแบบจำลอง ANN 1.3% และมากกว่าแบบจำลอง RF 2.23% แบบจำลอง RF มีค่า RMSE ที่ 2.5% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 0.1 %, มากกว่าค่าเฉลี่ยระยะยาว 0.08 %, น้อยกว่าแบบจำลอง ANN 0.07% และน้อยกว่าแบบจำลอง LSTM 0.3% และมีค่า Hitrate ที่ 53.21% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 3.61%, มากกว่าค่าเฉลี่ยระยะยาว 3.11%, น้อยกว่าแบบจำลอง ANN 0.93% และมากกว่าแบบจำลอง LSTM 2.23%

การพยากรณ์แบบรายเดือนให้ผลดังนี้ โดยแบบจำลอง ANN มีค่า RMSE ที่ 7.07% ซึ่งน้อยกว่าค่าเฉลี่ยระยะสั้น 7.16%, น้อยกว่าค่าเฉลี่ยระยะยาว 6.95%, น้อยกว่าแบบจำลอง LSTM 0.18% แต่มากกว่าแบบจำลอง RF 0.7% และมีค่า Hitrate ที่ 78.43% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 28.58%, มากกว่าค่าเฉลี่ยระยะยาว 27.59%, มากกว่าแบบจำลอง LSTM 0.16% แต่น้อยกว่าแบบจำลอง RF 1.99% แบบจำลอง LSTM มีค่า RMSE ที่ 7.25% ซึ่งน้อยกว่าค่าเฉลี่ยระยะสั้น 6.98%, น้อยกว่าค่าเฉลี่ยระยะยาว 6.77%, มากกว่าแบบจำลอง ANN 0.18% และกว่าแบบจำลอง RF 0.88% และมีค่า Hitrate ที่ 78.27% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 28.42%, มากกว่าค่าเฉลี่ยระยะยาว 27.43%, น้อยกว่าแบบจำลอง ANN 0.16% และน้อยกว่าแบบจำลอง RF 2.15% แบบจำลอง RF มีค่า RMSE ที่ 6.37% ซึ่งน้อยกว่าค่าเฉลี่ยระยะสั้น 7.86 %, น้อยกว่าค่าเฉลี่ยระยะยาว 7.65 %, น้อยกว่าแบบจำลอง ANN 0.7% และน้อยกว่าแบบจำลอง LSTM 0.88% และมีค่า Hitrate ที่ 80.42% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 30.57%, มากกว่าค่าเฉลี่ยระยะยาว 29.58%, มากกว่าแบบจำลอง ANN 1.99% และมากกว่าแบบจำลอง LSTM 2.15%

การพยากรณ์แบบรายไตรมาสให้ผลดังนี้ โดยแบบจำลอง ANN มีค่า RMSE ที่ 7.85% ซึ่งน้อยกว่าค่าเฉลี่ยระยะสั้น 18.81%, น้อยกว่าค่าเฉลี่ยระยะยาว 18.95%, น้อยกว่าแบบจำลอง LSTM 0.06% แต่มากกว่าแบบจำลอง RF 0.78% และมีค่า Hitrate ที่ 87.85% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 36.22%, มากกว่าค่าเฉลี่ยระยะยาว 36.42%, มากกว่าแบบจำลอง LSTM 0.10% แต่น้อยกว่าแบบจำลอง RF 1.11% แบบจำลอง LSTM มีค่า RMSE ที่ 7.91% ซึ่งน้อยกว่าค่าเฉลี่ยระยะสั้น 18.75%, น้อยกว่าค่าเฉลี่ยระยะยาว 18.89%, มากกว่าแบบจำลอง ANN 0.06% และกว่าแบบจำลอง RF 0.84% และมีค่า Hitrate ที่ 87.75% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 36.12%, มากกว่าค่าเฉลี่ยระยะยาว 36.32%, น้อยกว่าแบบจำลอง ANN 0.1% และน้อยกว่าแบบจำลอง RF 1.21% แบบจำลอง RF มีค่า RMSE ที่ 7.07% ซึ่งน้อยกว่าค่าเฉลี่ยระยะสั้น 19.59 %, น้อยกว่าค่าเฉลี่ยระยะยาว 19.73 %, น้อยกว่าแบบจำลอง ANN 0.78% และน้อยกว่าแบบจำลอง LSTM 0.84% และมีค่า Hitrate ที่ 88.96% ซึ่งมากกว่าค่าเฉลี่ยระยะสั้น 37.33%, มากกว่าค่าเฉลี่ยระยะยาว 37.53%, มากกว่าแบบจำลอง ANN 1.11% และมากกว่าแบบจำลอง LSTM 1.21%

ซึ่งทั้ง 3 แบบจำลองในการพยากรณ์ผลตอบแทนแบบรายวันนั้นแม่นยำน้อยกว่าค่าเฉลี่ยของผลตอบแทนช่วงเวลาก่อนหน้าทั้งระยะสั้น 5 วันและระยะยาว 20 วัน แต่ก็น้อยกว่าไม่มากกว่ากันนัก ในส่วนของความแม่นยำในการพยากรณ์ทิศทางของผลตอบแทนของหุ้นว่าผลตอบแทนจะเพิ่มขึ้น, ลดลงนั้นทั้ง 3 แบบจำลองให้การพยากรณ์ที่แม่นยำกว่าค่าเฉลี่ยของผลตอบแทนช่วงเวลาก่อนหน้าทั้งระยะสั้น 5 วันและระยะยาว 20 วัน แต่ก็แม่นยำไม่มากกว่ากันมากนักเช่นกัน ในส่วนของการพยากรณ์ที่มีช่วงยาวขึ้น คือ รายเดือน และรายไตรมาสพบว่าแบบจำลองที่มีความแม่นยำที่สุดก็คือแบบจำลอง Random Forest ซึ่งมีค่าความผิดพลาดต่ำที่สุดทั้งรายเดือนและรายไตรมาส และแม่นยำกว่าในการพยากรณ์ทิศทางของผลตอบแทนของหุ้นว่าผลตอบแทนจะเพิ่มขึ้น, ลดลง มากกว่าแบบจำลองอื่น หมายความว่าผลการพยากรณ์ตราสารทุนในตลาดหลักทรัพย์ของไทยแบบรายวันทั้ง 3 แบบจำลองอาจจะไม่เหมาะสมหรืออาจเกิดจากการคัดเลือกตัวแปรต้นที่ไม่ส่งผลต่อผลตอบแทนของหุ้นจึงทำให้ในการพยากรณ์รายวันได้ไม่ดี แต่ในทางกลับกันหากเป็นการพยากรณ์ในระยะยาว(รายเดือน รายไตรมาส) แบบจำลอง RF เหมาะสมกับการพยากรณ์ตราสารทุนในตลาดหลักทรัพย์ไทยกว่าแบบจำลอง ANN และ LSTM และหากสนใจในการคาดการณ์ทิศทางของผลตอบแทนของหุ้นว่าจะเพิ่มขึ้นหรือลดลงนั้นในการพยากรณ์ตราสารทุนในตลาดหลักทรัพย์ของไทยทั้ง 3 แบบ คือ แบบรายวัน, รายเดือน และรายไตรมาส แบบจำลองทั้ง 3 มีความเหมาะสมกับการทำนายทิศทางขึ้นลงของหุ้นในตลาดหลักทรัพย์ของไทย ซึ่งผลการวิจัยเป็นไปตามตารางที่ 1-3

สำหรับตารางที่ 1-3 เป็นการวัดผลแบบจำลองในขั้นตอนการคาดการณ์แบบจำลอง (Model Testing) โดยในขั้นตอนนี้มีค่าคลาดเคลื่อนที่คำนวณจาก RMSE และ MAE เพื่อเปรียบเทียบความแม่นยำในแต่ละแบบจำลองและในแต่ละช่วงผลตอบแทนที่พยากรณ์ โดยมี Benchmark เป็นค่าเฉลี่ยของผลตอบแทนช่วงเวลาก่อนหน้าระยะสั้นและระยะยาว

ตารางที่ 1 แสดงผลค่าคลาดเคลื่อนจากการทดสอบแบบจำลองหรือช่วงการใช้แบบจำลองคาดการณ์ผลตอบแทนของตราสารทุน (Model Testing) แบบ 1 วัน (1 Days) สำหรับทุกแบบจำลอง

Model	RMSE	MAE	HitRate	HitRate+	HitRate-
ANN	2.57%	1.85%	54.14%	42.77%	44.99%
LSTM	2.8%	1.99%	55.44%	43.70%	46.20%
RF	2.5%	1.8%	53.21%	42.14%	44.01%
Mean 5 day	2.4%	1.66%	49.60%	39.52%	40.94%
Mean 20 day	2.42%	1.67%	50.10%	39.79%	41.39%

ตารางที่ 2 แสดงผลค่าคลาดเคลื่อนจากผลการทดสอบแบบจำลองหรือช่วงการใช้แบบจำลองคาดการณ์ผลตอบแทนของตราสารทุน (Model Testing) แบบ 1 เดือน (1 Month) สำหรับทุกแบบจำลอง

Model	RMSE	MAE	HitRate	HitRate+	HitRate-
ANN	7.07%	4.94%	78.43%	76.36%	74.99%
LSTM	7.25%	5.02%	78.27%	76.49%	74.68%
RF	6.37%	4.51%	80.42%	78.57%	76.67%
Mean 5 day	14.23%	10.37%	49.85%	49.62%	45.59%
Mean 20 day	14.02%	10.25%	50.84%	50.58%	46.61%

ตารางที่ 3 แสดงผลค่าคลาดเคลื่อนจากผลการทดสอบแบบจำลองหรือช่วงการใช้แบบจำลองคาดการณ์ผลตอบแทนของตราสารทุน (Model Testing) แบบ 3 เดือน (3 Month) สำหรับทุกแบบจำลอง

Model	RMSE	MAE	HitRate	HitRate+	HitRate-
ANN	7.85%	5.3%	87.85%	89.42%	83.68%
LSTM	7.91%	5.34%	87.75%	88.93%	84.02%
RF	7.07%	4.85%	88.96%	89.99%	85.20%
Mean 5 day	26.66%	19.28%	51.63%	53.08%	46.30%
Mean 20 day	26.8%	19.33%	51.43%	52.85%	46.09%

สรุปงานวิจัย (Conclusion)

งานวิจัยนี้พบว่าในการพยากรณ์แบบรายวันในทุกแบบจำลองนั้นให้ผลที่แม่นยำน้อยกว่าค่าเฉลี่ยของผลตอบแทนช่วงเวลาก่อนหน้าทั้งระยะสั้นและระยะยาวตามลำดับ แต่ในส่วนของการพยากรณ์ที่มีช่วงยาวขึ้น คือ รายเดือน และรายไตรมาสพบว่าแบบจำลองที่มีความแม่นยำที่สุดก็คือแบบจำลอง Random Forest ซึ่งมีค่าความผิดพลาดต่ำที่สุดทั้งรายเดือนและรายไตรมาส และหากวัดความแม่นยำในการคาดการณ์ทิศทางของผลตอบแทนของหุ้นว่าผลตอบแทนจะเพิ่มขึ้น, ลดลง หรือว่าไปในทิศทางใดนั้น ผลที่ได้จากการทดลองพบว่าในการพยากรณ์แบบรายวันแบบจำลองที่ให้ค่าความแม่นยำของทิศทางผลตอบแทนของหุ้นแม่นยำที่สุดคือ LSTM แต่ในส่วนของการพยากรณ์ที่มีช่วงยาวขึ้น คือ รายเดือน และรายไตรมาสพบว่าแบบจำลองที่มีความแม่นยำที่สุดก็คือแบบจำลอง Random Forest ซึ่งมีค่าความแม่นยำมากที่สุดทั้งรายเดือนและรายไตรมาส

การพยากรณ์ตราสารทุนในตลาดหลักทรัพย์ของไทยแบบรายวันทั้ง 3 แบบจำลองอาจจะไม่เหมาะสมหรืออาจเกิดจากการคัดเลือกตัวแปรต้นที่ไม่ส่งผลต่อผลตอบแทนของหุ้นจึงทำให้ในการพยากรณ์รายวันได้ไม่ดี แต่ในทางกลับกันหากเป็นการพยากรณ์ในระยะยาว(รายเดือน รายไตรมาส) แบบจำลอง RF เหมาะสมกับการพยากรณ์ตราสารทุนในตลาดหลักทรัพย์ไทยกว่าแบบจำลอง ANN และ LSTM และหากสนใจในการคาดการณ์

ทิศทางของผลตอบแทนของหุ้นว่าจะเพิ่มขึ้นหรือลดลงนั้นในการพยากรณ์ตราสารทุนในตลาดหลักทรัพย์ของไทยแบบรายวัน แบบจำลองให้พยากรณ์ทิศทางแม่นยำที่สุดคือ LSTM แต่ถ้าหากเป็นการพยากรณ์แบบรายเดือนและรายไตรมาส นั้น แบบจำลอง Random Forest จะคาดการณ์ทิศทางของผลตอบแทนหุ้นได้แม่นยำที่สุด

บรรณานุกรม

- Baker, M. R., & Patil, R. B. (1998). Universal Approximation Theorem for Interval Neural Networks. *Reliable Computing*, 4, 235–239. doi:<https://doi-org.ejournal.mahidol.ac.th/10.1023/A:1009951412412>
- Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *North American Journal of Economics and Finance*, 47(47), 552–567. doi:<https://doi.org/10.1016/j.najef.2018.06.013>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. doi:<https://doi.org/10.1023/A:1010933404324>
- colah's blog. (2015). Understanding LSTM Networks. Retrieved from <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Han, S. (2019). *Stock Prediction with Random Forests and Long Short-term Memory*. (Master Degree). Iowa State University,
- Jiemwiriyakul, B., Sirianuntapiboon, P., & Lorsubkong, P. (2019). *Portfolio Return Prediction using Neural Network*. (Master degree Individual Study). Mahidol University
- Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). *An Empirical Exploration of Recurrent Network Architectures*. Paper presented at the Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research. <https://proceedings.mlr.press/v37/jozefowicz15.html>
- Kim, K.-j. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1), 307–319. doi:[https://doi.org/10.1016/S0925-2312\(03\)00372-2](https://doi.org/10.1016/S0925-2312(03)00372-2)
- Loh, W. Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1), 14-23. doi:<https://doi.org/10.1002/widm.8>
- Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, 165, 1-15. doi:<https://doi.org/10.1016/j.eswa.2020.113973>
- Nilson, M. (2019). A visual proof that neural nets can compute any function. Retrieved from: <http://neuralnetworksanddeeplearning.com/chap4.html>

- Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. *Expert Systems with Applications*, 42(4), 2162-2172. doi:<https://doi.org/10.1016/j.eswa.2014.10.031>
- Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232. doi:<https://doi.org/10.1109/TNNLS.2016.2582924>