

การเลือกหลักทรัพย์ในการจัดพอร์ตลงทุนโดยใช้ข้อมูล  
กูเกิลเทรนด์

STOCK PORTFOLIO SELECTION USING GOOGLE TREND DATA



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์  
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง  
คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง  
พ.ศ. 2566

KMITL-2023-SC-M-017-031

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# STOCK PORTFOLIO SELECTION USING GOOGLE TREND DATA



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE  
DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS  
KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE  
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2023

KMITL-2023-SC-M-017-031

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2023

SCHOOL OF SCIENCE

**KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การเลือกหลักทรัพย์ในการจัดพอร์ตลงทุนโดยกูเกิ้ลเทรนด์
ชื่อนักศึกษา	นายกิตติภาพ แซ่เตีย
รหัสนักศึกษา	63605057
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์)
	ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
พ.ศ	2566
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ดร.จิรภัทร์ หยกรัตน์ศักดิ์

### บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อศึกษาและระบุปัจจัยที่เกี่ยวข้องในการเลือกหลักทรัพย์สำหรับพอร์ตการลงทุน และเปรียบเทียบประสิทธิภาพของแบบจำลองในการทำนายการเคลื่อนไหวของหลักทรัพย์โดยใช้ตัวบ่งชี้ทางเทคนิค (Technical Indicators) และคำค้นหาของกูเกิ้ลเทรนด์ (Google Trends) โดยใช้ข้อมูลตลาดหลักทรัพย์ไทยเป็นฐานข้อมูล การศึกษานี้ใช้ชุดข้อมูล 7 ชุดและเปรียบเทียบประสิทธิภาพของแบบจำลองระหว่าง Logistic Regression, Random Forest, และ Extreme Gradient Boosting (XGBoost) นอกจากนี้ยังมีการเปรียบเทียบประสิทธิภาพการทำอัตราส่วนที่เหมาะสมต่อการจัดพอร์ต (Portfolio Optimization) ระหว่างการลงทุนแบบเท่าเทียม (Equity Investment), Sequential Least Squares Programming algorithm (SLSQP), Risk Parity Optimization (RPO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), และ Particle Swarm Optimization (PSO) ผลการวิจัยพบว่าชุดข้อมูลที่รวมตัวบ่งชี้ทางเทคนิค ข้อความค้นหาของกูเกิ้ลเทรนด์ และตัวชี้วัด Moving Average Convergence Divergence (MACD indicators) ได้ทำการปรับพารามิเตอร์แล้ว แสดงเส้นโค้ง ROC สูงสุดทั้งสามอัลกอริทึม และอัลกอริทึม Random Forest กับอัลกอริทึม XGBoost มีค่าอัตราการทำนายความสำเร็จและผลตอบแทนต่อปีที่คล้ายกัน แต่ยังมี ความแตกต่างกัน โดยอัลกอริทึม XGBoost ทำงานได้ดีในช่วงวิกฤตของตลาด (เช่นสถานการณ์ โควิด-19: Covid-19) ส่วนอัลกอริทึม Random Forest ทำงานได้ดีกว่าอัลกอริทึม XGBoost ในช่วงสภาวะตลาดปกติโดยเฉลี่ย จากการจำลองการจัดพอร์ตหลักทรัพย์ลงทุนในปี 2021 แนะนำ 4 อัลกอริทึมหลักคือ Equity Investment และ Genetic Algorithm สำหรับการลงทุนที่เน้นความเสี่ยงต่ำ และ SLSQP และ Ant Colony Optimization สำหรับการลงทุนที่ยอมรับความเสี่ยงได้สูง

**คำสำคัญ:** การเรียนรู้ของเครื่อง การจัดพอร์ตที่เหมาะสม คำค้นหาของกูเกิ้ลเทรนด์ โควิด-19

<b>Thesis Title</b>	Stock Portfolio Selection Using Google Trend Data
<b>Student</b>	Mr. Kittipob Saetia
<b>Student ID</b>	63605057
<b>Degree</b>	Master of Science (Data Science and Analytics) Kmitl Digital Analytics and Intelligence Center
<b>Year</b>	2023
<b>Thesis Advisor</b>	Dr. Jiraphat Yokrattanasak

### Abstract

This study aims to examine and identify factors related to the selection of securities for investment portfolios, develop portfolio allocation strategies for retail investors, and compare the performance of models in predicting the movement of securities using technical indicators and Google Trends data. The study utilizes data from the Thai stock market as the database and compares the performance of three models: Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost). Additionally, the study compares the performance of different portfolio optimization techniques, including Equity Investment, Sequential Least Squares Programming algorithm (SLSQP), Risk Parity Optimization (RPO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), and Particle Swarm Optimization (PSO). The research findings indicate that the dataset combining technical indicators, Google Trends search queries, and Moving Average Convergence Divergence (MACD) indicators with adjusted parameters show the highest ROC curves for all three algorithms. Both Random Forest and XGBoost algorithms exhibit similar prediction success rates and annual returns. However, XGBoost performs better during market crises such as the Covid-19 situation, while Random Forest outperforms XGBoost in normal market conditions on average. Based on the portfolio allocation simulations conducted in 2021, the researchers recommend using four main algorithms: Equity Investment and Genetic Algorithm for low-risk investment strategies, and SLSQP and Ant Colony Optimization for strategies that are willing to accept higher risk.

**Keywords:** Machine Learning, Portfolio Optimization, Covid-19, Google Trends

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จสมบูรณ์ได้ด้วยดีเพราะได้รับความอนุเคราะห์และเอาใจใส่ให้คำปรึกษาเป็นอย่างดีจาก อาจารย์ ดร.จิรภัทร์ หยกรัตนศักดิ์ อาจารย์ที่ปรึกษา ที่ได้ให้คำปรึกษาแนะนำแก้ไขปรับปรุงข้อบกพร่อง รวมทั้งให้คำแนะนำองค์ความรู้แนวทางการศึกษาค้นคว้ามาโดยตลอด ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

สุดท้ายนี้ ขอกราบขอบพระคุณสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ซึ่งเป็นสถานศึกษาที่ให้ความรู้แก่นักศึกษาทุกคน เพื่อให้นักศึกษาได้นำความรู้ไปใช้ประโยชน์ต่อตนเองและผู้อื่น รวมทั้งมีความรู้เพื่อที่จะนำไปประยุกต์ในการประกอบอาชีพสืบไป

กิตติภาพ แซ่เตีย



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

# สารบัญ

	หน้า
บทคัดย่อ.....	ก
Abstract.....	ข
กิตติกรรมประกาศ.....	ค
สารบัญ.....	ง
สารบัญตาราง.....	ฉ
สารบัญรูป.....	ช
<b>บทที่ 1 .....</b>	<b>1</b>
1.1 ที่มาและความสำคัญของงานวิจัย .....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	4
1.3 ขอบเขตของงานวิจัย .....	4
1.4 ระยะเวลาการดำเนินงานทำงานวิจัย .....	5
1.5 ประโยชน์ที่คาดว่าจะได้รับ .....	5
<b>บทที่ 2 .....</b>	<b>6</b>
2.1 ข้อมูล (Data).....	6
2.1.1. ข้อมูลคำค้นหาจากกูเกิลเทรนด์ (Google Trends Search Information).....	6
2.1.2. ข้อมูลตลาดหลักทรัพย์ (Stock Market Information).....	6
2.2 การเลือกปัจจัยที่ส่งผล (Feature Selection).....	9
2.3 แบบจำลอง (Modeling) .....	10
2.3.1 การจำแนกประเภท (Classification).....	11
2.3.2 การทำให้เหมาะสมที่สุด (Optimization).....	14
2.3 การประเมินผล (Evaluation) .....	20
2.4 งานวิจัยที่เกี่ยวข้อง (Research Methodology).....	22
<b>บทที่ 3 .....</b>	<b>28</b>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญ (ต่อ)

	หน้า
3.1 แผนผังขั้นตอนการดำเนินงาน (Process Diagram) .....	28
3.1.1 ขั้นตอนการดำเนินงาน .....	28
3.1.2 เครื่องมือหรือโปรแกรมที่ใช้ในการดำเนินงาน.....	29
3.2 ข้อมูล (Dataset).....	29
3.3 การเตรียมข้อมูล (Data Preprocessing).....	33
3.5 ผลลัพธ์และการอภิปราย (Results and Discussion).....	38
<b>บทที่ 4 .....</b>	<b>47</b>
4.1 แผนผังขั้นตอนการดำเนินงาน (Process Diagram) .....	47
4.2 การเลือกหลักทรัพย์ (Selections Stocks) .....	48
4.3 การจัดพอร์ตหลักทรัพย์ที่เหมาะสม (Portfolio Optimization) .....	48
4.3.1 การลงทุนแบบเท่าเทียม (Equity Investment - EI).....	48
4.3.2 Sequential Least Squares Programming algorithm (SLSQP) .....	49
4.3.3 การปรับปรุงความเสี่ยงแบบเทียบค่าความเสี่ยง (Risk Parity Optimization - RPO)...	50
4.3.4 ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm - GA).....	51
4.3.5 อาณาจักรมด (Ant Colony Optimization - ACO) .....	52
4.3.6 วิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization - PSO) .....	53
<b>บทที่ 5 .....</b>	<b>57</b>
5.1 สรุปและอภิปรายผล (Conclusions) .....	57
5.2 ข้อจำกัดและการทำงานในอนาคต (Limitation and Future Work) .....	58
เอกสารอ้างอิง .....	60
ภาคผนวก.....	65
ประวัติผู้เขียน.....	88

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญตาราง

ตารางที่	หน้า
2.1 สมการตัวบ่งชี้ทางเทคนิค (Technical Indicators).....	8
2.2 ตารางแสดงผลการทำนาย (Confusion Matrix).....	21
2.3 การเปรียบเทียบเชิงประจักษ์ของงานวิจัย.....	26
2.3 (ต่อ).....	27
3.1 ตัวอย่างข้อมูลหลักทรัพย์ของ บริษัท ดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน) (WHA).....	30
3.2 ตัวอย่างคำสำคัญ (Keywords).....	30
3.2 (ต่อ).....	31
3.2 (ต่อ).....	32
3.3 ตัวอย่างคำค้นหา Google Trends สำหรับแต่ละคำในประเทศไทย.....	33
3.4 ตัวอย่างข้อมูลหลักทรัพย์ และค่าตัวชี้วัดทางเทคนิค (Technical Indicators) ของบริษัท ดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน) (WHA).....	33
3.5 ตัวอย่างค่าพารามิเตอร์ของ MACD indicators สำหรับแต่ละหลักทรัพย์.....	34
3.5 (ต่อ).....	35
3.6 คำสำคัญ (Keywords) 20 อันดับแรกของบริษัท ดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน).....	35
3.6 (ต่อ).....	36
3.7 พารามิเตอร์ของการจำแนกประเภทแบบถดถอยโลจิสติก.....	37
3.8 พารามิเตอร์ของการจำแนกประเภทแบบการสุ่มป่าไม้.....	37
3.9 พารามิเตอร์ของการจำแนกประเภทแบบ Extreme Gradient Boosting.....	38
3.10 ประสิทธิภาพของชุดข้อมูล (Performance on dataset).....	38
3.11 ประสิทธิภาพสำหรับวิกฤต (Performance for Crisis).....	43
4.1 ตัวอย่างหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021.....	48
4.2 ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 เมื่อใช้การลงทุนแบบเท่าเทียม (Equity Investment).....	49
4.3 พารามิเตอร์ของ SLSQP.....	49
4.4 ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 (เมื่อจัดพอร์แบบ SLSQP).....	50

## สารบัญตาราง(ต่อ)

ตารางที่	หน้า
4.6 ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ในช่วงปี 2021 (เมื่อจัดพอร์ตแบบ RPO) .....	51
4.7 พารามิเตอร์ของ Genetic Algorithm.....	51
4.8 ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ในช่วงปี 2021 (เมื่อจัดพอร์ตแบบ Genetic Algorithm) .....	52
4.9 พารามิเตอร์ของ Ant Colony Optimization .....	52
4.10 ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ในช่วงปี 2021 (เมื่อจัดพอร์ตแบบ Ant Colony Optimization).....	53
4.11 พารามิเตอร์ของ Particle Swarm Optimization.....	53
4.12 ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ในช่วงปี 2021 (เมื่อจัดพอร์ตแบบ Particle Swarm Optimization).....	54
4.13 ประสิทธิภาพการจัดพอร์ตหลักทรัพย์ที่เหมาะสมและผลตอบแทนรวมประจำปี 2021 .....	55

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## สารบัญรูป

รูปที่	หน้า
2.1 แผนภาพขั้นตอนการสร้างแบบจำลองถดถอยโลจิสติก (Logistic Regression) .....	11
2.2 แผนภาพขั้นตอนการสร้างแบบจำลองการสุ่มป่าไม้ (Random Forests).....	12
2.3 แสดงแผนภาพขั้นตอนการสร้างแบบจำลอง XGBoost.....	14
3.1 แผนผังขั้นตอนการดำเนินงาน (Process Diagram) .....	28
3.2 เปอร์เซ็นต์การเลือกหลักทรัพย์ที่ประสบความสำเร็จ (Success Prediction Rate).....	40
3.3 เปอร์เซ็นต์ผลตอบแทนต่อปี (Annualized Return) .....	40
3.4 ผลตอบแทนสะสมตามอนุกรมเวลา (Time-Series Cumulative Return).....	41
3.5 วิฤตการณ์ในประเทศไทยระหว่างปี ค.ศ. 2019 – 2021 .....	42
4.1 แผนผังขั้นตอนการดำเนินงาน (Process Diagram) .....	47



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 1

### บทนำ

เนื่องจากในปัจจุบันการลงทุนทางการเงินมีความนิยมเป็นอย่างมาก ทางผู้วิจัยได้สืบค้นและศึกษาลักษณะการลงทุน พบว่าการเลือกหลักทรัพย์ที่ดี และการจัดพอร์ตหลักทรัพย์ที่เหมาะสมนั้นมีหลากหลายวิธี งานวิจัยนี้จะนำเสนออีกวิธีหนึ่งในการเลือกหลักทรัพย์ และการจัดพอร์ตลงทุน ในบทนี้จะกล่าวถึงที่มาและความสำคัญของงานวิจัย วัตถุประสงค์ของงานวิจัย ขอบเขตของงานวิจัย ระยะเวลาในการปฏิบัติงานวิจัย และประโยชน์ที่คาดว่าจะได้รับ โดยมีรายละเอียดดังต่อไปนี้

#### 1.1 ที่มาและความสำคัญของงานวิจัย

เป็นที่ทราบกันดีว่า ในตลาดหลักทรัพย์นั้น อาจเกิดความไม่มีประสิทธิภาพของตลาด (Market Inefficiencies) ถูกสังเกตเห็นเมื่อราคาของหลักทรัพย์ในจุดใดจุดหนึ่งไม่สะท้อนค่าความสำคัญทั้งหมดของสินทรัพย์ในตลาดหลักทรัพย์ สาเหตุอาจเกิดจากพฤติกรรมการตัดสินใจของนักลงทุนที่ไม่สามารถทำนายได้โดยใช้ข้อมูล (Data) และตัวชี้วัด (Indicators) ที่มีอยู่ในปัจจุบัน ดังนั้นการที่นักวิเคราะห์ (Analysts) และนักลงทุน (Investors) สามารถเข้าใจความสัมพันธ์ระหว่างตัวแปรที่เกี่ยวข้องกับปรากฏการณ์ต่าง ๆ ในตลาด เช่น ปริมาณการซื้อขาย (Trade Volume) แนวโน้มการค้นหาออนไลน์ (Online Search Trends) และพฤติกรรมของนักลงทุน (Investor Behaviors) จะ เป็นประโยชน์ในการสร้างโมเดลเพื่อทำนายพฤติกรรมการซื้อขายของนักลงทุน และ การเคลื่อนไหวของตลาดอย่างแม่นยำมากขึ้นได้

ในบริบทของตลาดการลงทุน (Investment Market) พฤติกรรมของนักลงทุนที่ไม่สมเหตุสมผล (Irrational Investor Behavior) สามารถเกิดขึ้นในรูปแบบต่าง ๆ ได้ ตัวอย่างหนึ่งคือนักลงทุนที่ออกจากตลาดเนื่องจากความกลัวเกี่ยวกับการชะลอตัวของตลาด (Market Downturn) ที่อาจเกิดขึ้นได้ในอนาคต การเคลื่อนไหวของราคาสินทรัพย์ในอนาคตและทัศนคติของตลาดโดยรวมจะได้รับผลกระทบจากแบบแผนพฤติกรรมตลาด จนถึงปัจจุบันการศึกษาที่ผ่านมาได้สำรวจเรื่องราวที่เกี่ยวข้องกับการจำแนกและการทำนายหลักทรัพย์ (Stock Prediction) และตลาดหลักทรัพย์โดยใช้ความเข้าใจในพฤติกรรมของมนุษย์และผลกระทบที่มีต่อตลาดหลักทรัพย์ ตัวชี้วัดทางเศรษฐกิจ (Economic) และ ตัวชี้วัด (Indicators), ข้อมูลตลาด (Market Information), ข่าวหัวข้อสำคัญ (Headline News), และคำค้นหาออนไลน์ (Online Search Terms) ถูกนำมาใช้ในการวิจัยเหล่านี้ (Alfonso & Ramirez, 2020; Antonio Agudelo Aguirre et al., 2020; Atkins et al., 2018; Chen et al., 2020; Dash & Dash, 2016; Huang et al., 2019; Li et al., 2014; Papadamou et al., 2022; Poutachidou & Papadamou, 2021; Teixeira & de Oliveira, 2010; Yu et al., 2013) ตัวบ่งชี้ที่กล่าวถึงส่วนใหญ่ได้รับการพิสูจน์แล้วว่าเป็นประโยชน์ในการทำนายราคารวมของหลักทรัพย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการศึกษาที่เกี่ยวข้องนี้ Teixeira and de Oliveira (2010) รวมตัวชี้วัดทางเทคนิค เช่น Simple Moving Average (SMA), Relative Strength Index (RSI), Stochastic Oscillator K, Stochastic Oscillator D, และ Bollinger bands พร้อมกับอัลกอริทึม K-Nearest Neighbor (KNN) เพื่อสร้างโมเดลการทำนาย การศึกษานี้สรุปว่าวิธีนี้มีประสิทธิภาพสูงกว่ากลยุทธ์ซื้อและถือไว้ (buy-and-hold strategy) ใน 12 จากทั้งหมด 15 หลักทรัพย์ในประเทศบราซิล (Brazil) ที่พิจารณาในการศึกษา เมื่อเปรียบเทียบผลลัพธ์ ดังนั้น การศึกษานี้มีความเห็นว่าเป็นไปได้ที่จะใช้วิธีนี้ในการทำนายแนวโน้มหลักทรัพย์ระยะสั้น

การศึกษ้อีกฉบับหนึ่งโดย Dash and Dash (2016) ในอินเดียใช้ตัวชี้วัดทางเทคนิคร่วมกับการเรียนรู้ของเครื่องจักร ได้ถูกทดสอบในบริบทของ S&P 500 และ BSE SENSEX ในวิธีวิเคราะห์ข้อมูลมากมาย และผลตอบแทนที่สูงที่สุดมาจากโครงข่ายประสาทเทียม ให้เปอร์เซ็นต์ผลตอบแทนที่ดีกว่าแบบจำลองอื่น ๆ เฉลี่ย 34.42 % ของดัชนี BSE SENSEX และ 45.58% ของดัชนี S&P 500 ตามลำดับ

นักวิจัยยังพยายามใช้ข้อมูลเชิงคุณภาพที่ไม่มีโครงสร้าง (Unstructured Qualitative Data) เพื่อทำนายราคาหลักทรัพย์และผลตอบแทนในตลาดหุ้น Li et al. (2014) ปรับใช้ข่าวหลัก บล็อก และกระทู้สนทนาทางการเงินร่วมกับ Support Vector Regression (SVR) การศึกษาสรุปว่าผลกระทบจากสื่อต่อบริษัทแตกต่างกันตามลักษณะของข้อมูล พื้นฐานของบทความ เช่น ข่าวที่เกี่ยวข้องกับบริษัทเฉพาะตัว สามารถเพิ่มความรู้นักลงทุนได้ อารมณ์ของสาธารณชนสามารถก่อให้เกิดความผันผวนในอารมณ์ของนักลงทุนและมีผลต่อการตัดสินใจของนักลงทุน ในทางตรงกันข้าม Atkins et al. (2018) รวมข้อมูลสองประเภท ข่าว (News) จาก Reuters USA และข้อมูลทางการเงิน (Financial Data) บน Yahoo Finance รวมกันใน Latent Dirichlet Allocation (LDA) การศึกษานี้สรุปว่าข้อมูลข่าวมีผลต่อการพยากรณ์ความผันผวนได้ดีกว่าการพยากรณ์ราคาปิดตลาดหลักทรัพย์

Bustos and Pomares-Quimbaya (2020) ได้เผยแพร่เกี่ยวกับกระบวนการโดยรวมในการทำนายการเคลื่อนไหวของตลาดหลักทรัพย์โดยใช้เทคนิคการเรียนรู้ของเครื่องจักร (Machine learning, ML) การเรียนรู้เชิงลึก (Deep Learning) การทำเหมืองข้อมูลจากข้อความ (Text Mining) และเทคนิคการจัดกลุ่ม (Clustering) เพื่อสร้างโมเดลการลงทุน การใช้เทคนิคเหล่านี้เป็นเครื่องมือในการวางแผนกลยุทธ์และการตัดสินใจของนักลงทุนที่ให้ผลลัพธ์ที่ดีกว่ากลยุทธ์การซื้อขายแบบดั้งเดิมอย่างชัดเจน

Huang et al. (2019) ได้ศึกษาความมีประสิทธิภาพของโมเดลการพยากรณ์ตลาดหลักทรัพย์สำหรับดัชนี S&P500 ด้วย Google Trends, Support Vector Machine (SVM), Ridge Regression, Lasso Regression, และ Elastic-Net Regression โดยโมเดลที่มีผลสูงสุดคือ Ridge Regression พร้อมกับเลือกปัจจัยที่สำคัญด้วย Vector Autoregression Model ด้วยความแม่นยำ 63.75%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Papadamou et al. (2022) ได้สำรวจความสัมพันธ์ระหว่างอารมณ์ของนักลงทุนที่แสดงโดย Google Trends และผลตอบแทนตลาดหลักทรัพย์ ความผันผวน และการซื้อขายหลักทรัพย์ในสถานการณ์ของหลักทรัพย์อุตสาหกรรมกัญชา โดยศึกษา Three-Factor Fama-French Model ผลตอบแทนจากหลักทรัพย์กัญชาและสภาพคล่องมีความสัมพันธ์กันในทางบวกทางสถิติ นั้นหมายความว่าความสนใจของนักลงทุนที่เพิ่มขึ้นจะเพิ่มผลตอบแทนในหลักทรัพย์

Poutachidou and Papadamou (2021) ศึกษานโยบาย Quantitative Easing (QE) กับตลาดหลักทรัพย์ในสหรัฐฯ โดยสรุปว่ามีความสัมพันธ์บวกระหว่างผลตอบแทนในดัชนีตลาดหลักทรัพย์และความสนใจของนักลงทุนในนโยบาย Quantitative Easing (QE) ในสหรัฐฯ ตามการวัดด้วย Google Metrics ซึ่งแนะนำว่าความสนใจของนักลงทุนใน QE จะลดความผันผวนในตลาดหลักทรัพย์และเพิ่มผลตอบแทนหลักทรัพย์

ด้วยเหตุนี้ มีช่องว่างในการวิจัยทางวิชาการ เนื่องจากมีเพียงบางบทความที่ได้ทำการสำรวจความมีประสิทธิภาพของการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่องจักร (Machine Learning) ไม่เพียงแต่ในการทำนายผลการเคลื่อนไหวของหลักทรัพย์ แต่ยังเป็นการอัตโนมัติในการเลือกสร้างพอร์ต (Automate Portfolio Selection) ให้ได้ผลตอบแทนสูงและการตัดสินใจการลงทุนที่ดีกว่า นอกจากนี้ การเลือกหลักทรัพย์ที่ "ถูกต้อง" ในการลงทุนต้องใช้ทักษะเฉพาะและประสบการณ์ เนื่องจากต้องพิจารณาเกณฑ์หลายอย่างและวิเคราะห์ข้อมูลจำนวนมากในขณะเลือกหลักทรัพย์ ดังนั้น ความสนใจของการวิจัยครั้งนี้ คือการใช้เทคนิคการเรียนรู้ของเครื่องจักร (Machine Learning) โดยเฉพาะ การถดถอยโลจิสติก (Logistic Regression), การสุ่มป่าไม้ (Random Forest), และ Extreme Gradient Boosting (Xgboost) เพื่อเลือกหลักทรัพย์ลงทุนโดยใช้ข้อมูล 7 ชุด: (1) ตัวชี้วัดเทคนิค (Technical Indicators), (2) ตัวชี้วัด Moving Average Convergence Divergence (MACD Indicators), (3) คำค้นใน Google Trends (Keyword), (4) Indicators & MACD, (5) Keyword & Indicators, (6) Keyword & MACD, และ (7) Keyword & Indicators & MACD

ดังนั้น การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อ (1) ศึกษาและระบุปัจจัยที่เกี่ยวข้องในการเลือกหลักทรัพย์สำหรับพอร์ตการลงทุน, (2) พัฒนาแนวทางการจัดพอร์ตการลงทุนสำหรับนักลงทุนรายย่อย, (3) เปรียบเทียบประสิทธิภาพความสามารถในการทำนายการเคลื่อนไหวของหลักทรัพย์โดยใช้แบบจำลองจากข้อมูล 7 ชุดด้านบน, และ (4) เปรียบเทียบประสิทธิภาพของแบบจำลองระหว่าง Logistic Regression, Random Forest, และ XGBoost (5) เปรียบเทียบประสิทธิภาพการหาอัตราส่วนที่เหมาะสมต่อจัดพอร์ต (Portfolio Optimization) ระหว่าง การลงทุนแบบเท่าเทียม (Equity Investment), Sequential Least Squares Programming algorithm (SLSQP), การปรับปรุงความเสี่ยงแบบเทียบค่าความเสี่ยง (Risk Parity Optimization - RPO), ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm - GA), อาณาจักรมด (Ant Colony Optimization - ACO), และวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization - PSO)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 1.2 วัตถุประสงค์ของงานวิจัย

วัตถุประสงค์ของงานวิจัยครั้งนี้ มีดังนี้

1. ศึกษาและระบุปัจจัยที่เกี่ยวข้องในการเลือกหลักทรัพย์สำหรับพอร์ตการลงทุน
2. สร้างแบบจำลองการให้คะแนนหลักทรัพย์ต่าง ๆ เพื่อสามารถจัดอันดับ และตัดสินใจเลือกหลักทรัพย์ในการลงทุนได้
3. เปรียบเทียบประสิทธิภาพความสามารถในการทำนายการเคลื่อนไหวของหลักทรัพย์จากข้อมูล 7 ชุด:
  - (1) ตัวชี้วัดเทคนิค (technical indicators),
  - (2) ตัวชี้วัด Moving Average Convergence Divergence (MACD indicators):  
MACD ได้มีการปรับค่าพารามิเตอร์ให้เหมาะสมของแต่ละหลักทรัพย์ โดยแต่ละหลักทรัพย์จะมีค่าพารามิเตอร์ ของ MACD ที่แตกต่างกัน
  - (3) คำค้นใน Google Trends (Keyword)
  - (4) Indicators & MACD
  - (5) Keyword & Indicators
  - (6) Keyword & MACD
  - (7) Keyword & Indicators & MACD
4. เปรียบเทียบประสิทธิภาพของแบบจำลองระหว่าง Logistic Regression, Random Forest, และ XGBoost
5. สร้างแบบจำลองการหาสัดส่วนในการลงทุนที่เหมาะสมสำหรับแต่ละหลักทรัพย์ เพื่อเพิ่มประสิทธิภาพของพอร์ตการลงทุน และเปรียบเทียบประสิทธิภาพการหาอัตราส่วนที่เหมาะสมต่อการจัดพอร์ตที่เหมาะสม (Portfolio Optimization) ระหว่าง การลงทุนแบบเท่าเทียม (Equity Investment), Sequential Least Squares Programming Algorithm (SLSQP), Risk Parity Optimization (RPO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), และ Particle Swarm Optimization (PSO)

## 1.3 ขอบเขตของงานวิจัย

ข้อมูลที่ใช้ศึกษาในงานวิจัยนี้ เป็นข้อมูลตลาดหลักทรัพย์ SET100 และคำค้นหาในอินเทอร์เน็ตในประเทศไทย ตั้งแต่วันที่ 1 มกราคม 2016 ถึง 31 ธันวาคม 2021 เพื่อศึกษาปัจจัยที่เกี่ยวข้อง การเลือกหลักทรัพย์ และการลงทุนที่เหมาะสม โดยมีรายละเอียดดังต่อไปนี้

1. ข้อมูลราคา ปริมาณในการซื้อขายของ ตลาดหลักทรัพย์ของประเทศไทย (SET100) เป็นรายสัปดาห์

2. คำ (Word) ซึ่งเป็นที่นิยมที่นักลงทุนใช้ค้นหาในกูเกิลติดอันดับในกูเกิลเทรนด์ (Google Trends) 484 คำ โดยคำในการค้นหานั้น อ้างอิงจากงานวิจัย Preis et al. (2013) และแหล่งข้อมูลจากเว็บไซต์ Finnomena SET Krungsri Encyclopedia และ stock2morrow

#### 1.4 ระยะเวลาการดำเนินงานทำงานวิจัย

ในการดำเนินงานวิจัยครั้งนี้มีขั้นตอนดังนี้

1. ศึกษาความเป็นมาและกำหนดขอบเขตของปัญหา
2. ศึกษาเทคนิควิธีการที่จะนำมาใช้แก้ปัญหา
3. ศึกษาปัจจัยที่เกี่ยวข้องและงานวิจัยเกี่ยวกับการเลือกหลักทรัพย์ลงทุน
4. รวบรวมและศึกษาข้อมูลตลาดหลักทรัพย์ของประเทศไทย และค้นหาในกูเกิลติดอันดับในกูเกิลเทรนด์ (Google Trends)
5. คัดเลือกปัจจัยที่ส่งผลต่อการเลือกหลักทรัพย์ลงทุน
6. สร้างแบบจำลอง เพื่อพยากรณ์ความเคลื่อนไหวของหลักทรัพย์
7. ตรวจสอบประสิทธิภาพของแบบจำลองพยากรณ์ความเคลื่อนไหวของหลักทรัพย์
8. สร้างแบบจำลอง เพื่อคำนวณหาอัตราส่วนในการลงทุน
9. ตรวจสอบประสิทธิภาพของแบบจำลอง คำนวณหาอัตราส่วนในการลงทุน
10. รายงานผลการดำเนินงานกับอาจารย์ที่ปรึกษา
11. แก้ไขรายงานที่ผ่านการตรวจสอบจากอาจารย์ที่ปรึกษา
12. ตรวจสอบความถูกต้องของการจัดทำรายงานฉบับสมบูรณ์
13. ส่งงานวิจัยตีพิมพ์
14. นำเสนองานวิจัยต่อคณะกรรมการ

#### 1.5 ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัยนี้ มีดังนี้

เพื่อให้ได้แบบจำลองการตัดสินใจเลือกการลงทุนในหลักทรัพย์ และสัดส่วนการลงทุนที่เหมาะสมในช่วงเวลาต่าง ๆ เพื่อลดความเสี่ยงในการลงทุนและเพิ่มผลตอบแทนที่สูงขึ้น

## บทที่ 2

# ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในบทนี้คือทฤษฎีและงานวิจัยที่เกี่ยวข้องซึ่งประกอบไปด้วย ข้อมูล (Data), การเลือกปัจจัยที่ส่งผล (Feature Selection), แบบจำลอง (Modeling), การประเมินผล (Evaluation) และงานวิจัยที่เกี่ยวข้อง (Research Methodology)

### 2.1 ข้อมูล (Data)

ข้อมูลในงานวิจัยนี้เป็นประเภทข้อมูลที่มีโครงสร้าง (Structured Data) หมายถึงกลุ่มข้อมูลที่มีโครงสร้างที่กำหนดไว้ ซึ่งจัดระเบียบในรูปแบบตารางซึ่ง สามารถอธิบายลักษณะเป็นคอลัมน์ของตารางได้ โครงสร้างดังกล่าวทำให้สามารถเข้าถึงข้อมูลได้ง่าย และการค้นหาข้อมูลที่เรียบง่ายหรือไม่ซับซ้อน แต่สามารถทำได้โดยไม่ต้องใช้ความพยายามเพิ่มเติม โดยเก็บข้อมูลจาก 2 แหล่ง คือ ข้อมูลคำค้นหาจากกูเกิลเทรนด์ (Google Trends Search Information) และ ข้อมูลตลาดหลักทรัพย์ (Market Information)

#### 2.1.1. ข้อมูลคำค้นหาจากกูเกิลเทรนด์ (Google Trends Search Information)

กูเกิลเทรนด์ (Google Trends) คือ เว็บไซต์หนึ่งของกูเกิล (Google) ที่ถูกผลิตออกมาเพื่อช่วยเหลือการตรวจสอบความนิยมของคำสำคัญ (Keywords) บนโลกออนไลน์ โดยที่ของคำสำคัญนั้นจะเป็นคำว่าอะไรก็ได้ อาทิเช่น ชื่อสินค้า ชื่อบุคคล หรืออาจจะเป็นชื่อเว็บไซต์ ซึ่งสามารถตรวจสอบความนิยมเหล่านี้ตามตำแหน่งที่ต้องการได้ ไม่ว่าจะเป็นเทรนด์ระดับโลก เทรนด์ระดับประเทศ หรือ เทรนด์ระดับจังหวัด อีกทั้งยังสามารถดูเทรนด์ย้อนอดีตรายวันได้อีกด้วย หรือแม้แต้ออนดูความนิยมในหลาย ๆ ปีที่ผ่านมาเพื่อวิเคราะห์โอกาสเติบโตของสิ่งนั้นได้อีกด้วย โดยกูเกิลเทรนด์นั้นตอบโจทย์สำหรับผู้ใช้งานที่จำเป็นต้องพึ่งพาอาศัยหรือได้ประโยชน์จากการจับเทรนด์ ไม่ว่าจะเป็นนักการตลาด เจ้าของร้านค้าออนไลน์ หรือแม้แต่คนที่ต้องการเริ่มทำบล็อก (Blog) หรือ วิล็อก (Vlog) ของตัวเองแต่ไม่รู้ว่าควรจะนำไอเดียไหนมาขายดี หรือเอาไว้อัปโหลดวิดีโอที่มีคนค้นหาสูงสุดในแต่ละวันหรือแต่ละช่วง ทางด้านการตลาด (Marketers) จะช่วยวางแผนธุรกิจ เพื่อพัฒนาแผนธุรกิจในแต่ละช่วง กูเกิล เทรนด์สามารถเข้าช่วยวิเคราะห์ทิศทางตลาดบนโลกออนไลน์ได้ดียิ่งขึ้น และด้านการเงิน Huang et al. (2019) ได้ใช้คำค้นหาทางอินเทอร์เน็ต เพื่อมองหาความสนใจของนักลงทุน และส่งผลต่อการเคลื่อนไหวดัชนี S&P 500 พบว่าการเคลื่อนไหวในทิศทางต่าง ๆ ของ S&P 500 นั้นเปลี่ยนแปลงตามปริมาณการค้นหาโดยขึ้นอยู่กับคำเฉพาะที่กำลังค้นหา และตามส่วนขยายความรู้สึกของคำนั้น

#### 2.1.2. ข้อมูลตลาดหลักทรัพย์ (Stock Market Information)

ด้วยแนวทางการวิเคราะห์ทางเทคนิคราคาหลักทรัพย์จะสะท้อนข้อมูลทั้งหมดที่จำเป็นในการทำความเข้าใจพฤติกรรมของตลาด ด้วยวิธีนี้สิ่งที่สำคัญคือการวิเคราะห์อนุกรมเวลา (Time Series) เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Analyze) ที่สอดคล้องกับราคา โดยทั่วไปข้อมูลนี้เป็นข้อมูลสาธารณะสามารถดาวน์โหลดได้จากหน้าเพจของตลาดหลักทรัพย์

ในบางงานวิจัยจะใช้ข้อมูลหลักทรัพย์รายวันซึ่งประกอบด้วยราคาเปิด (OP) ราคาปิด (CP) ราคาสูงสุด (MAX) และราคาต่ำสุด (MIN) และปริมาณ (VOL) ของธุรกรรมที่ดำเนินการ ราคาปิดเป็นข้อมูลที่ใช้กันมากที่สุด แต่ปริมาณและช่วงต่าง ๆ ก็แสดงมูลค่าในการทำนายเช่นกัน การศึกษาส่วนใหญ่ใช้ระยะเวลาของข้อมูลประมาณ 1,000 วัน ซึ่งสามารถจัดการได้ง่ายสำหรับการเรียนรู้ด้วยเครื่องจักร (Machine Learning)

#### ตัวชี้วัดทางเทคนิค (Technical Indicators)

โดยทั่วไปแล้ว ตัวชี้วัดทางเทคนิคจะได้รับการศึกษาและวางแผนบนแผนภูมิเพื่อช่วยในการให้ข้อมูลทิศทางเกี่ยวกับราคาของหลักทรัพย์ เมื่อใช้ดัชนีบนกราฟ ดัชนีจะแสดงเป็นเส้นและแสดงค่าระหว่างเส้นเหล่านั้นควบคู่ไปกับราคาเพื่อกำหนดทิศทางและการเคลื่อนไหวที่เป็นไปได้ถัดไป

ตัวบ่งชี้ทางเทคนิคสามารถใช้เพื่อสรุปพฤติกรรมหรือแนวโน้มของอนุกรมเวลา และสามารถคาดการณ์ราคาหลักทรัพย์ได้ (Alfonso & Ramirez, 2020; Anghel, 2015; Basak et al., 2019; Bhargavi. R, 2017; Bustos & Pomares-Quimbaya, 2020; Dash & Dash, 2016; Perry, 2011; Vaidya, 2018) การศึกษานี้แสดงให้เห็นว่าเทคนิคต่างๆ สามารถทำการสรุปแทนที่จะเป็นชุดราคาหลักทรัพย์แบบเต็มเวลาได้อย่างไร ซึ่งทำให้การเรียนรู้ของเครื่องจักร (Machine Learning) เรียนรู้ง่ายขึ้น ตัวบ่งชี้ทางเทคนิคแสดงไว้ในตารางที่ 2.1

ตารางที่ 2.1 สมการตัวบ่งชี้ทางเทคนิค (Technical Indicators)

Technical Indicators	Formula
Simple Moving Average (SMA)	$\frac{C_t + C_{t-1} + C_{t-2} + \dots + C_{t-n-1}}{n} \quad (2.1)$
Weighted Moving Average (WMA)	$\frac{(n)(C_t) + (n-1)(C_{t-1}) + \dots + C_{t-n-1}}{(n) + (n-1) + \dots + 1} \quad (2.2)$
Exponential Moving Average (EMA)	$\alpha C_t + (1-\alpha)C_{t-1} + (1-\alpha)^2 C_{t-2} + \dots + (1-\alpha)^{t-n+2} C_{t-n+2} + (1-\alpha)^{t-n+1} EMA_{t-n+1} \quad (2.3)$
Moving Average Convergence Divergence (MACD)	$EMA_{12}_t - EMA_{26}_t \quad (2.4)$
Relative Strength Index (RSI)	$100 - \frac{100}{1 + \left( \frac{\sum_{i=0}^{n-1} UP_{t-i}/n}{\sum_{i=0}^{n-1} DW_{t-i}/n} \right)} \quad (2.5)$
Stochastic Oscillator K (K)	$\frac{C_t - L_{t-n}}{H_{t-n} - L_{t-n}} \times 100 \quad (2.6)$
Stochastic Oscillator D (D)	$\frac{\sum_{t=0}^{n-1} K_t}{n} \quad (2.7)$

หมายเหตุ:  $C_t$  คือราคาปิด,  $L_t$  คือราคาต่ำ,  $H_t$  คือราคาสูง ณ เวลา  $t$ ,  $\alpha$  คือ smoothing factor,  $UP_t$  หมายถึงการเปลี่ยนแปลงของราคาที่สูงขึ้น ในขณะที่  $DW_t$  คือการเปลี่ยนแปลงของราคาที่ลดลง ณ เวลา  $t$

ตัวบ่งชี้ทางเทคนิคที่ใช้ในแบบจำลองการคาดการณ์ มีดังต่อไปนี้

- 1) ค่าเฉลี่ยเคลื่อนที่ (Moving Average: MA) ตัวบ่งชี้ทางเทคนิคที่ระบุแนวโน้มเฉลี่ยในช่วงเวลาที่กำหนดโดยการปรับให้เรียบและกรองสัญญาณผิดปกติต่าง ๆ และค่าเฉลี่ยเคลื่อนที่ที่เกิดขึ้นจากการใช้ราคาปิดเฉลี่ยในช่วงเวลาที่กำหนดค่าเฉลี่ย
- 2) ค่าเฉลี่ยเคลื่อนที่แบบง่าย (Simple Moving Average: SMA) ค่าเฉลี่ยนี้เหมาะสมในสถานการณ์ความผันผวนปกติในอนุกรมเวลา เพื่อช่วยให้นักวิเคราะห์ข้อมูลตีความรูปแบบและบริบทของสภาวะตลาดหลักทรัพย์ได้ดีขึ้น (Perry, 2011)
- 3) ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก (Weighted Moving Average: WMA) ค่าเฉลี่ยนี้เหมาะสมในสถานการณ์ที่มีความผันผวนสูงในอนุกรมเวลา เพื่อช่วยให้นักวิเคราะห์ข้อมูลตีความรูปแบบและบริบทของแนวโน้มในอดีตได้ดีขึ้น (Perry, 2011)
- 4) ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักแบบ Exponential (Exponential Moving Average: EMA) ค่าเฉลี่ยนี้จัดลำดับความสำคัญของข้อมูลล่าสุด ดังนั้นจึงตอบสนองต่อการเปลี่ยนแปลงของราคาได้เร็วกว่าค่าเฉลี่ยเคลื่อนที่อย่างง่าย (Bustos & Pomares-Quimbaya, 2020; Praekhaow, 2010)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 5) Moving Average Convergence Divergence (MACD) MACD indicates มีแนวคิดในการวาดเส้นค่าเฉลี่ยเคลื่อนที่ 2 เส้นพร้อมกัน จากนั้นจึงวิเคราะห์ลักษณะของเส้นค่าเฉลี่ยเคลื่อนที่ทั้งสองเส้น จุดที่น่าสนใจเป็นพิเศษของ MACD คือเป็น indicator ที่สามารถให้ข้อมูลได้ 2 มุมมองพร้อมกันคือทิศทางแนวโน้มของราคาหลักทรัพย์ และโมเมนตัมของราคาหลักทรัพย์ มาตรฐาน MACD คือ EMA 12 ลบด้วย EMA 26 (Alfonso & Ramirez, 2020; Anghel, 2015; Bustos & Pomares-Quimbaya, 2020; Dash & Dash, 2016) โดย EMA 12 และ EMA 26 คือ เส้นค่าเฉลี่ยเคลื่อนที่ลักษณะเส้นเป็นแบบ Exponential โดยใช้ราคาปิดย้อนหลัง 12 วัน และ 26 วัน ตามลำดับ
- 6) ดัชนีความสัมพัทธ์สัมพัทธ์ (The Relative Strength Index, RSI) เป็นตัวบ่งชี้ทางเทคนิคที่ระบุว่าสินทรัพย์แข็งแกร่งหรืออ่อนแอเมื่อเทียบกับราคาปิดล่าสุด นอกจากนี้ยังใช้เพื่อกำหนดความน่าจะเป็นของการกลับตัวของแนวโน้ม (Bhargavi. R, 2017)
- 7) ออสซิลเลเตอร์สโตแคสติก (Stochastic Oscillator) คือ Stochastic Oscillator เป็นตัวบ่งชี้ทางเทคนิคแบบโมเมนตัมที่ใช้ในการเปรียบเทียบราคาปิดในช่วงเวลาที่กำหนดกับช่วงราคาต่าง ๆ ออสซิลเลเตอร์นี้มีความไวอย่างมากต่อการเปลี่ยนแปลงของราคาในตลาด แม้ว่าความผันผวนของตัวบ่งชี้จะสามารถทำให้ราบบรรลุขึ้นได้บ้าง โดยการเปลี่ยนช่วงเวลาที่ใช้วัด Stochastic Oscillator ที่ใช้บ่อยที่สุดคือ Stochastic Oscillator K (K) และ Stochastic Oscillator D (D) (Vaidya, 2018) เส้น K จะเปรียบเทียบค่าต่ำสุดและค่าสูงสุดในช่วงเวลาที่กำหนดเพื่อกำหนดช่วงราคา จากนั้นแสดงราคาปิดล่าสุดเป็นเปอร์เซ็นต์ของช่วงนี้ เส้น D คือเส้นค่าเฉลี่ยเคลื่อนที่ของ K

## 2.2 การเลือกปัจจัยที่ส่งผล (Feature Selection)

### ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson Correlation Coefficient, $r$ )

สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เป็นการศึกษาค่าความสัมพันธ์เชิงเส้น (Linear Relationship) ระหว่างตัวแปร 2 ตัวหรือข้อมูล 2 ชุด โดยมีข้อตกลงเบื้องต้น คือ ตัวแปรทั้งสองต้องเป็นตัวแปรต่อเนื่อง อยู่ในมาตราอันตรภาคชั้น (Interval Scale) ขึ้นไป และมีการแจกแจงปกติสองตัวแปร (Bivariate Normal Distribution) โดยค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน จะมีค่าอยู่ระหว่าง - 1 ถึง + 1 ถ้า เข้าใกล้  $\pm 1$  หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันอยู่ในระดับมาก ถ้าสัมประสิทธิ์สหสัมพันธ์มีค่าเข้าใกล้ 0 หมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันในระดับน้อยหรือไม่สัมพันธ์กันเลย เครื่องหมายบวกลบ ( $\pm$ ) จะแสดงทิศทางของความสัมพันธ์ว่ามีทิศทางเป็นแบบใด โดยถ้าสัมประสิทธิ์สหสัมพันธ์มีเครื่องหมายบวก (+) หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์

ไปทิศทางเดียวกัน หากมีเครื่องหมายลบ (-) หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์ไม่ไปทิศทางเดียวกันหรือตรงกันข้าม สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน สามารถคำนวณได้ดังนี้

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (2.8)$$

เมื่อ	$r_{xy}$	คือ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันระหว่างตัวแปร $x$ และ $y$
	$\sum x$	คือ ผลรวมของข้อมูลที่วัดจากตัวแปรที่ 1 หรือตัวแปร $x$
	$\sum y$	คือ ผลรวมของข้อมูลที่วัดจากตัวแปรที่ 2 หรือตัวแปร $y$
	$\sum xy$	คือ ผลรวมของผลคูณของตัวแปรที่ 1 และ 2
	$\sum x^2$	คือ ผลรวมของกำลังสองของข้อมูลที่วัดจากตัวแปรที่ 1
	$\sum y^2$	คือ ผลรวมของกำลังสองของข้อมูลที่วัดจากตัวแปรที่ 2
	$n$	คือ จำนวนของข้อมูล

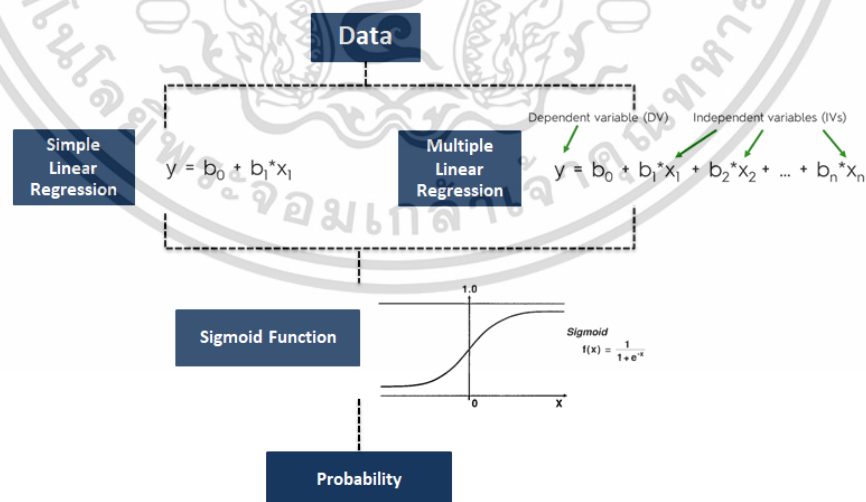
### 2.3 แบบจำลอง (Modeling)

การสร้างแบบจำลองในการศึกษาครั้งนี้ประกอบด้วยแบบจำลองการจำแนกประเภท (Classification Model) ซึ่งเป็นแบบจำลองการเรียนรู้แบบมีผู้สอนเพื่อสร้างแบบจำลองที่ต้องมีตัวแปรที่ใช้วัดเป้าหมายเป็นจุดเริ่มต้นในการเรียนรู้ เป้าหมายของการจำแนกจะเป็นแบบกลุ่มหรือไม่ต่อเนื่อง เช่น Yes/No, A/B/C เป็นต้น ดังนั้นในการประเมินผลลัพธ์ที่ได้จากแบบจำลองการจำแนกจึงสามารถวัดความแม่นยำ (Accuracy) ได้ เช่น Confusion Matrix โดยสร้างแบบจำลองต่อไปนี้จะใช้เพื่อทำนายการเลือกหลักสูตรฯ ในการศึกษาเลือกการถดถอยโลจิสติก (Logistic Regressions) การสุ่มป่าไม้ (Random Forest) และ Extreme Gradient Boosting (XGBoost) สาเหตุที่เลือกแบบจำลองทั้งสามนี้เนื่องจาก (1) แบบจำลองทั้งสามถูกใช้โดยทั่วไปและอ้างอิงถึงในการศึกษาที่เกี่ยวข้องจำนวนมาก (Ananthakumar & Sarkar, 2017; Basak et al., 2019; Geurts et al., 2006; Ghatasheh, 2014; Preis et al., 2013; Sadowsky, 2021) และ (2) สามารถจัดการข้อมูลความไม่สมดุลตามที่กล่าวไว้ (Le et al., 2022; Wang et al., 2022)

### 2.3.1 การจำแนกประเภท (Classification)

**การถดถอยโลจิสติก (Logistic Regression)** ถูกจัดเป็นแบบจำลองประเภทการเรียนรู้แบบมีผู้สอน (Supervised Learning) ซึ่งเป็นแบบจำลองที่จะนำมาใช้ทำนายความน่าจะเป็นที่จะเกิดเหตุการณ์ที่สนใจจากชุดข้อมูลของตัวแปรอิสระ (Independent Variables) ที่เหมาะสมและเป็นอัลกอริทึมที่มีพื้นฐานการสร้างมาจากสมการทางคณิตศาสตร์ซึ่งแบ่งตามชนิดของตัวแปรอิสระสามารถแบ่งได้ 2 ประเภท คือ การถดถอยโลจิสติกแบบง่าย (Simple Logistic Regression) มีตัวแปรอิสระเพียงตัวเดียว และ การถดถอยพหุแบบโลจิสติก (Multiple Logistic Regression) มีตัวแปรอิสระหลายตัว โดยมีการนำมาประยุกต์ใช้อย่างแพร่หลาย อาทิเช่น การทำนายโอกาสที่ลูกค้าจะเกิดหนี้เสียให้กับบริษัทของธุรกิจประเภทธนาคาร การทำนายโอกาสที่ลูกค้าจะย้ายเครือข่ายของธุรกิจประเภทโทรศัพท์ และมีการใช้ความสามารถของแบบจำลอง ในการประยุกต์ทางการตลาดโดยทำแบบจำลองในการระบุการแบ่งของส่วนตลาด (Market Segmentation) เพื่อเพิ่มโอกาสในการเสนอแคมเปญ (Campaign) ให้กับส่วนของตลาดในแต่ละส่วนได้ (Ananthakumar & Sarkar, 2017; Robles et al., 2008) ขั้นตอนในการสร้างแบบจำลองหลักแสดงดังรูปที่ 2.1 และมีรายละเอียดดังนี้

1. การนำชุดข้อมูลมาสร้างสมการถดถอยตัวแปรเดียว (Simple Linear Regression) หรือ สมการถดถอยหลายตัวแปร (Multiple Linear Regression) โดยขึ้นอยู่กับตัวแปรอิสระที่ถูกนำมาใช้ในประเภทของงานที่ทำ
2. นำสมการถดถอยที่ได้ไปเข้าสู่ฟังก์ชันซิกมอยด์ (Sigmoid) เพื่อปรับค่าที่ได้ให้อยู่ในช่วง 0-1 เนื่องจากสมการถดถอยที่ได้อาจมีค่ามากกว่า 1 หรือน้อยกว่า 0 ได้ ซึ่งกฎความน่าจะเป็น จะต้องอยู่ระหว่าง 0-1 เท่านั้น จึงได้มีการนำฟังก์ชันนี้มาใช้งาน
3. เมื่อผ่านฟังก์ชันซิกมอยด์จะได้ความน่าจะเป็นของเหตุการณ์ที่สนใจ



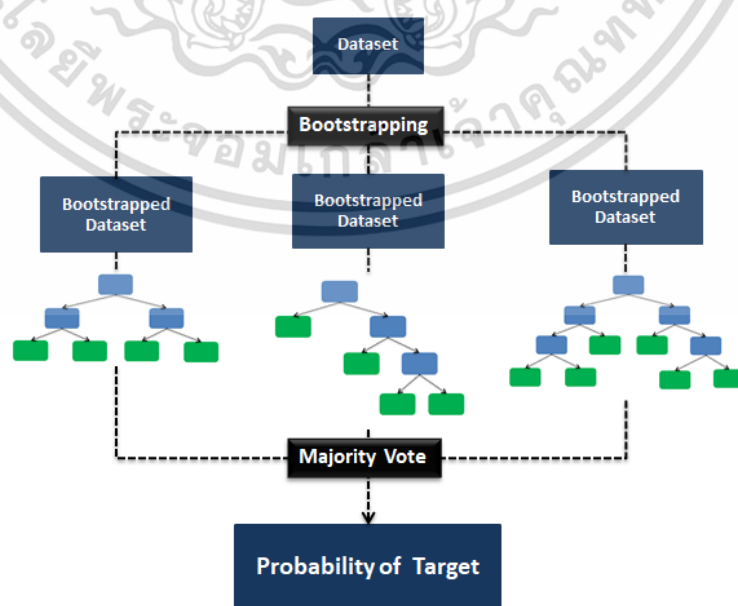
รูปที่ 2.1 แผนภาพขั้นตอนการสร้างแบบจำลองถดถอยโลจิสติก (Logistic Regression)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**การสุ่มป่าไม้ (Random Forest)** เป็นแบบจำลองที่มีพื้นฐานการสร้างมาจาก Tree-Based นั่นคือสร้างมาจากต้นไม้ตัดสินใจ (Decision Tree) หลาย ๆ ต้นรวมกัน ซึ่งเป็นแบบจำลองที่มีการนำเทคนิคความฉลาดของฝูงชน (Wisdom of the Crowd) มาประยุกต์ใช้กับการทำนายผลข้อมูลโดยถูกเรียกว่า Ensemble Technique และจัดอยู่ในประเภท Bootstrap Aggregation คือการสุ่มตัวอย่างข้อมูลออกมาแล้วทำการสร้างแบบจำลองทำนายผล ซึ่งการสุ่มใช้วิธีการสุ่มแบบแทนที่ (Random with Replacement) จึงทำให้ขนาดของชุดข้อมูลที่ได้นั้นมีขนาดเท่าชุดข้อมูลเดิม จุดเด่นของแบบจำลองนี้ คือให้ความถูกต้อง (Accuracy) สูง และความแปรปรวน (Variance) น้อย ซึ่งดีกว่าต้นไม้ตัดสินใจเพียงต้นเดียวแต่ก็ใช้เวลาในการคำนวณที่ซับซ้อนมากกว่าหลายเท่า (Basak et al., 2019; Sadorsky, 2021) โดยมีการศึกษาที่ได้นำแบบจำลองนี้ไปพยากรณ์ความเสี่ยงทางการใช้บัตรเครดิต (Credit Risk Prediction) และมีการเปรียบเทียบกับแบบจำลองอื่น ๆ มากกว่า 10 แบบจำลอง (Ghatasheh, 2014) พบว่าแบบจำลอง สุ่มป่าไม้ นั้นให้ค่าความถูกต้อง (Accuracy) ค่าความไว (Sensitivity) และ ค่า F-Measure สูงที่สุด

ขั้นตอนในการสร้างแบบจำลองหลักแสดงดังรูปที่ 2.2 และมีรายละเอียดดังนี้

1. สร้างชุดข้อมูลใหม่ด้วยวิธีการสุ่มแบบแทนที่โดยข้อมูลชุดใหม่จะถูกเรียกว่า Bootstrapped Dataset
2. สร้างต้นไม้ตัดสินใจโดยใช้ Bootstrapped Dataset ที่ได้จากการกำหนดจำนวนของต้นไม้ตัดสินใจที่จะสร้าง พร้อมทั้งจำนวนตัวแปรอิสระที่จะถูกพิจารณาในการสร้างต้นไม้ตัดสินใจแต่ละขั้นตอน
3. แต่ละต้นไม้ตัดสินใจจะสร้างผลลัพธ์
4. ผลลัพธ์สุดท้ายจะพิจารณาจากการลงคะแนนเสียงข้างมาก



รูปที่ 2.2 แผนภาพขั้นตอนการสร้างแบบจำลองการสุ่มป่าไม้ (Random Forests)

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น เมื่ออนุญาตให้เผยแพร่ขึ้นด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**Extreme Gradient Boosting (XGBoost)** เป็นแบบจำลองที่จัดอยู่ในประเภท Boosting คือการนำ Weak Classifier ที่มีความแม่นยำต่ำมาทำนายผลข้อมูล จากนั้นจะนำ Weak Classifier ที่ได้มาแก้ไขความคลาดเคลื่อน (Error) ของตัวแรกที่ถูกรัง โดยผลรวมของการทำนายจะเกิดเป็นการจำแนกหมวดหมู่ใหม่ขึ้นมาโดยจะทำซ้ำไปเรื่อย ๆ จนกว่าจะได้แบบจำลองที่ดีที่สุดจากผลรวมของการทำนาย ซึ่งแบบจำลองนี้เป็นการนำ Ensemble Technique มาประยุกต์ใช้เช่นเดียวกับแบบจำลอง Random Forests Classification แต่กระบวนการสร้างแตกต่างกันอย่างชัดเจน จุดเด่นของแบบจำลองชนิดนี้ คือสามารถพยากรณ์ข้อมูลได้ซับซ้อนและแม่นยำมากเมื่อเทียบกับแบบจำลองที่ถูกจัดอยู่ในประเภทเรียนรู้แบบมีผู้สอนอื่น ๆ สามารถทำงานได้ดีทั้งการถดถอย (Regression) และการจำแนกหมวดหมู่ (Classification) แต่การคำนวณและการประมวลผลนั้นใช้เวลานานเนื่องจากแบบจำลองเป็นการทำงานแบบลำดับขั้น ซึ่งต่างจาก Bootstrap Aggregation ที่สามารถทำงานพร้อมกันได้ (Geurts et al., 2006)

ขั้นตอนในการสร้างแบบจำลองหลักแสดงดังรูปที่ 2.3 และมีรายละเอียดดังนี้

1. กำหนดเกณฑ์พยากรณ์เริ่มต้นโดยใช้หลักเกณฑ์ทั่วไปซึ่งแบ่งตามความน่าจะเป็นที่มากกว่าหรือเท่ากับ 0.5 จะถูกพยากรณ์ว่าเป็น Positive Class และความน่าจะเป็นที่น้อยกว่า 0.5 แต่มากกว่า 0 นั้นเป็น Negative Class หรือแบ่งตามสัดส่วนของประเภท ที่ต้องการตามประเภทงาน
2. หาค่าคลาดเคลื่อน (Residual) จากการแปลงค่าจริงของตัวแปรพยากรณ์โดยกำหนดให้ Positive Class เป็น 1 และ Negative Class เป็น 0 จากนั้นนำค่าจริงของตัวแปรพยากรณ์ที่ได้ลบกับเกณฑ์ที่ได้จากขั้นตอนที่ 1
3. สร้างต้นไม้ตัดสินใจจากตัวแปรอิสระทั้งหมดโดยต้นไม้ตัดสินใจนั้นจะเป็นการพยากรณ์ค่าคลาดเคลื่อนที่ได้จากขั้นตอนที่ 2 โดยปกติจะกำหนดจำนวนสูงสุดของโหนดใบของต้นไม้ตัดสินใจอยู่ระหว่าง 8 ถึง 32 ใบ ซึ่งจะช่วยลดการเกิด Overfitting ได้
4. คำนวณผลรวมของ Residual ของโหนดใบทุกใบ โดยคำนวณจากสูตร

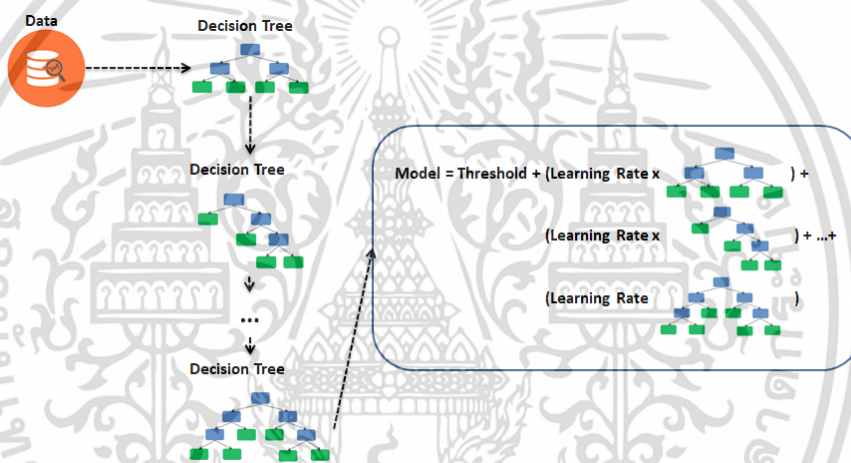
$$\text{Residual}_{\text{New}} = \frac{\sum \text{Residual}}{\sum (\text{Previous Probability} \times (1 - \text{Previous Probability})) + \lambda} \quad (2.9)$$

โดย

Residual	คือ ค่าความคลาดเคลื่อนของแต่ละโหนดใบ
Previous Probability	คือ ค่าความน่าจะเป็นก่อนหน้าของการคำนวณในแต่ละรอบ
$\lambda$	คือ Regularization Parameter

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. รวมเกณฑ์ที่ได้จากขั้นตอนที่ 1 โดยผ่านฟังก์ชันอัตราต่อรองลอการิทึม (Log Odds) กับค่าเกณฑ์ที่กำหนดไว้พร้อมกับต้นไม้ตัดสินใจที่ได้จากขั้นตอนที่ 3 เข้าด้วยกัน โดยจะใช้อัตราเรียนรู้ (Learning Rate) ซึ่งเป็นพารามิเตอร์ที่ช่วยในการปรับต้นไม้ตัดสินใจไม่ให้ Overfit กับข้อมูลมากเกินไป ซึ่งมีค่าอยู่ระหว่าง 0-1 และจะได้ Weak Classifier ออกมา
6. คำนวณค่า Residual ใหม่อีกครั้งจาก Weak Classifier ในขั้นตอนที่ 5 โดยนำข้อมูลทั้งหมดไปคำนวณ จากนั้นนำค่าที่ได้ของแต่ละข้อมูลไปผ่านฟังก์ชัน Sigmoid ซึ่งจะได้ค่าความน่าจะเป็นของแต่ละข้อมูลทั้งหมด
7. ทำขั้นตอนที่ 2 ไปเรื่อย ๆ จนกระทั่งค่า Residual ไม่เปลี่ยนแปลงจากเดิมหรือครบตามจำนวนต้นไม้ตัดสินใจที่ได้กำหนดไว้



รูปที่ 2.3 แสดงแผนภาพขั้นตอนการสร้างแบบจำลอง XGBoost

### 2.3.2 การทำให้เหมาะที่สุด (Optimization)

มาตรวัดตามตัวแบบ Sharpe (Sharpe Ratio) เป็นการวัดผลตอบแทนของหลักทรัพย์ที่มากกว่าหรือเหนือกว่าอัตราผลตอบแทนของหลักทรัพย์ที่ปราศจากความเสี่ยง สำหรับนักลงทุนในการลงทุน ไม่ว่าจะเป็นการลงทุนในตลาดหลักทรัพย์ (Stock Market) หรือตลาดเทรดเงิน (Forex Exchange) จะมีการวัดประสิทธิภาพของพอร์ตการลงทุนโดยใช้เครื่องมือหรือดัชนีชี้วัดต่าง ๆ เป็นตัววัดความสำเร็จในการบริหารพอร์ตการลงทุน นักลงทุนจะดูที่อัตราผลตอบแทนต่อปีแค่ขณะตลาดได้ก็เพียงพอแล้ว แต่ถ้าจะลงทุนในระยะยาวจะพิจารณาเพียงแค่ผลตอบแทน (Return) อย่างเดียวไม่เพียงพอ จะต้องมามีเครื่องมือที่来帮助ชี้วัดประสิทธิภาพของพอร์ตการลงทุน (Chen et al., 2021; Gatta et al., 2023; Gunjan & Bhattacharyya, 2022) นั่นก็คือวิธีคำนวณค่ามาตรวัดตามตัวแบบ Sharpe ซึ่งแสดงดังสมการที่ (2.10)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{Sharpe Ratio} = \frac{\sum_{i=1}^n r_i x_i - r_f}{\sqrt{\sum_{i=1}^n \sum_{j=1}^n x_i x_j \sigma_{ij}}} \quad (2.10)$$

โดยที่

$r_i$	แทน ผลตอบแทนของหลักทรัพย์ตัวที่ $i$
$x_i$	แทน สัดส่วนในการลงทุนในหลักทรัพย์ตัวที่ $i$ ต่อการลงทุนทั้งหมดในหลักทรัพย์ของนักลงทุน
$r_f$	แทน อัตราดอกเบี้ยที่ปราศจากความเสี่ยง เช่น อัตราดอกเบี้ยพันธบัตร (Risk Free Rate)
$\sigma_{ij}$	แทน ความแปรปรวนร่วมของผลตอบแทนระหว่างหลักทรัพย์ตัวที่ $i$ และตัวที่ $j$

ในการนำค่ามาตรวัดตามตัวแบบชาร์ปเรโซ มาใช้เพื่อเปรียบเทียบพอร์ตการลงทุน ควรใช้เปรียบเทียบระหว่างพอร์ตการลงทุนที่มีนโยบายการลงทุนเหมือนกันเป็นสำคัญ โดยพอร์ตที่มีค่ามาตรวัดตามตัวแบบชาร์ปเรโซ สูงกว่า แสดงว่า พอร์ตนั้นสร้างผลตอบแทนได้สูงกว่า ณ หนึ่งหน่วยความเสี่ยงที่เท่ากัน

### การปรับปรุงความเสี่ยงแบบเทียบค่าความเสี่ยง (Risk Parity Optimization - RPO)

เป็นวิธีการจัดการลงทุนที่มุ่งหวังให้มีการกระจายความเสี่ยงที่เท่าเทียมกันในสินทรัพย์ต่าง ๆ ในพอร์ตการลงทุน การลงทุนแบบ Risk Parity มองหาวิธีการแบ่งสัดส่วนการลงทุนในสินทรัพย์ต่าง ๆ ในพอร์ต โดยทำให้มีค่าความเสี่ยงเทียบเท่ากันในทุก ๆ สินทรัพย์ แทนการแบ่งสัดส่วนตามมูลค่าเงินลงทุน โดยปรับให้สินทรัพย์ที่มีความเสี่ยงสูงมีสัดส่วนการลงทุนน้อยลง ในขณะที่สินทรัพย์ที่มีความเสี่ยงต่ำมีสัดส่วนการลงทุนสูงขึ้น ซึ่งการกระจายความเสี่ยงแบบนี้ถือว่ามีเหตุผลและสามารถลดความผันผวนของผลตอบแทนได้ (Chaves et al., 2011; Gambeta & Kwon, 2020) กระบวนการปรับปรุงความเสี่ยงแบบเทียบค่าความเสี่ยงนั้นใช้วิธีการทางคณิตศาสตร์เพื่อคำนวณและปรับแก้สัดส่วนการลงทุนในสินทรัพย์ โดยการใช้งานข้อมูลปริมาณความผันผวนราคา (Volatility) และความสัมพันธ์ระหว่างสินทรัพย์ต่าง ๆ เป้าหมายของ RPO คือให้ผู้ลงทุนสามารถทำให้ความเสี่ยงกระจายอย่างเท่าเทียมกันในทุก ๆ สินทรัพย์ โดยไม่ต้องพึ่งพาเพียงค่ามูลค่าเงินลงทุน นี่เป็นวิธีที่ใช้กันอย่างแพร่หลายในการจัดการพอร์ตการลงทุนในสถาบันการเงินและองค์กรทางการเงิน หากต้องการดำเนินการ RPO จะต้องระบุหลักการและพารามิเตอร์ที่เกี่ยวข้อง เช่น ช่วงเวลาที่สำคัญในการวิเคราะห์ความผันผวนราคา หรือวิธีการคำนวณค่าความสัมพันธ์ระหว่างสินทรัพย์ต่าง ๆ อย่างไร เนื่องจากเป็นกระบวนการ

ทางคณิตศาสตร์ที่ซับซ้อน บางครั้งอาจใช้การคำนวณและการปรับแก้ที่ซับซ้อน จึงควรใช้วิธีการทางคอมพิวเตอร์หรือโปรแกรมที่สามารถช่วยในกระบวนการปรับปรุงเหล่านี้ได้

$$\text{Risk Parity} = \sum_{i=1}^n x_i^2 \sigma_i^2 \quad (2.11)$$

โดยที่

- $x_i$  แทน สัดส่วนในการลงทุนในหลักทรัพย์ตัวที่  $i$  ต่อการลงทุนทั้งหมดในหลักทรัพย์ของนักลงทุน
- $\sigma_{ij}$  แทน ความแปรปรวนร่วมของผลตอบแทนระหว่างหลักทรัพย์ตัวที่  $i$

**Sequential Least Squares Programming (SLSQP)** เป็นอัลกอริทึมการปรับให้เหมาะสมซึ่งใช้เพื่อแก้ปัญหาค่าเหมาะที่สุดแบบไม่เชิงเส้น เดิมทีเปิดตัวโดย Dieter Kraft ในปี 1988 SLSQP ได้รับการออกแบบมาเพื่อจัดการกับปัญหาที่ฟังก์ชันวัตถุประสงค์และข้อจำกัดไม่เป็นเชิงเส้น โดย SLSQP ประมาณวิธีแก้ปัญหาค่าเหมาะที่สุด โดยแก้ลำดับของปัญหาย่อยของโปรแกรมกำลังสอง ในการวนซ้ำแต่ละครั้ง ระบบจะสร้างการประมาณกำลังสองของฟังก์ชันวัตถุประสงค์และข้อจำกัดรอบ ๆ จุดปัจจุบัน จากนั้นแก้ปัญหาย่อยเพื่อค้นหาการวนซ้ำถัดไป อัลกอริทึมจะอัปเดตการวนซ้ำจนกว่าจะถึงเกณฑ์การบรรจบกันหรือถึงจำนวนการวนซ้ำสูงสุด และ SLSQP มีข้อได้เปรียบในการจัดการทั้งข้อจำกัดด้านความเท่าเทียมและความไม่เท่าเทียมกัน ทำให้เหมาะสมสำหรับปัญหาค่าเหมาะที่สุดที่หลากหลาย (Chen et al., 2021; Gatta et al., 2023) ขั้นตอนในการสร้าง SLSQP และมีรายละเอียดดังนี้

1. กำหนดค่าเริ่มต้นสำหรับตัวแปรของอัลกอริทึมแบบสุ่ม
2. สร้างการประมาณกำลังสอง (Quadratic Approximation) ของฟังก์ชันวัตถุประสงค์ (Objective Function) และข้อจำกัดรอบ ๆ จุดปัจจุบัน
3. แก้ปัญหาย่อยของโปรแกรมกำลังสองเพื่อรับการวนซ้ำครั้งต่อไป
4. ตรวจสอบการบรรจบกัน หากเป็นไปตามเกณฑ์การบรรจบกัน ให้หยุด มิฉะนั้น ให้ไปที่ขั้นตอนที่ 5
5. อัปเดตการวนซ้ำและกลับไปขั้นตอนที่ 2

**ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm - GA)** เป็นอัลกอริทึมการเพิ่มประสิทธิภาพแบบเมตาฮีวิริสติกที่ได้รับแรงบันดาลใจจากระบบการคัดเลือกโดยธรรมชาติและพันธุกรรม สามารถใช้สำหรับการเพิ่มประสิทธิภาพพอร์ตโฟลิโอเพื่อค้นหาการจัดสรรสินทรัพย์ที่เหมาะสมที่สุดในพอร์ตโฟลิโอที่เพิ่มผลตอบแทนสูงสุดหรือลดความเสี่ยงให้น้อยที่สุด หรือบรรลุการแลกเปลี่ยนระหว่างทั้งสอง เป็นภาพรวมทั่วไปของวิธีการใช้ Genetic Algorithm เพื่อเพิ่มประสิทธิภาพพอร์ตการลงทุน (Portfolio Optimization) (Gunjan & Bhattacharyya, 2022) มีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. กำหนดโครโมโซม (Chromosome): โครโมโซมแต่ละตัวแสดงถึงวิธีแก้ปัญหาที่เป็นไปได้หรือการจัดสรรพอร์ตโฟลิโอ โดยทั่วไปจะถูกเข้ารหัสเป็นสตริงไบนารี โดยแต่ละยีนจะเป็นตัวแทนว่ามีสินทรัพย์รวมอยู่ด้วยหรือไม่
2. ประชากรเริ่มต้น (Population): สร้างประชากรเริ่มต้นของโครโมโซมแบบสุ่มหรือใช้กลยุทธ์ที่กำหนดไว้ล่วงหน้า ขนาดของประชากรสามารถกำหนดได้จากความซับซ้อนของปัญหา
3. ประเมินสมรรถภาพ (Evaluate Fitness): คำนวณความเหมาะสมของโครโมโซมแต่ละตัวในประชากร และ ค่าความเหมาะสม (Fitness Value) แสดงถึงประสิทธิภาพของการจัดสรรพอร์ตโฟลิโอตามเกณฑ์ที่กำหนดไว้ล่วงหน้า เช่น ความเสี่ยง ผลตอบแทน หรือทั้งสองอย่างรวมกัน สำหรับการเพิ่มประสิทธิภาพพอร์ตโฟลิโอ ค่าความเหมาะสมที่ซักรันทั่วไป ได้แก่ การเพิ่มประสิทธิภาพค่าเฉลี่ยความแปรปรวน อัตราส่วน Sharpe หรือผลตอบแทนที่ปรับตามความเสี่ยง
4. การคัดเลือก (Selection): เลือกโครโมโซม (Chromosome) จากประชากรปัจจุบันเพื่อสร้างรุ่นต่อไป โดยทั่วไปกระบวนการคัดเลือกจะขึ้นอยู่กับค่าความเหมาะสม โดยบุคคลที่มีรูปร่างดีกว่าจะมีโอกาสได้รับเลือกสูงกว่า วิธีการเลือกทั่วไป ได้แก่ การเลือกวงล้อสุ่ม การเลือกการแข่งขัน หรือการเลือกตามอันดับ
5. ผู้ดำเนินการทางพันธุกรรม (Genetic Operators):
  - ก. ครอสโอเวอร์ (Crossover): ทำการครอสโอเวอร์หรือรวมกันใหม่ระหว่างโครโมโซมที่เลือกเพื่อสร้างลูกหลาน ครอสโอเวอร์เกี่ยวข้องกับ การแลกเปลี่ยนข้อมูลทางพันธุกรรมระหว่างโครโมโซมแม่สองตัวเพื่อสร้างโซลูชันใหม่ สามารถใช้เทคนิคการครอสโอเวอร์แบบต่าง ๆ เช่น ครอสโอเวอร์หนึ่งจุด ครอสโอเวอร์สองจุด หรือยูนิฟอร์มครอสโอเวอร์
  - ข. การกลายพันธุ์ (Mutation): แนะนำการเปลี่ยนแปลงแบบสุ่มหรือการกลายพันธุ์ในโครโมโซมลูกหลานเพื่อรักษาความหลากหลายและสำรวจพื้นที่ใหม่ของพื้นที่การแก้ปัญหา การกลายพันธุ์ช่วยป้องกันไม่ให้อัลกอริทึมติดอยู่ใน optima ในเครื่อง การกลายพันธุ์อาจเกี่ยวข้องกับการพลิกหรือเปลี่ยนยีนเฉพาะในโครโมโซม
6. การแทนที่ (Replacement): การแทนที่ส่วนหนึ่งของประชากรที่มีอยู่ด้วยลูกหลานที่เกิดจากตัวดำเนินการทางพันธุกรรม กลยุทธ์การเปลี่ยนอาจขึ้นอยู่กับความได้เปรียบ (การรักษาวิธีแก้ปัญหาที่ดีที่สุด) การทดแทนรุ่นต่อรุ่น หรือทั้งสองอย่างรวมกัน
7. ทำซ้ำขั้นตอนที่ 3 ถึง 6: ทำซ้ำขั้นตอนตามจำนวนรุ่นที่กำหนดหรือจนกว่าจะถึงเกณฑ์การยกเลิก อัลกอริทึมยังคงพัฒนาประชากรต่อไป โดยแต่ละรุ่นจะปรับปรุงสมรรถภาพโดยรวมของประชากร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8. การสิ้นสุด (Termination): หยุดอัลกอริทึมเมื่อผ่านเกณฑ์การยุติ เช่น ถึงจำนวนรุ่นสูงสุด บรรลุระดับความเหมาะสมที่ต้องการ หรือเมื่ออัลกอริทึมรวมเข้าด้วยกัน

ด้วยการใช้ตัวดำเนินการทางพันธุกรรมและกลไกการเลือกซ้ำ ๆ Genetic Algorithm จะสำรวจพื้นที่ของโซลูชันและแนะนำการเพิ่มประสิทธิภาพเพื่อค้นหาการจัดสรรพอร์ตโฟลิโอที่ดีที่สุดซึ่งตรงตามวัตถุประสงค์ที่กำหนดไว้ และสิ่งสำคัญคือต้องทราบว่า การนำ Genetic Algorithm ไปใช้เฉพาะสำหรับการเพิ่มประสิทธิภาพพอร์ตโฟลิโออาจแตกต่างกันไปตามข้อกำหนดของปัญหาและวัตถุประสงค์ของการเพิ่มประสิทธิภาพ สามารถใช้ฟังก์ชันความเหมาะสมที่แตกต่างกัน วิธีการเลือกเทคนิคการครอสโอเวอร์ และเกณฑ์การยุติ เพื่อปรับแต่งอัลกอริทึมให้ตรงกับความต้องการเฉพาะ

**อาณาจักรมด (Ant Colony Optimization, ACO)** วิธีการนี้เป็นวิธีการที่จำลองวิธีการทำงานของมดงานในการเดินทางจากรังไปสู่แหล่งอาหาร ปกติแล้วเมื่อมดเดินทางมดจะปล่อยกลิ่นฟีโรโมนทิ้งไว้รายทางเพื่อนำทางให้มดงานตัวถัดไป จากการทดลองเพื่อศึกษาการเดินทางของมดด้วยการสร้างสะพานที่มี 2 ทางแยก โดยการทดลองในช่วงแรกจะใช้สะพานที่มีความยาวเท่ากันไม่ว่าจะแยกไปทางใดก็ตาม เมื่อมีมดเดินข้ามผ่านสะพานมากเพียงพอ จะพบว่าทาง 1 เส้นบนสะพานจะมีความเข้มข้นของฟีโรโมนสูงกว่าอีกฝั่งหนึ่ง นั่นเท่ากับเป็นการดึงดูดให้มดงานตัวถัด ๆ ไปเดินผ่านสะพานด้านที่มีความเข้มข้นของฟีโรโมนสูงมากกว่าอีกฝั่ง (Gunjan & Bhattacharyya, 2022) จึงนำมาเพื่อทดสอบการคัดเลือกผลประกอบการของหลักทรัพย์และจัดอันดับหลักทรัพย์ที่นำลงทุน และหลักทรัพย์ที่เหมาะสมในการลงทุน การคำนวณค่าในแบบจำลองอาณาจักรมด ด้วยฟังก์ชันวัตถุประสงค์ มาตราวัดตามตัวแบบชาร์ปเรโซ เพิ่มกำไรสูงสุดให้กับพอร์ตการลงทุนจากผลตอบแทน

ขั้นตอนต่อไปนี้เป็นภาพรวมทั่วไปของวิธีการใช้ Ant Colony Optimization เพื่อเพิ่มประสิทธิภาพพอร์ตโฟลิโอ (Portfolio Optimization):

1. ระบุปัญหา (Define The Problem): กำหนดฟังก์ชันวัตถุประสงค์ (Objective Function) และข้อจำกัด (Constraints) สำหรับปัญหาการเพิ่มประสิทธิภาพพอร์ตโฟลิโอ ซึ่งคล้ายกับอัลกอริทึมการปรับให้เหมาะสมอื่นๆ ฟังก์ชันวัตถุประสงค์แสดงถึงการวัดสมรรถรูปที่ต้องขยายหรือย่อให้เล็กสุด เช่น ผลตอบแทนที่ปรับตามความเสี่ยงหรืออัตราส่วน Sharpe ข้อจำกัดอาจรวมถึงข้อจำกัดเกี่ยวกับน้ำหนักสินทรัพย์ ความเสี่ยงของภาคธุรกิจ หรือข้อจำกัดเฉพาะพอร์ตโฟลิโออื่น ๆ
2. เริ่มต้นฝูงมด (Ant Colony): สร้างอาณาจักรมดเริ่มต้น โดยที่มดแต่ละตัวแสดงถึงการจัดสรรพอร์ตโฟลิโอที่เป็นไปได้ มดแต่ละตัวติดตามทรัพย์สินที่เลือกไว้และน้ำหนักที่สอดคล้องกัน ในขั้นต้นมดแต่ละตัวอาจมีการจัดสรรแบบสุ่มหรือแบบเดียวกัน

3. กำหนดฟีโรโมนเมทริกซ์ (Pheromone Matrix): สร้างเมทริกซ์ฟีโรโมนที่แสดงถึงความน่าดึงดูดใจหรือความปรารถนาในการเลือกสินทรัพย์เฉพาะสำหรับพอร์ตโฟลิโอ เมทริกซ์ฟีโรโมนนำทางมดในกระบวนการตัดสินใจ
  4. การตัดสินใจของมด (Ant Decision-Making): มดแต่ละตัวจะเลือกสินทรัพย์ซ้ำ ๆ สำหรับพอร์ตโฟลิโอตามกฎการตัดสินใจที่น่าจะเป็น กฎการตัดสินใจจะพิจารณาทั้งความน่าดึงดูดใจของสินทรัพย์ตามเมทริกซ์ฟีโรโมน อาจรวมถึงประสิทธิภาพในอดีตหรือลักษณะความเสี่ยงของสินทรัพย์
  5. ประเมินความเหมาะสม (Evaluate Fitness): หลังจากที่มีมดแต่ละตัวสร้างการจัดสรรพอร์ตโฟลิโอแล้ว ให้ประเมินความเหมาะสมของพอร์ตโฟลิโอโดยการคำนวณค่าฟังก์ชันวัตถุประสงค์ ฟิตเนสแสดงให้เห็นว่าการจัดสรรพอร์ตโฟลิโอทำงานได้ดีเพียงใดตามวัตถุประสงค์ที่กำหนดไว้
  6. อัปเดตเมทริกซ์ฟีโรโมน (Update Pheromone Matrix): หลังจากประเมินความเหมาะสมของมดทั้งหมดแล้ว ให้อัปเดตเมทริกซ์ฟีโรโมนเพื่อสะท้อนถึงคุณภาพของการจัดสรรพอร์ตโฟลิโอ มดที่สร้างพอร์ตโฟลิโอที่ดีกว่าจะปล่อยฟีโรโมนในระดับที่สูงกว่าในสินทรัพย์ที่เกี่ยวข้อง ซึ่งมีอิทธิพลต่อการตัดสินใจของมดตัวต่อ ๆ ไป
  7. ทำซ้ำขั้นตอนที่ 4 ถึง 6: ทำซ้ำขั้นตอนการตัดสินใจของมด การประเมินพอร์ตโฟลิโอ และการอัปเดตเมทริกซ์ฟีโรโมน การทำซ้ำแต่ละครั้งช่วยให้มดสามารถสำรวจและใช้ประโยชน์จากชุดค่าผสมของสินทรัพย์ต่าง ๆ โดยค่อย ๆ ผสานเข้ากับการจัดสรรพอร์ตโฟลิโอที่ดีขึ้น
  8. การสิ้นสุด (Termination): หยุดอัลกอริทึมเมื่อเป็นไปตามเกณฑ์การสิ้นสุด เช่น การวนซ้ำถึงจำนวนสูงสุดหรือการบรรลุระดับความเหมาะสมที่ต้องการ
- ใน ACO เมทริกซ์ฟีโรโมนมีบทบาทสำคัญในการขึ้นำกระบวนการตัดสินใจของมด กลไกการตอบรับเชิงบวกของการเสริมฟีโรโมนช่วยให้มดรวมตัวกันเพื่อจัดสรรพอร์ตโฟลิโอที่ดีขึ้น สิ่งสำคัญคือต้องสังเกตว่าการใช้ ACO เฉพาะสำหรับการเพิ่มประสิทธิภาพพอร์ตโฟลิโออาจแตกต่างกันไปตามข้อกำหนดของปัญหาและวัตถุประสงค์ของการเพิ่มประสิทธิภาพ การออกแบบฟีโรโมนเมทริกซ์ กฎการตัดสินใจที่น่าจะเป็น ฟีโรโมนเมทริกซ์ (Pheromone Matrix) และพารามิเตอร์อื่น ๆ สามารถปรับแต่งให้เหมาะกับปัญหาการเพิ่มประสิทธิภาพพอร์ตโฟลิโอเฉพาะได้

วิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization - PSO) เป็นอัลกอริทึมการปรับให้เหมาะสมแบบเมตาฮิวริสติกที่ได้รับแรงบันดาลใจจากพฤติกรรมของฝูงนกหรือฝูงปลา นอกจากนี้ยังสามารถนำไปใช้กับการเพิ่มประสิทธิภาพพอร์ตโฟลิโอเพื่อค้นหาการจัดสรรสินทรัพย์ที่เหมาะสมที่สุดเพื่อเพิ่มผลตอบแทนสูงสุดหรือลดความเสี่ยง (Chen et al., 2021; Gunjan & Bhattacharyya, 2022)

ขั้นตอนต่อไปนี้เป็นภาพรวมทั่วไปของวิธีการใช้ Particle Swarm Optimization เพื่อเพิ่มประสิทธิภาพพอร์ตโฟลิโอ (Portfolio Optimization):

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. กำหนดปัญหา: กำหนดหน้าที่วัตถุประสงค์และข้อจำกัดสำหรับปัญหาการเพิ่มประสิทธิภาพพอร์ตโฟลิโอ ฟังก์ชันวัตถุประสงค์ (Objective Function) แสดงถึงการวัดสมรรถรูปที่ต้องขยายหรือย่อให้เล็กสุด เช่น ผลตอบแทนที่ปรับตามความเสี่ยงหรืออัตราส่วน Sharpe ข้อจำกัดอาจรวมถึงข้อจำกัดเกี่ยวกับน้ำหนักสินทรัพย์ ความเสี่ยงของภาคธุรกิจ หรือข้อจำกัดเฉพาะพอร์ตโฟลิโออื่น ๆ
2. Initialize the Swarm: สร้างฝูงเริ่มต้นของอนุภาค โดยที่แต่ละอนุภาคแสดงถึงการจัดสรรพอร์ตโฟลิโอที่เป็นไปได้ แต่ละอนุภาคมีลักษณะเฉพาะตามตำแหน่งและความเร็วในพื้นที่ โดยตำแหน่งของอนุภาคสอดคล้องกับน้ำหนักสินทรัพย์ในพอร์ตโฟลิโอ
3. ประเมินความเหมาะสม (Evaluate Fitness): การประเมินความเหมาะสมของแต่ละอนุภาค โดยการคำนวณค่าฟังก์ชันวัตถุประสงค์ตามตำแหน่ง การประเมินความเหมาะสมแสดงให้เห็นว่าการจัดสรรพอร์ตโฟลิโอทำงานได้ดีเพียงใดตามวัตถุประสงค์ที่กำหนดไว้
4. อัปเดตความเร็วของอนุภาค (Particle Velocity) และตำแหน่งของอนุภาค (Particle Position): การอัปเดตความเร็วและตำแหน่งของแต่ละอนุภาคตามตำแหน่งที่มีค่าความเหมาะสมที่ดีที่สุด (best-known positions (pbest)) และตำแหน่งที่มีค่าความเหมาะสมที่ดีที่สุดของฝูงทั้งหมด (global best-known position (gbest)) ความเร็วกำหนดการเคลื่อนที่ของอนุภาคในพื้นที่การแก้ปัญหา ในขณะที่ตำแหน่งแสดงถึงน้ำหนักสินทรัพย์ในพอร์ตโฟลิโอ
5. ทำซ้ำขั้นตอนที่ 3 และ 4: ทำซ้ำขั้นตอนการประเมินความเหมาะสมและอัปเดตตำแหน่งและความเร็วของอนุภาค การทำซ้ำแต่ละครั้งมีจุดมุ่งหมายเพื่อปรับปรุงความเหมาะสมของอนุภาค และนำฝูงไปสู่แนวทางแก้ไขที่ดีขึ้น
6. การสิ้นสุด (Termination): หยุดอัลกอริทึมเมื่อเป็นไปตามเกณฑ์การสิ้นสุด เช่น การวนซ้ำถึงจำนวนสูงสุดหรือการบรรลุระดับความเหมาะสมที่ต้องการ  
สิ่งสำคัญคือต้องสังเกตว่าการใช้งาน PSO เฉพาะสำหรับการเพิ่มประสิทธิภาพพอร์ตโฟลิโออาจแตกต่างกันไปตามข้อกำหนดของปัญหาและวัตถุประสงค์ของการเพิ่มประสิทธิภาพ ฟังก์ชันความเหมาะสม จำนวนอนุภาค สมการอัปเดตสำหรับความเร็วและตำแหน่ง และพารามิเตอร์อื่น ๆ สามารถปรับแต่งให้เหมาะกับปัญหาการเพิ่มประสิทธิภาพพอร์ตโฟลิโอเฉพาะได้

### 2.3 การประเมินผล (Evaluation)

การประเมินผลเป็นเครื่องมือสำคัญในการวัดประสิทธิผลของการวิจัยนี้ เครื่องมือที่ใช้ในการประเมินผลต้องสอดคล้องกับการวัดผลการปฏิบัติงานในการศึกษา ได้ดังนี้

#### Receiver Operating Characteristic curves (ROC curves)

ROC curves ใช้เพื่อประเมินประสิทธิภาพของอัลกอริทึมการจำแนกประเภทไบนารี แสดงประสิทธิภาพของการจำแนกประเภท (Classification) แทนที่จะเป็นค่าเดียวเหมือนเมตริกอื่น ๆ ส่วน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ใหญ่ (Ananthakumar & Sarkar, 2017; Basak et al., 2019; Bustos & Pomares-Quimbaya, 2020; Geurts et al., 2006; Huang et al., 2019) ตัวแปรตาม (Y) ในกรณีนี้คือตัวแปรเชิงคุณภาพ ซึ่งแบ่งได้เป็น 2 กรณี คือ  $Y = 1$  เมื่อเหตุการณ์ที่น่าสนใจหรือผลการทดสอบเป็นบวก และ  $Y = 0$  เมื่อเหตุการณ์ถูกละเว้น หรือผลการทดสอบเป็นเชิงลบ จุดตัด (cut-off point) คือจุดที่ใช้จำแนกเหตุการณ์ ออกเป็นเหตุการณ์ที่น่าสนใจและไม่น่าสนใจ การเปรียบเทียบระหว่างค่าพยากรณ์และการสังเกตแบ่งได้ เป็น 4 กรณี ดังตารางที่ 2.2

ตารางที่ 2.2 ตารางแสดงผลการทำนาย (Confusion Matrix)

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positive (TP)	False Positive (FP)
Predicted Negative (0)	False Negative (FN)	True Negative (TN)

TP, TN, FP และ FN ในตารางแสดงด้วยค่าความถี่:

- True Positive (TP) คือการตอบสนองเมื่อสิ่งที่คาดการณ์ตรงกับสิ่งที่เกิดขึ้นจริง ในกรณีคำทำนายที่ “จริง” สิ่งที่เกิดขึ้นนั้น “จริง”
- False Positive (TN) คือเมื่อการคาดคะเนตรงกับสิ่งที่เกิดขึ้น ในกรณีที่คำทำนาย “ไม่จริง” สิ่งที่เกิดขึ้นนั้น “ไม่จริง”
- False Positive (FP) คือการคาดคะเนที่ไม่ตรงกับสิ่งที่เกิดขึ้น กล่าวคือ การคาดคะเนนั้น “จริง” แต่สิ่งที่เกิดขึ้นนั้น “ไม่จริง”
- False Negative (FN) คือการคาดคะเนที่ไม่ตรงกับสิ่งที่เกิดขึ้นจริง นั่นคือการคาดคะเนว่า บางสิ่ง “ไม่จริง” แต่สิ่งที่เกิดขึ้นนั้น “จริง”

True Positive Rate (TPR) คือการวัดสิ่งที่คาดการณ์และตรงกับสิ่งที่เกิดขึ้นจริงกับ เหตุการณ์จริงทั้งหมด (True positive (TP) + False Negative (FN)) ดังแสดงในสมการที่ 2.12

$$TPR = \frac{TP}{TP+FN} \quad (2.12)$$

False Positive Rate (FPR) เป็นตัวชี้วัดของสิ่งที่คาดการณ์ไม่ตรงกับสิ่งที่เกิดขึ้นกับ เหตุการณ์ที่เป็นเท็จทั้งหมด (True Negative (TN) + False Positive (FP)) ดังแสดงในสมการที่ 2.13

$$FPR = \frac{FP}{TN+FP} \quad (2.13)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Precision คือค่าเฉลี่ยถ่วงน้ำหนักของค่าบวกและค่าลบของความแม่นยำ ในขณะที่ Recall คือค่าเฉลี่ยถ่วงน้ำหนักของค่าบวกการเรียกคืน ดังแสดงในสมการ 2.14 – 2.17

$$\text{Precision}_{\text{positive}} = \frac{TP}{TP+FP} \quad (2.14)$$

$$\text{Precision}_{\text{negative}} = \frac{TN}{TN+FN} \quad (2.15)$$

$$\text{Recall}_{\text{positive}} = \frac{TP}{TP+FN} \quad (2.16)$$

$$\text{Recall}_{\text{negative}} = \frac{TN}{TN+FP} \quad (2.17)$$

ความแม่นยำ (Accuracy) และ F1-score ดังแสดงในสมการ 2.18 – 2.19

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.18)$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.19)$$

ROC curves ถูกพล็อตระหว่าง TPR (Sensitivity) และ FPR (1-Specificity) สำหรับค่าจุดตัดที่เป็นไปได้ทั้งหมด แกน y แทน TPR ในขณะที่แกน x แทน FPR ที่ค่าจุดตัดแต่ละค่า จะแยกผลการพยากรณ์ออกเป็นสองกลุ่ม: เหตุการณ์ที่  $P(Y = 1)$  มากกว่าหรือเท่ากับจุดตัด และเหตุการณ์ที่  $P(Y = 1)$  น้อยกว่า จุดตัดขาดทุน โดยที่  $P(Y = 1)$  คือความน่าจะเป็นที่เหตุการณ์ที่น่าสนใจหรือผลการทดสอบเป็นบวก พื้นที่ใต้เส้นโค้ง ROC คือดัชนีที่ใช้ระบุความถูกต้องหรือความน่าเชื่อถือของแบบจำลอง ดังนั้น แบบจำลองที่มีพื้นที่มากที่สุดได้ ROC curves จึงเป็นแบบจำลองที่แม่นยำที่สุด (Elizabeth et al., 1988)

## 2.4 งานวิจัยที่เกี่ยวข้อง (Research Methodology)

Teixeira et al (2010) ได้วิจัยการเลือกหลักทรัพย์อัตโนมัติ โดยข้อมูลในการวิจัยนี้พิจารณาจากประวัติของราคาปิดของหลักทรัพย์รายวันและปริมาตรการซื้อขายของตลาดหลักทรัพย์บราซิล (Bovespa) ในช่วง 4 มกราคม 1998 ถึง 3 กันยายน 2009 โดยนำตัวชี้วัดทางเทคนิค (Technical Indicators) ดังต่อไปนี้ในการสร้างแบบจำลอง ค่าเฉลี่ยเคลื่อนที่แบบง่าย (Simple Moving Average: SMA) ดัชนีความแข็งแรงสัมพัทธ์ (Relative Strength Index: RSI) ออสซิลเลเตอร์สโตแคสติก เค (Stochastic Oscillator K: K) ออสซิลเลเตอร์สโตแคสติก ดี (Stochastic Oscillator D: D) และ โบลลิงเจอร์แบนด์ (Bollinger Bands) มาใช้สร้างแบบจำลองการพยากรณ์ราคาหลักทรัพย์ด้วยขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้สุด k ตัว (K-Nearest Neighbor Algorithm: KNN) ผลวิจัยพบว่าวิธีการนี้สามารถบรรลุผลลัพธ์ที่ดีในแง่ของผลกำไรสำหรับหลักทรัพย์ 12 จาก 15 ตัวที่พิจารณาใน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทดลอง โดยใช้การเปรียบเทียบผลลัพธ์กับผลกำไรที่จะได้รับด้วยกลยุทธ์ Buy-and-hold ดังนั้นจึงเชื่อว่าการใช้แนวทางนี้ในการคาดการณ์แนวโน้มหลักทรัพย์ระยะสั้น นั้นสามารถเป็นไปได้

Li et al (2014) ได้วิจัยผลกระทบของข่าวและอารมณ์ต่อการเคลื่อนไหวของดัชนีหลักทรัพย์จีน (CSI 100) โดยข้อมูลในการวิจัยนี้คือ เว็บไซต์ (Web News) และบล็อก (Blog) หรือ กระดานสนทนาทางการเงิน (Financial Discussion Board) ในช่วง 1 มกราคม ถึง 31 ธันวาคม 2011 และสร้างแบบจำลองการพยากรณ์หลักทรัพย์ด้วย การถดถอยซัพพอร์ตเวกเตอร์ (Support Vector Regression) ผลวิจัยว่า (1) ข้อมูลพื้นฐานของบทความ เช่น ข่าวเฉพาะของบริษัทสามารถเพิ่มพูนความรู้ของนักลงทุน และส่งผลกระทบต่อกิจกรรมการซื้อขายของนักลงทุน (2) ความรู้สึกสาธารณะทำให้เกิดความผันผวนทางอารมณ์ของนักลงทุนและเข้าไปแทรกแซงในการตัดสินใจของพวกนักลงทุน และ (3) ผลกระทบของสื่อต่อบริษัทแตกต่างกันไปตามลักษณะของบริษัทและเนื้อหาของบทความ

Dash et al (2016) ได้วิจัยการซื้อขายหลักทรัพย์แบบไฮบริดที่รวมตัวชี้วัดทางเทคนิค (Technical Indicators) เข้ากับ เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) โดยข้อมูลในการวิจัยนี้คือ ราคาปิดรายวันและปริมาตรการซื้อขายของดัชนี S&P 500 และ BSE SENSEX ในช่วง 4 มกราคม 2010 ถึง 31 ธันวาคม 2014 ตัวชี้วัดทางเทคนิค (Technical Indicators) ที่นำมาใช้ในการสร้างปัจจัยมีดังต่อไปนี้ ค่าเฉลี่ยเคลื่อนที่แบบง่าย (Simple Moving Average: SMA) ค่าเฉลี่ยเคลื่อนที่แบบคอนเวอร์เจนซ์ไดเวอร์เจนซ์ (Moving Average Convergence Divergence: MACD) ออสซิลเลเตอร์สโตแคสติก เค (Stochastic Oscillator K: K) ออสซิลเลเตอร์สโตแคสติก ดี (Stochastic Oscillator D: D) ดัชนีความแข็งแรงสัมพัทธ์ (Relative Strength Index: RSI) และ ออสซิลเลเตอร์สโตแคสติก อี (Stochastic Oscillator R: R or Larry Williams R%) และสร้างแบบจำลองการพยากรณ์หลักทรัพย์ด้วย โครงข่ายประสาทเทียม (Artificial Neural Network: ANN) ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) แบบจำลองเบย์อย่างง่าย (Naive Bayesian) ขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้สุด k ตัว (K-Nearest Neighbor Algorithm: KNN) และ แบบจำลองต้นไม้ตัดสินใจ (Decision Tree: DT) เพื่อเปรียบเทียบประสิทธิภาพ ซึ่งผลปรากฏว่าโครงข่ายประสาทเทียมให้เปอร์เซ็นต์ผลตอบแทนที่ดีกว่าแบบจำลองอื่น ๆ เฉลี่ย 34.42 % ของดัชนี BSE SENSEX และ 45.58% ของดัชนี S&P 500

Mitra Thakur et al (2018) ได้วิจัยการเลือกหลักทรัพย์ในการจัดพอร์ตลงทุน โดยแบ่งออกเป็น 2 ส่วนด้วยกัน คือ การเลือกหลักทรัพย์ และการจัดการสัดส่วนในการลงทุน โดยการเลือกหลักทรัพย์ใช้ทฤษฎีหลักฐาน Dempster และ Shafer (Dempster-Shafer Evidence Theory) โดยพิจารณาจากปัจจัย 10 ปัจจัยต่อไปนี้ กำไรต่อหลักทรัพย์ (Earn Per Share ; EPS) อัตราส่วนทางการเงิน (Price to Earning Ratio ; P/E) อัตราการจ่ายเงินปันผล (Payout Ratio ; PR) อัตราส่วน ราคาต่อยอดขาย (Price to Sales Ratio ; P/S) อัตราส่วนหนี้สินระยะยาว (Long Term Debt to Equity Ratio ; LTDER) ราคาหลักทรัพย์ต่อมูลค่าทางบัญชี (Price to Book Value ; P/B)

อัตราส่วนเงินทุนหมุนเวียน (Current Ratio ; CR) อัตราส่วนราคาต่อกระแสเงินสดต่อหลักทรัพย์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Price to Cash Flow Ratio ; P/CF) อัตรากำไร (Profit Margin ; PM) และ อัตราการหมุนเวียนของลูกหนี้ (Accounts Receivable Turnover ; ART) และในส่วนของจัดการสัดส่วนในการลงทุน ใช้ขั้นตอนวิธีอาณานิคมด (Ant Colony Optimization; ACO) สำหรับการเพิ่มประสิทธิภาพสัดส่วนในการจัดพอร์ตลงทุน ดังนั้นงานวิจัยนี้เสนอวิธีเลือกหลักทรัพย์ และหาสัดส่วนในการจัดพอร์ตลงทุน เพื่อใช้เป็นเครื่องมือในการตัดสินใจที่ดีขึ้นสำหรับการลงทุน

Atkins et al (2018) ได้วิจัยการคาดการณ์ความผันผวนและราคาปิดของตลาดหลักทรัพย์ โดยข้อมูลในงานวิจัยนี้สามารถแยกออกเป็น 2 ส่วน คือ ข่าว (News data) จากสำนักข่าวรอยเตอร์ สหรัฐอเมริกาช่วงเดือนกันยายน 2011 ถึง เดือนกันยายน 2012 โดยจัดการข้อมูลด้วยแบบจำลอง Latent Dirichlet Allocation และ ข้อมูลทางการเงิน (Financial data) โดยใช้ข้อมูลรายวันตลาดหลักทรัพย์สหรัฐอเมริกาจากฐานข้อมูล Yahoo Finance ระหว่างวันที่ 9 กันยายน 2011 ถึง วันที่ 7 กันยายน 2012 และในการเลือกปัจจัยสำหรับการพยากรณ์ด้วยไคสแควร์ (Chi-squared) และสร้างแบบจำลองการพยากรณ์ความผันผวนและราคาปิดของตลาดหลักทรัพย์ด้วยแบบจำลองเบย์อย่างง่าย (Simple Naive Bayes) ซึ่งผลปรากฏว่า ข้อมูลข่าวส่งผลต่อคาดการณ์ความผันผวนนั้นดีกว่าการคาดการณ์ราคาปิดของตลาดหลักทรัพย์ ดังนั้นงานวิจัยนี้สามารถคาดการณ์ความผันผวนของตลาดหลักทรัพย์ เพื่อประกอบการตัดสินใจในการลงทุน

Huang et al (2019) ได้วิจัยการคาดการณ์การเคลื่อนไหวของ S&P 500 index โดยใช้ Google Trend Search โดยข้อมูลในงานวิจัยนี้สามารถแยกออกเป็น 2 ส่วน คือ จำนวนการค้นหาใน Google Trend โดยมีคำค้นหา 98 คำ และ ข้อมูล การเคลื่อนไหวของ S&P 500 index โดยตัวงานวิจัยนี้ได้แบ่งเป็น 2 ส่วนใหญ่ ๆ คือ การเลือกปัจจัยที่ส่งผล ด้วยแบบจำลอง Vector Autoregressive (VAR) และ Mallows  $C_p$  และ การเลือกแบบจำลองในการคาดการณ์การเคลื่อนไหวของ S&P 500 index ด้วยแบบจำลองซัพพอร์ตเวกเตอร์แมทซิน (Support Vector Machine : SVM) การถดถอยริดจ์ (Ridge Regression) แลสโซ (Lasso) และการถดถอยยืดหยุ่น (Elastic Net Regression) เพื่อเปรียบเทียบประสิทธิภาพ โดยแบบจำลองที่มีประสิทธิภาพดีที่สุดคือการถดถอยริดจ์กับการเลือกปัจจัยที่ส่งผลคือ VAR มีความแม่นยำ 63.75%

Alberto et al (2020) ได้วิจัยการคาดการณ์การเคลื่อนไหวของดัชนี NASDAQ ด้วยการเรียนรู้ของเครื่อง (Machine Learning) ผ่านตัวชี้วัดทางเทคนิคอย่างค่าเฉลี่ยเคลื่อนที่แบบคอนเวอร์เจนซ์ไดเวอร์เจนซ์ (Moving Average Convergence Divergence: MACD) โดยข้อมูลในการวิจัยนี้คือ ราคาปิดรายวันและปริมาณของดัชนี NASDAQ ในช่วง 9 มกราคม 2014 ถึง 31 ธันวาคม 2019 โดยได้มีการหาค่าพารามิเตอร์ที่เหมาะสมของค่าเฉลี่ยเคลื่อนที่แบบคอนเวอร์เจนซ์ไดเวอร์เจนซ์จากขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithms: GA) ซึ่งผลปรากฏว่าสามารถหาค่าที่เหมาะสมที่สุดของพารามิเตอร์ของตัวชี้วัดทางเทคนิค ซึ่งส่งผลให้ได้รับผลตอบแทนจากการลงทุนที่สูงขึ้นผ่านขั้นตอนวิธีทางพันธุกรรม และประสิทธิภาพดีกว่า Buy-and-hold ประมาณ 4%

Bustos et al (2020) ได้ศึกษาค้นคว้ากระบวนการ การคาดการณ์ของตลาดหลักทรัพย์ ด้วยการใช้การเรียนรู้ด้วยเครื่องจักร (Machine Learning) การเรียนรู้เชิงลึก (Deep Learning) การทำเหมืองข้อความ (Text Mining) และวิธีการรวมกลุ่ม (Ensemble Techniques) ในการรู้เรียนข้อมูลที่มีโครงสร้าง (Structured Data) และข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) โดยข้อมูลที่มีโครงสร้างนั้นประกอบไปด้วย ข้อมูลข่าวสารทางการตลาด (Market Information) ดัชนีชี้วัดทางเทคนิค (Technical Indicators) และดัชนีเศรษฐกิจ (Economic Indicators) และข้อมูลที่ไม่มีโครงสร้างนั้นประกอบไปด้วย ข่าว (News) เครือข่ายสังคม (Social Network) และบล็อก (Blogs) ดังนั้นการศึกษาค้นคว้าดังกล่าวทำให้ทราบถึงกระบวนการ การคาดการณ์การเคลื่อนไหวของตลาดหลักทรัพย์ในอีกรูปแบบหนึ่งที่น่าการเรียนรู้ด้วยเครื่องจักร การเรียนรู้เชิงลึก การทำเหมืองข้อความ และวิธีการรวมกลุ่มนำมาประยุกต์ใช้ เพื่อให้ได้รูปแบบการลงทุนในตลาดหลักทรัพย์ที่มีความแม่นยำมากขึ้น สามารถช่วยให้นักลงทุนมีเครื่องมือในการตัดสินใจที่ดีขึ้น

Gunjan & Bhattacharyya (2022) ได้ศึกษาค้นคว้ากระบวนการเพิ่มประสิทธิภาพพอร์ตโฟลิโอ การเพิ่มประสิทธิภาพพอร์ตโฟลิโอช่วยอำนวยความสะดวกในการเลือกพอร์ตการลงทุนในสถานการณ์ตลาดที่ผันผวน โดยการเปรียบเทียบกับแบบคลาสสิก: Genetic Algorithm (GA), Particle swarm optimization (PSO), Combinatorial Optimization Problems (COPs), Ant colony optimization (ACO) และแบบการเพิ่มประสิทธิภาพเชิงควอนตัม (Quantum computing ,QC) โดยการเพิ่มประสิทธิภาพเชิงควอนตัม นั้นให้ผลดีเมื่อพูดถึงการจัดสรรพอร์ตการลงทุนที่เหมาะสมที่สุด

โดยสามารถเปรียบเทียบลักษณะข้อมูล แบบจำลอง หลักทรัพย์/ดัชนี และช่วงเวลา ของงานวิจัยข้างต้นได้จาก ตารางที่ 2.3

ตารางที่ 2.3 การเปรียบเทียบเชิงประจักษ์ของงานวิจัย

ปี	ข้อมูล	แบบจำลอง	หลักทรัพย์/ ดัชนี	ช่วงเวลา	อ้างอิง
2010	Technical Indicator: SMA, RSI, K, D and BB	KNN	Bovespa	04/01/1998- 03/09/2009	(Teixeira & de Oliveira, 2010)
2014	Web News and Financial Discussion Board	Support Vector Regression (SVR)	CSI100	01/01/2011- 31/11/2011	(Li et al., 2014)
2016	Technical Indicator: SMA, MACD, K, D, RSI and R	ANN, NB, SVM, KNN and DT	BSE SENSEX, S&P500	01/01/2010- 01/12/2014	(Dash & Dash, 2016)
2018	Web News: Financial, General and Company Financial: Market, Technical Indicator and Economic	Naïve Bayes	Dow Jones	09/09/2011- 07/09/2012	(Atkins et al., 2018)
2018	Financial: Market	Dempster- Shafer Evidence Theory	Bombay Stock Exchange (BSE)	01/01/1998- 31/12/2009	(Mitra Thakur et al., 2018)
2019	Google Trends	SVM, Ridge Regression, LASSO and Elastic Net	S&P 500	01/01/2004- 31/12/2017	(Huang et al., 2019)
2020	Technical Indicator: MACD	Genetic Algorithms (GA)	NASDAQ	01/01/2008- 31/12/2019	(Antonio Agudelo Aguirre et al., 2020)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.3 (ต่อ)

ปี	ข้อมูล	แบบจำลอง	หลักทรัพย์/ ดัชนี	ช่วงเวลา	อ้างอิง
2022	Financial: Market	Genetic Algorithm (GA), Particle swarm optimization (PSO), Combinatorial Optimization Problems (COPs), Ant colony optimization (ACO) Quantum computing (QC)	Hang Sang DAX 100 FTSE 100 S & P 100 Nikkei	07/09/2016 27/04/2017	(Gunjan & Bhattacharyya, 2022)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 3

### วิธีการดำเนินงานวิจัยการเลือกหลักทรัพย์

บทนี้จะกล่าวถึงวิธีการดำเนินงานวิจัยในครั้งนี้ ซึ่งมีวัตถุประสงค์เพื่อเลือกหลักทรัพย์ที่ดี และการจัดพอร์ตหลักทรัพย์ที่เหมาะสม โดยเนื้อหาในบทนี้ประกอบด้วยหัวข้อหลัก ๆ คือ แผนผังขั้นตอนการดำเนินงาน (Process Diagram) ข้อมูล (Dataset) การเตรียมข้อมูล (Data Preprocessing) แบบจำลอง (Modelling) และผลลัพธ์และการอภิปราย (Results and Discussion)

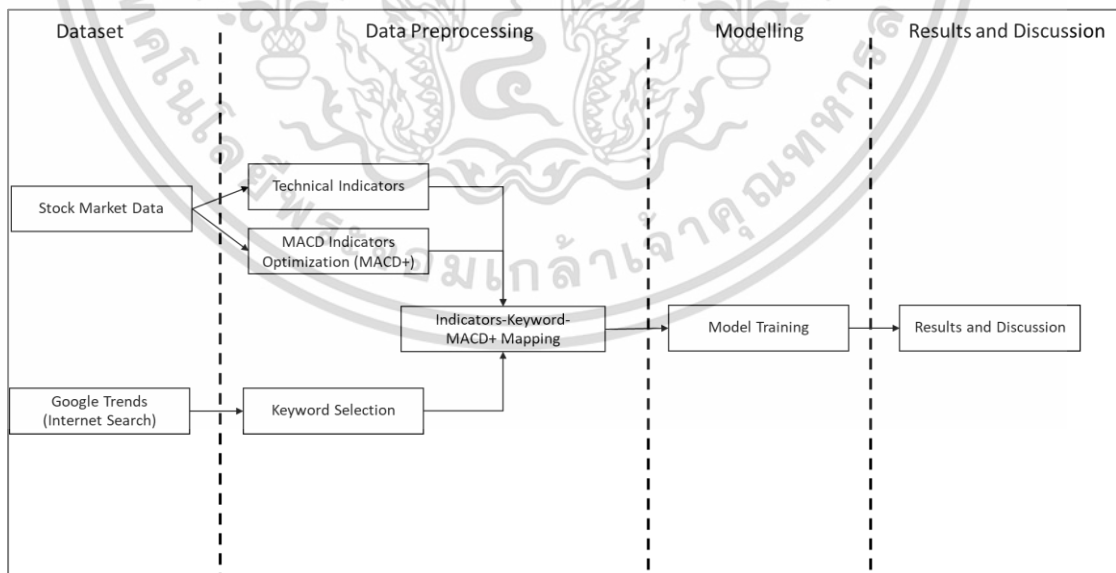
#### 3.1 แผนผังขั้นตอนการดำเนินงาน (Process Diagram)

แผนผังขั้นตอนการดำเนินงานจะแบ่งออกเป็น 2 ส่วน คือ แผนผังขั้นตอนการดำเนินงาน และเครื่องมือหรือโปรแกรมที่ใช้ในการดำเนินงาน

##### 3.1.1 ขั้นตอนการดำเนินงาน

การดำเนินงานวิจัยในครั้งนี้ได้มีกระบวนการในการทำงานวิจัยหลายกระบวนการแสดงในรูปแบบที่ 3.1 โดยมีขั้นตอนในการดำเนินงาน ต่อไปนี้

1. ข้อมูล (Dataset)
2. การเตรียมข้อมูล (Data Preprocessing)
3. แบบจำลอง (Modelling)
4. ผลลัพธ์และการอภิปราย (Results and Discussion)



รูปที่ 3.1 แผนผังขั้นตอนการดำเนินงาน (Process Diagram)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.1.2 เครื่องมือหรือโปรแกรมที่ใช้ในการดำเนินงาน

ด้วยการดำเนินงานในการจัดการข้อมูลนั้นต้องมีเครื่องมือและโปรแกรมในการจัดการโดยทางผู้วิจัยได้ใช้เครื่องมือ และโปรแกรมดังต่อไปนี้

1. เครื่องคอมพิวเตอร์
2. ซอฟต์แวร์ (Software)
  - 2.1 Microsoft Excel 365
  - 2.2 Microsoft Word 365
  - 2.3 Browser Google Chrome
  - 2.4 Jupyter Notebook

## 3.2 ข้อมูล (Dataset)

ข้อมูลที่ใช้ในการศึกษานี้เป็นประเภทข้อมูลที่มีโครงสร้าง (Structured Data) ซึ่งจัดอยู่ในรูปแบบตารางพร้อมคอลัมน์ที่กำหนดได้ รูปแบบนี้ช่วยให้เข้าถึงข้อมูลได้ง่ายและค้นหาได้ง่ายหรือไม่ซับซ้อน แต่ไม่ต้องทำงานเพิ่มเติม ในงานวิจัยนี้รวบรวมข้อมูลจากสองแหล่ง: ตลาดหลักทรัพย์ (Stock Market) และการค้นหาคำสำคัญเฉพาะ (Specific Keyword Searches)

### 3.2.1 ข้อมูลตลาดหลักทรัพย์ (Stock Market Data)

สำหรับข้อมูลตลาดหลักทรัพย์ ข้อมูลนี้เป็นข้อมูลสาธารณะที่สามารถดาวน์โหลดได้จากเว็บไซต์ของตลาดหลักทรัพย์ ซึ่งประกอบด้วย

- ราคาเปิด (Opening Price) เป็นราคาแรกของหลักทรัพย์จดทะเบียนเมื่อเริ่มต้นวันซื้อขาย
- ค่าสูง (High) และต่ำ (Low) แสดงถึงราคาสูงสุดและต่ำสุดของหลักทรัพย์ในวันนั้น ๆ โดยทั่วไปเทรดเดอร์จะใช้สถิติเหล่านี้เพื่อกำหนดความผันผวนของหลักทรัพย์
- ราคาปิด (Closing Price) คือราคาของหลักทรัพย์ ณ สิ้นวันทำการซื้อขาย
- ราคาปิดที่ปรับปรุง (Adjusted Close Price) แล้วถือเป็นราคาจริงของหลักทรัพย์นั้น เนื่องจากสะท้อนถึงมูลค่าของหลักทรัพย์หลังจากจ่ายเงินปันผลแล้ว

ราคาหลักทรัพย์ได้รับอิทธิพลจากหลายปัจจัยและมักถูกพิจารณาว่าเป็นหนึ่งในตัวบ่งชี้ในการศึกษาพฤติกรรมของตลาด ด้วยเหตุนี้ การใช้ราคาหลักทรัพย์เพื่อศึกษาตัวบ่งชี้ทางเทคนิคจึงช่วยเพิ่มประสิทธิภาพในการทำความเข้าใจพฤติกรรมของตลาด ข้อมูลตลาดหลักทรัพย์แห่งประเทศไทย (SET100) รวบรวมทุกสัปดาห์จาก Yahoo Finance ตัวอย่างราคาและปริมาณหลักทรัพย์ย้อนหลังรายวันที่รวบรวมจาก SET100 แสดงในตารางที่ 3.1 สำหรับหลักทรัพย์ “บริษัท ดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน)” (WHA)

ตารางที่ 3.1 ตัวอย่างข้อมูลหลักทรัพย์ของ บริษัท ดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน) (WHA).

Date	High	Low	Open	Close	Volume	Adj Close	Symbol
1/4/2017	3	2.96	2.96	2.98	25413400	2.378223	WHA
1/5/2017	3.04	2.98	3	3.02	82795000	2.410144	WHA
1/6/2017	3.04	3	3.02	3.02	48678800	2.410144	WHA
1/9/2017	3.1	3.02	3.02	3.08	1.52E+08	2.458028	WHA
1/10/2017	3.12	3.04	3.1	3.04	90063300	2.426106	WHA
1/11/2017	3.1	3.06	3.08	3.06	74300900	2.442067	WHA
1/12/2017	3.2	3.1	3.1	3.16	3.55E+08	2.521873	WHA
1/16/2017	3.26	3.18	3.26	3.24	90585100	2.585719	WHA
1/17/2017	3.26	3.18	3.24	3.2	83648500	2.553796	WHA
1/18/2017	3.24	3.18	3.22	3.2	53264600	2.553796	WHA

### 3.2.2 Keywords (Internet Search)

คำว่า "Keywords" ในงานวิจัยนี้ หมายถึง คำศัพท์ที่นักลงทุนใช้ในการติดต่อสื่อสารในภาคการลงทุนด้านหลักทรัพย์ รวมถึง คำศัพท์ที่ผู้ใช้อินเทอร์เน็ตใช้ในการค้นหาข้อมูลหลักทรัพย์ การศึกษานี้ประกอบด้วยคำสำคัญ 484 คำ มีแปดหมวดหมู่ตามการวิจัย (Preis et al., 2013) และแหล่งข้อมูลเช่น Finnomena, SET, Krungsri, Encyclopedia website และ stock2morrow นอกจากนี้ยังสามารถจำแนกเป็นหมวดหมู่ได้ดังนี้ การลงทุนขั้นพื้นฐาน (Basic Investment Terms), กลุ่มอุตสาหกรรม (Industry Groups), ชื่อหลักทรัพย์ (Stock Names), วิธีการซื้อขาย (Trading Methods), การค้นหาทั่วโลก (Global Search), คำยอดนิยม (Popular Words), สำนวน (Idiom) และคำค้นหารายปี (Yearly Search Term) ตัวอย่างและคำจำกัดความของคำสำคัญแสดงไว้ในตารางที่ 3.2

ตารางที่ 3.2 ตัวอย่างคำสำคัญ (Keywords)

Keywords		
Type	Keyword	Definition
Basic Investment Term	P/E	อัตราส่วนราคาต่อกำไร (P/E): อัตราส่วนสำหรับการประเมินมูลค่าบริษัทที่วัดราคาหลักทรัพย์ปัจจุบันเทียบกับกำไรต่อหลักทรัพย์ (EPS)
Basic Investment Term	P/BV	อัตราส่วนราคาต่อมูลค่าตามบัญชี (P/BV): การประเมินมูลค่าตลาดของบริษัทเมื่อเทียบกับมูลค่าตามบัญชี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ตารางที่ 3.2 (ต่อ)

Keywords		
Type	Keyword	Definition
Basic Investment Term	EPS	กำไรต่อหลักทรัพย์ (EPS): จำนวนจากกำไรของบริษัทหารด้วยจำนวนหลักทรัพย์ที่จำหน่ายได้แล้วของหลักทรัพย์สามัญของบริษัท
Industry Group	Agribusiness	ธุรกิจการเกษตร
Industry Group	Food & Beverage	อาหารและเครื่องดื่ม
Industry Group	Insurance	ประกันภัย
Stock Name	ADVANC	บมจ.แอดวานซ์ อินโฟร์ เซอร์วิส (ADVANC.BKK)
Stock Name	BBL	บมจ. ธนาคารกรุงเทพ (BBL.BKK)
Stock Name	CPN	บมจ. เซ็นทรัลพัฒนา (CPN.BKK)
Trading Method	Technical	กลยุทธ์การซื้อขายที่อาศัยตัวบ่งชี้ทางเทคนิคเป็นหลัก
Trading Method	Day Trade	กลยุทธ์การเทรดที่มักได้รับข้อมูลจากการวิเคราะห์ทางเทคนิคของการเคลื่อนไหวของราคา และต้องมีวินัยในตนเองและความเที่ยงธรรมในระดับสูง
Trading Method	Swing Trade	กลยุทธ์การซื้อขายที่มุ่งเน้นไปที่การทำกำไรเล็กน้อยในแนวโน้มระยะสั้นและตัดขาดทุนได้เร็วกว่า
Global Search	Economics	เศรษฐศาสตร์
Global Search	Politics	การเมือง
Global Search	Conflict	ขัดแย้ง
Popular Word	กอง (Kong)	กองทุนรวม
Popular Word	ป๊อป (Pop)	นายหน้า
Popular Word	หรั่ง (Rang)	นักลงทุนต่างชาติ
Idiom	ลำไย (Lamyai)	กำไร
Idiom	ซื้อควาย (Sue Khwai)	ซื้อหลักทรัพย์ก่อนที่ราคาหลักทรัพย์จะลง
Idiom	ขายหมู (Khai Mu)	ขายหลักทรัพย์ก่อนที่ราคาหลักทรัพย์จะขยับขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### ตารางที่ 3.2 (ต่อ)

Keywords		
Type	Keyword	Definition
Yearly Search Term	คนละครึ่ง (Khon La Khrueng)	การรณรงค์ช่วยเหลือทางการเงิน COVID-19 ของรัฐบาลไทย
Yearly Search Term	โควิด-19 (Covid - 19)	โรคติดเชื้อที่เกิดจากไวรัส SARS-CoV-2
Yearly Search Term	ชิมช้อปใช้ (Chim Chop Chai)	การรณรงค์ช่วยเหลือทางการเงิน COVID-19 ของรัฐบาลไทย

Google Trends เป็นเว็บไซต์ของ Google ที่สร้างขึ้นเพื่อช่วยกำหนดความนิยมของคำสำคัญออนไลน์ เช่น ชื่อสินค้า ชื่อบุคคล หรืออาจเป็นชื่อเว็บไซต์ ผู้ใช้สามารถดูความนิยมของคำสำคัญเหล่านี้ตามสถานที่ ไม่ว่าจะเป็นระดับโลก ระดับชาติ หรือระดับจังหวัด และยังสามารถดูแนวโน้มยอดนิยมรายวัน (Huang et al., 2019; Nishimura & Acoba, 2022; Papadamou et al., 2022; Poutachidou & Papadamou, 2021; Sycinska-Dziarnowska et al., 2022; Tudor, 2022) Google Trends ได้รับการออกแบบมาเพื่อตรวจสอบความนิยมของคำสำคัญเมื่อเวลาผ่านไปเพื่อกำหนดโอกาสในการพัฒนาที่อาจเกิดขึ้น Google Trends สร้างขึ้นสำหรับผู้ใช้ที่พึงพาหรือได้รับประโยชน์จากเทรนด์ ไม่ว่าจะเป็นนักการตลาดหรือเจ้าของร้านค้าออนไลน์ หรือแม้แต่ผู้ที่ต้องการเริ่มต้นบล็อกหรือวิดีโอบล็อกของตนเอง แต่ไม่แน่ใจว่าจะขายอะไรดี หรืออะไรที่ถูกค้นหามากที่สุด คำค้นหาในแต่ละวันหรือเดือน นอกจากนี้ยังช่วยในการพัฒนากลยุทธ์ทางการตลาดและการพัฒนาแผนธุรกิจอย่างค่อยเป็นค่อยไป ในด้านการเงิน Huang et al. (2019) ค้นพบว่าการเคลื่อนไหวในทิศทางต่าง ๆ ของ S&P 500 เปลี่ยนแปลงปริมาณการค้นหาตามแต่ละวลีที่ค้นหา และโดยการขยายความหมายของคำ

Google Trends ปรับข้อมูลการค้นหาให้เป็นมาตรฐาน (Normalizes) เพื่อให้การเปรียบเทียบคำง่ายขึ้น โดย normalizes ตามเวลาและตำแหน่งของคำค้นหา จากนั้น ตัวเลขผลลัพธ์จะถูกปรับขนาดจาก 0 ถึง 100 เปอร์เซ็นต์ของการค้นหาทั้งหมดในหัวข้อ พื้นที่ที่แสดงความสนใจในการค้นหาเดียวกันสำหรับคำ ๆ หนึ่ง (Huang et al., 2019; Nishimura & Acoba, 2022; Papadamou et al., 2022; Poutachidou & Papadamou, 2021; Sycinska-Dziarnowska et al., 2022; Tudor, 2022) ตารางที่ 3.3 แสดงตัวอย่างคำยอดนิยมที่ประชาชนใช้ค้นหบน Google ติดอันดับ Google Trends ระหว่างปี 2017 ถึง 2021 ในประเทศไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 ตัวอย่างคำค้นหา Google Trends สำหรับแต่ละคำในประเทศไทย

Date	Debt	Color	Stocks	Restaurant	Portfolio	Inflation	Housing	Dow Jones
2017-01-01	3	21	17	60	30	17	11	3
2017-01-08	22	31	32	51	29	24	7	11
2017-01-15	9	39	25	43	36	15	9	6
2017-01-22	37	41	42	43	38	24	32	9
2017-01-29	12	29	20	50	41	38	17	8
2017-02-05	34	38	21	51	25	55	30	8

### 3.3 การเตรียมข้อมูล (Data Preprocessing)

การประมวลผลข้อมูลแบ่งออกเป็นสามส่วน ดังนี้ (1) ตัวบ่งชี้ทางเทคนิค (technical indicators) (2) การเพิ่มประสิทธิภาพตัวบ่งชี้ MACD (MACD) และ (3) การเลือกคำสำคัญ (keyword selection) ตัวบ่งชี้ทางเทคนิคในตลาดหลักทรัพย์หมายถึงสูตรทางคณิตศาสตร์ตามการเปลี่ยนแปลงและทิศทางของตลาด เช่น ราคาและปริมาณ สำหรับการเลือกคำสำคัญ ผู้เขียนเลือกศึกษาคำค้นหาที่นักลงทุนในตลาดหลักทรัพย์ใช้ รวมถึงคำที่ผู้ใช้อินเทอร์เน็ตใช้ในการค้นหาข้อมูลที่เกี่ยวข้องกับหลักทรัพย์

#### 3.3.1 ตัวบ่งชี้ทางเทคนิค (Technical Indicators)

ตัวชี้วัดทางเทคนิค (Technical Indicators) มีขั้นตอนในการดำเนินการข้อมูล ดังนี้

1. รวบรวมข้อมูลของตลาดหลักทรัพย์ของประเทศไทย (SET 100) เป็นรายสัปดาห์ จากยาฮูไฟแนนซ์ (Yahoo Finance)
2. คำนวณค่าตัวชี้วัดทางเทคนิค (Technical Indicators) ต่าง ๆ

โดยได้แสดงผลดังตารางที่ 3.4

ตารางที่ 3.4 ตัวอย่างข้อมูลหลักทรัพย์ และค่าตัวชี้วัดทางเทคนิค (Technical Indicators) ของบริษัทดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน) (WHA)

Date	SMA	WMA	EMA	MACD	RSI	%K	%D
2021-01-15	2.999	3.012	3.019	0.025	49.695	47.619	68.158
2021-01-22	3.004	3.005	3.003	0.026	36.991	38.095	60.975
2021-01-29	2.999	2.979	2.978	0.024	25.940	28.571	52.432
2021-02-05	3.004	3.010	3.017	0.032	72.998	76.471	49.736
2021-02-11	3.047	3.086	3.088	0.048	85.846	82.609	53.501

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 3.4 แสดงถึงตัวอย่างข้อมูลหลักทรัพย์ และค่าตัวชี้วัดทางเทคนิค (Technical Indicators) ของบริษัท ดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน) ค่าเฉลี่ยเคลื่อนที่แบบง่าย (SMA: window = 4) ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก (WMA: window = 4) ค่าเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนักแบบ Exponential (EMA: window = 4) ค่าเฉลี่ยเคลื่อนที่แบบคอนเวอร์เจนซ์ไดเวอร์เจนซ์ (MACD: window fast=12, window slow=26) ดัชนีความสัมพันธ์สัมพัทธ์ (RSI: window=4, method="EMA") ออสซิลเลเตอร์สโตแคสติก เค (%K: window=7) และออสซิลเลเตอร์สโตแคสติก ดี (%D: window=7)

### 3.3.2 การเพิ่มประสิทธิภาพตัวบ่งชี้ MACD (MACD)

ตัวชี้วัด Moving Average Convergence Divergence (MACD indicators): MACD ได้มีการปรับค่าพารามิเตอร์ให้เหมาะสมในของแต่ละหลักทรัพย์ โดยแต่ละหลักทรัพย์จะมีค่าพารามิเตอร์ของ window fast และ window slow ที่แตกต่างกัน โดยมีการปรับค่าพารามิเตอร์ให้เหมาะสมด้วยแบบจำลองการสุ่มป่าไม้ (Random Forest) แสดงผลลัพธ์ดังตารางที่ 3.5

ตารางที่ 3.5 ตัวอย่างค่าพารามิเตอร์ของ MACD indicators สำหรับแต่ละหลักทรัพย์

Symbol	Indicators	Version	Window fast	Window slow	STD
ADVANC	MACD	V2	2	25	16.48
PTT	MACD	V2	2	25	6.31
ORI	MACD	V6	3	7	2.86
KTC	MACD	V1	2	34	14.07
AEONTS	MACD	V0	2	30	38.46
SCB	MACD	V2	2	25	28.10
TOP	MACD	V1	2	34	19.37
TASCO	MACD	V0	2	30	3.19
CK	MACD	V2	2	25	4.25
BTS	MACD	V2	2	25	1.71
CPALL	MACD	V2	2	25	8.61
BBL	MACD	V2	2	25	42.09
EA	MACD	V7	3	6	8.64
DELTA	MACD	V6	3	7	126.43
CRC	MACD	V2	2	25	1.36
MINT	MACD	V5	3	11	7.88
MAJOR	MACD	V5	3	11	5.14
SAWAD	MACD	V5	3	11	12.46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.5 (ต่อ)

Symbol	Indicators	Version	Window fast	Window slow	STD
SUPER	MACD	V15	9	19	0.23
STPI	MACD	V12	5	13	1.39
BEC	MACD	V5	3	11	2.97
AMATA	MACD	V18	20	50	4.35
CPF	MACD	V5	3	11	2.49
HANA	MACD	V3	2	27	8.65
GFPT	MACD	V13	5	8	2.01
EGCO	MACD	V5	3	11	51.91
BGRIM	MACD	V3	2	27	9.69
PRM	MACD	V5	3	11	1.26
GULF	MACD	V5	3	11	6.08

### 3.3.3 การเลือกคำสำคัญ (keyword selection)

นี่คือการเลือกปัจจัยที่ส่งผลต่อความปลอดภัยแต่ละอย่าง ปัจจัยที่กล่าวถึงคือคำสำคัญที่ใช้ค้นหาทางอินเทอร์เน็ตและค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (r) คำสำคัญ 20 อันดับแรก ทั้งความสัมพันธ์เชิงบวกและเชิงลบจะถูกเลือกสำหรับการรักษาความปลอดภัยแต่ละรายการ คำค้นหาสูงสุด 20 อันดับแรกทั้งเชิงบวกและเชิงลบของบริษัท ดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน) แสดงไว้ในตารางที่ 3.6

ตารางที่ 3.6 คำสำคัญ (Keywords) 20 อันดับแรกของบริษัท ดับบลิวเอชเอ คอร์ปอเรชั่น จำกัด (มหาชน)

Symbol	Positive	Correlation	Negative	Correlation
WHA	wha	0.2322	restaurant	-0.2028
WHA	sta	0.2262	major	-0.1999
WHA	tcap	0.2161	holiday	-0.1818
WHA	sawad	0.2066	fun	-0.1771
WHA	settrade	0.1885	แขก (kaek)	-0.1693
WHA	dow jones	0.1878	hybride	-0.1671
WHA	banpu	0.1851	food& beverage	-0.1642

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 (ต่อ)

Symbol	Positive	Correlation	Negative	Correlation
WHA	tisco	0.1840	bec	-0.1567
WHA	toa	0.1832	ประชัย (bpra chai)	-0.1511
WHA	aot	0.1824	forex	-0.1430
WHA	bcp	0.1822	short selling	-0.1411
WHA	hmpro	0.1766	water	-0.1362
WHA	ปอด (bpot)	0.1753	mbk	-0.1340
WHA	lh	0.1724	thani	-0.1339
WHA	amata	0.1712	เบาะหนัง (bor nang)	-0.1331
WHA	bdms	0.1689	travel	-0.1249
WHA	cpf	0.1679	top	-0.1224
WHA	bbf	0.1672	ตรกรถ (dtok rot)	-0.1218
WHA	บริการ (bor-ri-gaan)	0.1645	markets	-0.1214
WHA	cpn	0.1615	bts	-0.1198

หมายเหตุ: แวก: ส่วนของ Indorama Ventures PCL (IVL) stock , ประชัย (bpra chai): ส่วนของ TPI Polene PCL (TPIPL) stock, เบาะหนัง: ส่วนของ Interhides PCL (IHL) stock, ตรกรถ (dtok rot): กำลังจะซื้อหลักทรัพย์ตอนที่ราคาตกลงมากแต่กลับขึ้น ปอด (bpot): พอร์ตหลักทรัพย์ และ บริการ: กลุ่มอุตสาหกรรมบริการ

### 3.4 แบบจำลอง (Modelling)

แบบจำลอง (Model) ในการพยากรณ์การเคลื่อนไหวของราคาแต่ละหลักทรัพย์ ได้เปรียบเทียบกัน 3 แบบจำลอง คือ การจำแนกประเภทแบบถดถอยโลจิสติก (Logistic Regression) การจำแนกประเภทแบบการสุ่มป่าไม้ (Random Forest) และ การจำแนกประเภทแบบ Extreme Gradient Boosting (XGBoost)

### 3.4.1 การจำแนกประเภทแบบถดถอยโลจิสติก (Logistic Regression)

การกำหนดพารามิเตอร์ของการจำแนกประเภทแบบถดถอยโลจิสติก มีรายละเอียดแสดงในตารางที่ 3.7

ตารางที่ 3.7 พารามิเตอร์ของการจำแนกประเภทแบบถดถอยโลจิสติก

ชื่อพารามิเตอร์	ความหมาย	ค่า
Inverse	ค่าผกผัน ของ Regularization Strength	1.0
Penalty	พารามิเตอร์ที่กำหนดบรรทัดฐานของแบบจำลองการทำนายผลโดยเก็บไว้ 2 ตัว คือ Lasso Regression (l1) เป็นการตัดบางตัวแปรที่ส่งผลน้อยต่อแบบจำลองออกไปและ Ridge Regression (l2) จะเก็บทุกตัวแปรไว้ไม่มีการตัดทิ้ง	l2

### 3.4.2 การจำแนกประเภทแบบการสุ่มป่าไม้ (Random Forest)

การกำหนดพารามิเตอร์ของการจำแนกประเภทแบบการสุ่มป่าไม้ มีรายละเอียดแสดงในตารางที่ 3.8

ตารางที่ 3.8 พารามิเตอร์ของการจำแนกประเภทแบบการสุ่มป่าไม้

ชื่อพารามิเตอร์	ความหมาย	ค่า
Max Depth	ความสูงของต้นไม้ตัดสินใจ	70
Number of Trees	จำนวนแบบจำลองต้นไม้ตัดสินใจ (Decision Tree) ในการสุ่มป่าไม้ (Random Forest)	1200
Min Samples Split	จำนวนตัวอย่างขั้นต่ำที่ต้องใช้ในการแยกโหนด	10
Min Child Weight	จำนวนผลรวมข้อมูลที่ต่ำที่สุดที่อยู่ที่อยู่ในโหนดใบ	4

### 3.4.3 การจำแนกประเภทแบบ Extreme Gradient Boosting (XGBoost)

การกำหนดพารามิเตอร์ของการจำแนกประเภทแบบ Extreme Gradient Boosting มีรายละเอียดแสดงในตารางที่ 3.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.9 พารามิเตอร์ของการจำแนกประเภทแบบ Extreme Gradient Boosting

ชื่อพารามิเตอร์	ความหมาย	ค่า
Max Depth	ความสูงของต้นไม้ตัดสินใจ	10
Min Child Weight	จำนวนผลรวมข้อมูลที่ต่ำที่สุดที่อยู่ที่อยู่ในโหนดใบ	1
Gamma	จำนวนต่ำสุดของ loss reduction ที่จะใช้แบ่งโหนดใบของต้นไม้ตัดสินใจ	0.01
Learning Rate	อัตราการเรียนรู้ของแบบจำลองโดยมีหน้าที่ป้องกันไม่ให้แบบจำลอง overfit มากเกินไป	0.001

### 3.5 ผลลัพธ์และการอภิปราย (Results and Discussion)

#### 3.5.1 ประสิทธิภาพของชุดข้อมูล (Performance on dataset)

ตารางที่ 3.10 ประสิทธิภาพของชุดข้อมูล (Performance on dataset)

Model	Type dataset	Test					Unknown				
		Accuracy	Precision	Recall	F1-score	AUC	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	Indicators	0.9699	0.9589	0.9699	0.9587	0.8457	0.9818	0.9739	0.9818	0.9745	0.8251
	MACD	0.9663	0.9337	0.9663	0.9497	0.6478	0.9814	0.9638	0.9814	0.9726	0.6616
	Keyword	0.9727	0.9664	0.9727	0.9653	0.8999	0.9840	0.9843	0.9840	0.9779	0.8880
	Indicators & MACD	0.9706	0.9649	0.9706	0.9623	0.9208	0.9775	0.9753	0.9775	0.9763	0.9385
	Keyword & Indicators	0.9813	0.9794	0.9813	0.9790	0.9767	0.9834	0.9827	0.9834	0.9830	0.9733
	Keyword & MACD	0.9733	0.9690	0.9733	0.9674	0.9283	0.9824	0.9769	0.9824	0.9776	0.9326
	Keyword & Indicators & MACD	0.9799	0.9781	0.9799	0.9786	0.9718	0.9824	0.9833	0.9824	0.9828	0.9740
Random Forest	Indicators	0.9694	0.9555	0.9694	0.9553	0.9348	0.9818	0.9638	0.9818	0.9727	0.9324
	MACD	0.9663	0.9511	0.9663	0.9508	0.8445	0.9818	0.9638	0.9818	0.9727	0.7770
	Keyword	0.9727	0.9734	0.9727	0.9619	0.9120	0.9814	0.9717	0.9814	0.9737	0.9045
	Indicators & MACD	0.9666	0.9540	0.9666	0.9520	0.9459	0.9818	0.9638	0.9818	0.9727	0.9447
	Keyword & Indicators	0.9734	0.9708	0.9734	0.9642	0.9645	0.9831	0.9833	0.9831	0.9758	0.9623
	Keyword & MACD	0.9727	0.9714	0.9727	0.9642	0.9374	0.9821	0.9824	0.9821	0.9735	0.9201
	Keyword & Indicators & MACD	0.9741	0.9725	0.9741	0.9671	0.9693	0.9831	0.9833	0.9831	0.9758	0.9638
XGBoost	Indicators	0.9732	0.9674	0.9732	0.9681	0.9577	0.9818	0.9767	0.9818	0.9782	0.9499
	MACD	0.9648	0.9503	0.9648	0.9541	0.8349	0.9818	0.9638	0.9818	0.9727	0.7753
	Keyword	0.9743	0.9694	0.9743	0.9682	0.9339	0.9850	0.9840	0.9850	0.9802	0.9456
	Indicators & MACD	0.9712	0.9653	0.9712	0.9653	0.9568	0.9801	0.9737	0.9801	0.9760	0.9572
	Keyword & Indicators	0.9823	0.9810	0.9823	0.9796	0.9787	0.9879	0.9871	0.9879	0.9855	0.9810
	Keyword & MACD	0.9762	0.9731	0.9762	0.9721	0.9651	0.9831	0.9803	0.9837	0.9785	0.9616
	Keyword & Indicators & MACD	0.9817	0.9803	0.9817	0.9794	0.9838	0.9860	0.9842	0.9860	0.9848	0.9810

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อเลือกหลักทรัพย์ลงทุนโดยใช้ข้อมูล 7 ชุด: (1) ตัวชี้วัดเทคนิค (Technical Indicators), (2) ตัวชี้วัด Moving Average Convergence Divergence (MACD indicators), (3) คำค้นใน Google Trends (Keyword), (4) Indicators & MACD, (5) Keyword & Indicators, (6) Keyword & MACD, และ (7) Keyword & Indicators & MACD

ข้อมูลเปรียบเทียบแบ่งออกเป็น 7 ชุด: (1) ตัวชี้วัดเทคนิค (Technical Indicators) ได้แก่ SMA, WMA, EMA, MACD, RSI, K และ D; ), (2) ตัวชี้วัด Moving Average Convergence Divergence (MACD indicators), (3) คำค้นใน Google Trends (Keyword), (4) Indicators & MACD, (5) Keyword & Indicators, (6) Keyword & MACD, และ (7) Keyword & Indicators & MACD นอกจากนี้ ข้อมูลสำหรับการเรียนรู้ของเครื่องจักร (Machine Learning) ยังแบ่งออกเป็นสามชุด ได้แก่ ฝึกฝนและทดสอบ (Train and Test): 1 มกราคม 2017 ถึง 30 พฤษภาคม 2021 ด้วยอัตราส่วน 80:20 และ Unknown: 1 มิถุนายน 2021 ถึง 31 ธันวาคม 2021 ชุดข้อมูลที่ไม่รู้จักคือ ใช้เพื่อจุดประสงค์ในการทราบประสิทธิภาพของข้อมูลที่ไม่เคยเจอมาก่อน

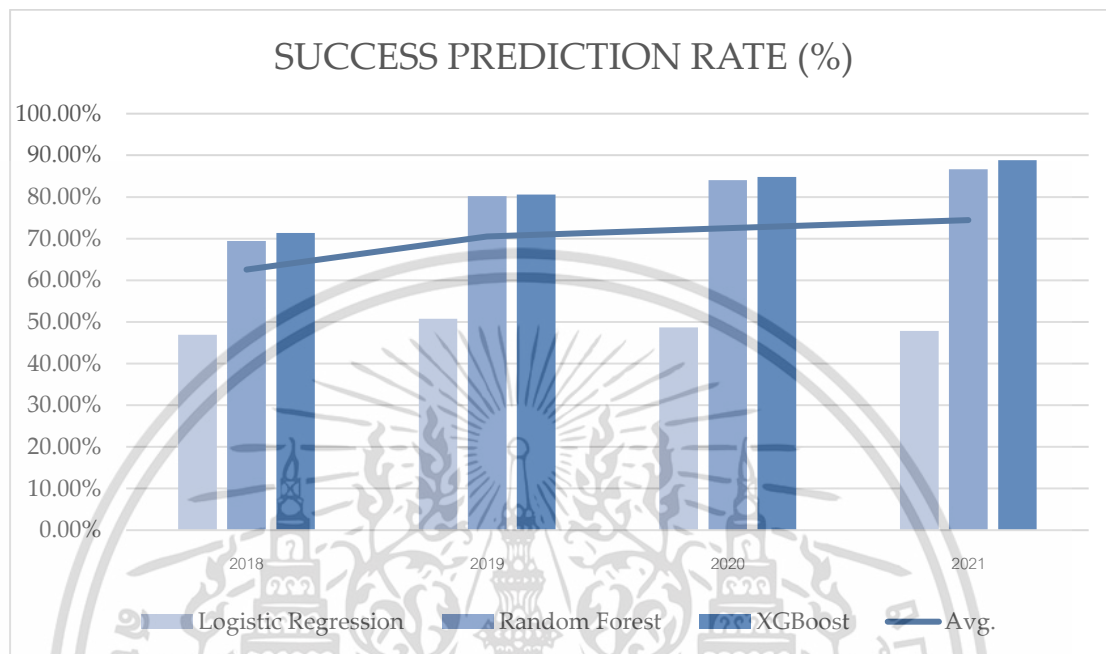
ตารางที่ 3.10 แสดงให้เห็นว่าสำหรับข้อมูลทดลอง (Data Test) ข้อมูลชุดที่ (7): Keyword & Indicators & MACD มีประสิทธิภาพเหนือกว่าชุดข้อมูลอื่นๆ ในแง่ของ accuracy, precision, recall, F1-score and ROC-AUC score สำหรับข้อมูลที่ไม่รู้จัก (Data Unknown) Random Forest และ XGBoost โดยใช้ Keyword & Indicators & MACD เป็นชุดข้อมูลยังมีประสิทธิภาพดีกว่าชุดข้อมูลอื่นๆ โดยเฉพาะอย่างยิ่ง ใช้ Keyword & Indicators & MACD ควบคู่ไปกับ Logistic Regression ทำงานได้ดีขึ้นเมื่อเปรียบเทียบคะแนน F1 และคะแนน ROC-AUC

โดยรวมแล้ว เมื่อเปรียบเทียบประสิทธิภาพการคาดการณ์ของชุดข้อมูลทั้ง 7 ชุด Keyword & Indicators & MACD มีคะแนน F1 และคะแนน ROC-AUC สูงสุดในทั้ง 7 ชุด ดังนั้นบอกเป็นนัยว่าชุดข้อมูลสามารถเปิดเผยแนวโน้มการค้นหออนไลน์ (Online Search Trends) และอนุกรมเวลาทางการเงิน (Financial Time Series) ที่สามารถนำไปสู่การวิเคราะห์ได้ ดังนั้น ตารางที่ 3.11 การเลือกหลักทรัพย์ประสิทธิภาพจึงใช้ชุดข้อมูล Keyword & Indicators & MACD เพื่อการวิเคราะห์เพิ่มเติม

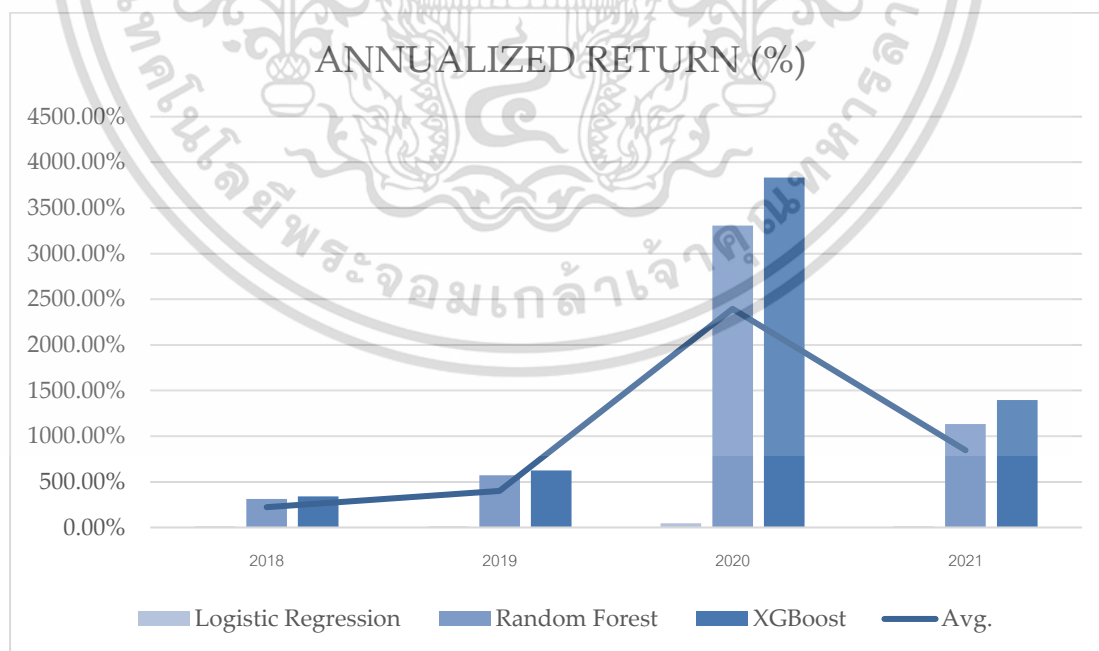
### 3.5.2 ประสิทธิภาพ (Performance)

การวัดประสิทธิภาพด้วยการจำลองตลาดหลักทรัพย์ตามการซื้อขายรายสัปดาห์ กำหนดซื้อที่ราคาปิดของวันทำการสุดท้ายของสัปดาห์ก่อนหน้าและขายที่ราคาปิดของวันสุดท้ายของสัปดาห์ เมื่อจำลองในตลาดหลักทรัพย์ไทย สัปดาห์แรก และการลงทุนในสัปดาห์ต่อ ๆ ไป จะเพิ่มขึ้นหรือลดลงขึ้นอยู่กับผลการลงทุนของสัปดาห์ก่อน โดยแบบจำลองจะคัดเลือกหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรก นอกจากนี้ แบ่งการซื้อขายในแต่ละปีและทุกปีที่เริ่มต้นด้วยเงิน 10,000 บาทตั้งแต่ปี 2018 ถึง 2021 ประสิทธิภาพถูกกำหนดโดยเปอร์เซ็นต์ของการเลือกหลักทรัพย์ที่ประสบความสำเร็จ (เปอร์เซ็นต์ของการเลือกหลักทรัพย์ที่เหมาะสมประจำสัปดาห์) และเปอร์เซ็นต์ผลตอบแทนต่อปี เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(ผลตอบแทนรวมในแต่ละสัปดาห์รวม 1 ปี) ผลลัพธ์ของ Logistic Regression, Random Forest และ XGBoost ถูกนำมาเปรียบเทียบในรูปที่ 3.2 – 3.4 ซึ่งให้รายละเอียดเกี่ยวกับประสิทธิภาพการเลือกหลักทรัพย์และผลตอบแทนรวมประจำปีตั้งแต่ปี 2018 ถึง 2021

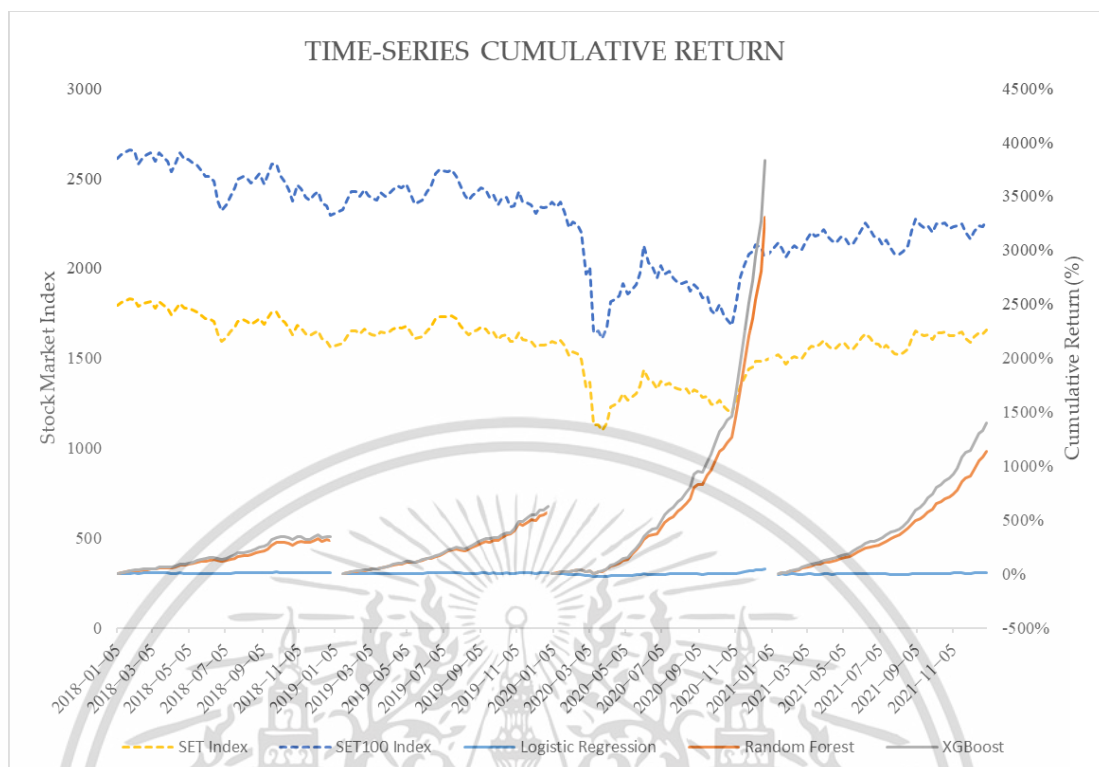


รูปที่ 3.2 เปอร์เซ็นต์การเลือกหลักทรัพย์ที่ประสบความสำเร็จ (Success Prediction Rate)



รูปที่ 3.3 เปอร์เซ็นต์ผลตอบแทนต่อปี (Annualized Return)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 ผลตอบแทนสะสมตามอนุกรมเวลา (Time-Series Cumulative Return)

ต่อไปนี้เป็นผลลัพธ์ประจำปีตั้งแต่ปี 2018 ถึง 2021 (รูปที่ 3.2 – 3.3): อัตราความสำเร็จเฉลี่ยสำหรับแบบจำลอง Logistic Regression คือ 49.55% และผลตอบแทนเฉลี่ยต่อปีคือ 19.13% โดยอัตราการทำนายความสำเร็จสูงสุดในปี 2019 คือ 50.80 % และผลตอบแทนต่อปีสูงสุดในปี 2020 อยู่ที่ 45.42 % แบบจำลอง Random Forest มีอัตราความสำเร็จเฉลี่ย 80.06% และผลตอบแทนเฉลี่ยต่อปี 1330.78% โดยอัตราการทำนายความสำเร็จที่ดีที่สุดในปี 2021 คือ 86.67% และผลตอบแทนต่อปีที่ดีที่สุดในปี 2020 คือ 3,305.68% แบบจำลอง XGBoost มีอัตราความสำเร็จเฉลี่ย 81.39% และผลตอบแทนเฉลี่ยต่อปีที่ 1,549.06% โดยมีอัตราความสำเร็จ 88.82% สำหรับปีที่ดีที่สุด 2021 และ 3,834.02% สำหรับปีที่ดีที่สุด 2020 จากรูปที่ 3.4 Logistic Regression, Random Forest และ XGBoost แสดงให้เห็นถึงความโดดเด่นเหนือดัชนีมาตรฐาน SET และดัชนี SET100 ในพอร์ตโฟลิโอตลอดระยะเวลาจำลอง (ค.ศ. 2018 ถึง ค.ศ. 2021) ความโดดเด่นของผลตอบแทนมีความสำคัญมากขึ้นโดยเริ่มตั้งแต่ช่วงการระบาดของ COVID-19 ในปี 2020 โดยสรุปการเปรียบเทียบประสิทธิภาพกับการประเมินทางการเงินย้อนหลังประจำปีเผยให้เห็นความแตกต่างระหว่างทั้งสามแบบจำลอง เห็นได้ชัดว่าการใช้ Random Forest มีเปอร์เซ็นต์ความสำเร็จโดยเฉลี่ยสูงสุด ในขณะที่ XGBoost มีผลตอบแทนต่อปีโดยเฉลี่ยสูงสุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในปี 2020 เป็นปีที่มีผลตอบแทนมากที่สุด เนื่องจากตลาดหลักทรัพย์ไทยมีความผันผวนอย่างมากจากเหตุการณ์โควิด-19 และวิกฤตการณ์ต่าง ๆ ทุก ๆ ปีจะมีเหตุการณ์ที่มีผลกระทบต่อตลาดหลักทรัพย์ทั้งในเชิงบวกและเชิงลบ ดังนั้น ตารางที่ 3.11 จึงแสดงประสิทธิภาพของแบบจำลองสำหรับวิกฤตต่าง ๆ ในประเทศไทย

### 3.5.3 ประสิทธิภาพในช่วงวิกฤต (Performance for Crisis)

ในช่วงวิกฤตต่าง ๆ ในประเทศไทย ประสิทธิภาพของโมเดลจะแตกต่างกันไปตามช่วงเวลา จากปี 2018 ถึง 2021 ตามรูปที่ 3.5 มีวิกฤต 6 ประเภท ได้แก่ Quantitative Easing (QE), Controversy, Foreign Investors, MSCI, Protestation และ Covid-19 สำหรับวิกฤตการณ์โควิด-19 แบ่งได้เป็น 5 สถานะของโควิด เนื่องจากแต่ละสถานะค่อนข้างมีเอกลักษณ์เฉพาะตัวในแง่ของการแสดงตัวแบบในตลาดหลักทรัพย์ไทย ดังนั้นจึงอธิบายตามสถานะของการระบาดของโควิดและความสนใจไปที่ผลกระทบในแต่ละช่วง

	2018	2019	2020	2021
Jan	Quantitative Easing (QE)			Covid-19 (3)
Feb			Covid-19 (1)	
Mar				
Apr				
May		MSCI		
Jun	Controversy			
Jul				
Aug				Covid-19 (4)
Sep				
Oct			Protestation	
Nov				Covid-19 (5)
Dec	Foreign Investors		Covid-19 (2)	

รูปที่ 3.5 วิกฤตการณ์ในประเทศไทยระหว่างปี ค.ศ. 2019 – 2021

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.11 ประสิทธิภาพสำหรับวิกฤต (Performance for Crisis)

% Success	Model			Win Model
	Crisis	Logistic Regression	Random Forest	
QE	55.00%	85.00%	83.75%	Random Forest
Controversy	51.11%	78.89%	65.56%	Random Forest
Foreign Investors	45.00%	71.67%	63.33%	Random Forest
MSCI	57.78%	86.67%	82.22%	Random Forest
Covid-19 (1)	51.54%	73.85%	79.23%	XGBoost
Protestation	65.56%	94.44%	90.00%	Random Forest
Covid-19 (2)	42.86%	90.00%	91.43%	XGBoost
Covid-19 (3)	40.00%	82.22%	84.44%	XGBoost
Covid-19 (4)	56.25%	92.50%	93.75%	XGBoost
Covid-19 (5)	51.11%	90.00%	90.00%	Random Forest
AVERAGE	51.62%	84.52%	82.37%	Random Forest

#### Quantitative Easing (QE)

มาตรการผ่อนคลายเชิงปริมาณ (Quantitative Easing QE) เป็นรูปแบบหนึ่งของนโยบายการเงินที่ธนาคารกลางของแต่ละประเทศกำหนดขึ้น ในเดือนมกราคม 2018 ดัชนีตลาดหลักทรัพย์แห่งประเทศไทยปิดที่สูงสุดที่ 1,838.96 จุด ซึ่งเป็นปัจจัยสำคัญในการลดลงของราคาทองคำ (ชัย et al., 2021) เนื่องจากผลกระทบของมาตรการ QE ในสหรัฐอเมริกาและญี่ปุ่นและการปรับลดอัตราดอกเบี้ยของธนาคารกลางยุโรป (ECB) ซึ่งกระตุ้นให้นักลงทุนลงทุนในตลาดหลักทรัพย์เพื่อหาผลตอบแทนที่สูงขึ้น ผลลัพธ์ของ Logistic Regression, Random Forest และ XGBoost ตั้งแต่เดือนมกราคมถึงกุมภาพันธ์ 2018 ดังแสดงในตารางที่ 3.11 เผยให้เห็นอัตราการเลือกหลักทรัพย์ที่ประสบความสำเร็จที่ 55.0%, 85.0% และ 83.75% ตามลำดับ

#### การขัดแย้ง (Controversy)

การตอบโต้ทางการค้าของสหรัฐฯและจีนเป็นสงครามการค้าโลก โดยเฉพาะมาตรการตอบโต้ทางการค้าระหว่างสหรัฐฯ และจีน ส่งผลให้ดัชนี SET ปรับตัวลงเกือบทั้งสัปดาห์ แม้ว่าจะมีปัจจัยสนับสนุนหลักทรัพย์กลุ่มพลังงานจากราคาน้ำมันในตลาดโลกที่ปรับตัวสูงขึ้น อย่างไรก็ตาม ยังไม่เพียงพอที่จะชดเชยแรงกดดันจากการขายสุทธิหลักทรัพย์ไทยของนักลงทุนต่างชาติตั้งแต่ต้นถึงกลางสัปดาห์ อย่างไรก็ตาม ตลาดหลักทรัพย์ไทยยังคงปิดสีแดงสวนทางกับตลาดหลักทรัพย์ภูมิภาค ในเดือนมิถุนายนและกรกฎาคม 2018 อัตราการคัดเลือกหลักทรัพย์ที่ประสบความสำเร็จสำหรับ

Logistic Regression, Random Forest และ XGBoost คือ 51.11%, 78.89% และ 65.55% ตามลำดับ ดังแสดงในตารางที่ 3.11

### Foreign Investors (นักลงทุนต่างชาติ)

SET ปิดลบ 1,548.37 จุด ต่ำสุดรอบ 19 เดือน ณ วันที่ 27 ธันวาคม 2018 สาเหตุหลักมาจากนักลงทุนต่างชาติขายสุทธิ 1,832.50 ล้านบาท (Smart, 2018; Today, 2018) ดังแสดงในตารางที่ 3.11 ผลลัพธ์ของ Logistic Regression, Random Forest และ XGBoost ในช่วงเดือนธันวาคม 2018 ถึงมกราคม 2019 แสดงอัตราการเลือกหลักทรัพย์ที่ 45.00%, 71.67% และ 63.33% ตามลำดับ

### MSCI

MSCI Index คือเกณฑ์มาตรฐานของ Morgan Stanley Capital International พัฒนาขึ้นเพื่อเป็นตัวชี้วัดสำหรับนักลงทุนต่างชาติเพื่อใช้ในการเลือกหลักทรัพย์และผลตอบแทน วันที่ 28 พฤษภาคม 2019 ตลาดหลักทรัพย์ไทยปิดที่ 1,632.04 จุด เพิ่มขึ้น 7.20 จุด มูลค่าการซื้อขาย 204,855.67 ล้านบาท สูงสุดเป็นประวัติการณ์ เนื่องจากการปรับน้ำหนักของดัชนี MSCI Thailand ซึ่งเพิ่มน้ำหนักของหลักทรัพย์ไทยบางส่วนในการลงทุนของดัชนี จากตารางที่ 3.11 อัตราการคัดเลือกหลักทรัพย์ที่คาดการณ์โดย Logistic Regression, Random Forest และ XGBoost ในช่วงเดือนพฤษภาคม-มิถุนายน 2019 อยู่ที่ 57.78%, 86.67% และ 82.22% ตามลำดับ

### โควิด-19 (COVID-19) (1)

วิกฤตโควิด-19 ส่งผลกระทบต่อเศรษฐกิจโลกหลายประการ โดยมาตรการล็อกดาวน์ส่งผลกระทบต่อเศรษฐกิจมากที่สุด เมื่อมีการล็อกดาวน์ ทำให้ผู้คนและธุรกิจต่าง ๆ ชื้อของในที่สาธารณะยากลำบากขึ้น เช่น ห้างสรรพสินค้าถูกสั่งปิด เริ่มส่งผลกระทบต่อตลาดหลักทรัพย์ไทยในวันที่ 26 กุมภาพันธ์ 2020 ตลาดหลักทรัพย์ไทยร่วงหนัก 72.69 จุด หรือ -5.05% ในวันเดียว ปิดที่ 1,366.41 จุด SET ปิดการซื้อขายวันที่ 9 มี.ค. ที่ 1,255.94 จุด ลดลง 108.63 จุด หรือ -7.96% SET วันที่ 12 มี.ค. ปิดที่ 1,114.91 จุด ลดลง 134.98 จุด หรือขาดทุน 10.80% ตลาดหลักทรัพย์ปิดวันที่ 16 มี.ค. อยู่ที่ 1,046.08 จุด ลดลง 82.83 จุด หรือ 7.34% ตามที่ระบุไว้ในตารางที่ 3.11 ประสิทธิภาพของการเลือกหลักทรัพย์ตั้งแต่เดือนกุมภาพันธ์ถึงเมษายน 2563 โดยใช้ Logistic Regression, Random Forest และ XGBoost อยู่ที่ 51.54%, 73.85% และ 79.22% ตามลำดับ

### การประท้วง (Protestation)

การประท้วงในกรุงเทพมหานคร ประเทศไทย เมื่อวันที่ 30 ตุลาคม 2020 นำไปสู่การประกาศภาวะฉุกเฉินในเมืองหลวง ส่งผลให้ดัชนี SET ปิดการซื้อขายวันนี้ที่ 1,214.95 จุด การใช้เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Logistic Regression, Random Forest และ XGBoost จากตารางที่ 3.11 อัตราการทำนายความสำเร็จสำหรับเดือนตุลาคมและพฤศจิกายน 2020 เป็นดังนี้ 65.56%, 94.44% และ 90% ตามลำดับ

### โควิด-19 (COVID-19) (2)

การระบาดของโรคโควิด-19 เริ่มขึ้นที่จังหวัดสมุทรสาครในประเทศไทย พร้อมๆ กับที่ประเทศได้ออกมาตรการกักกันเพื่อหยุดไม่ให้ไวรัสแพร่กระจายต่อไป ด้วยความกังวลของนักลงทุนเกี่ยวกับมาตรการกักกันของจังหวัดสมุทรสาคร ดัชนีตลาดหลักทรัพย์แห่งประเทศไทยในวันที่ 21 ธันวาคม 2020 ลดลงมากกว่า 5% ปิดตลาดโดยขาดทุน 80.60 จุด หรือ -5.44% ที่ 1,401.78 จุด จากตารางที่ 3.11 อัตราความสำเร็จของการเลือกหลักทรัพย์ที่คาดการณ์โดย Logistic Regression, Random Forest และ XGBoost ในช่วงเดือนธันวาคม 2020 ถึงมกราคม 2021 อยู่ที่ 42.86%, 90% และ 91.4% ตามลำดับ

### โควิด-19 (COVID-19) (3)

ผลจากการผ่อนคลายมาตรการล็อกดาวน์ ดัชนีตลาดหลักทรัพย์แห่งประเทศไทย ปิดวันที่ 11 มิ.ย. 2021 อยู่ที่ 1,636.56 จุด เพิ่มขึ้น 11.29 จุด สูงสุดในรอบ 1 ปี 4 เดือน 18 วัน อย่างไรก็ตามเมื่อมาตรการล็อกดาวน์ผ่อนคลายลง การแพร่ระบาดก็รุนแรงขึ้น และตลาดหลักทรัพย์ก็ปรับตัวลงอย่างต่อเนื่อง จากตารางที่ 3.11 ประสิทธิภาพของการเลือกหลักทรัพย์ที่ประสบความสำเร็จโดยใช้ Logistic Regression, Random Forest และ XGBoost ในช่วงเดือนมิถุนายนถึงกรกฎาคม 2021 คือ 40%, 82.22% และ 84.44% ตามลำดับ

### โควิด-19 (COVID-19) (4)

เนื่องจากจำนวนผู้ป่วยโควิด-19 ในประเทศยังคงสร้างแรงกดดันมากขึ้น ตลาดหลักทรัพย์แห่งประเทศไทยปิดที่ 1521.72 จุดในวันที่ 6 สิงหาคม 2021 ลดลง 5.94 จุดและปิดต่ำสุดในรอบ 5 เดือน สาเหตุมาจากตลาดหลักทรัพย์ไทยยังเผชิญกับแรงกดดันจากตัวเลขผู้ติดเชื้อโควิด-19 ในประเทศ ที่ยังสูงกว่าวันละ 20,000 ราย ตลาดตระหนักดีถึงประเด็นนี้ แต่ก็ยังมีความเสี่ยง นักลงทุนจึงระมัดระวัง จากการรวบรวมโดยตลาดหลักทรัพย์แห่งประเทศไทย (ตลท.) เมื่อวันที่ 6 สิงหาคม 2021 นักลงทุนสถาบันขายหลักทรัพย์ 135.90 ล้านบาท โบรกเกอร์ขายหลักทรัพย์ 363.93 ล้านบาท นักลงทุนต่างชาติขายหลักทรัพย์ 2,078.31 ล้านบาท และนักลงทุนรายย่อยซื้อหลักทรัพย์มูลค่า 2,578.14 ล้านบาท ยอดขายของนักลงทุนต่างชาติในช่วง 6 วันแรกของเดือนสิงหาคมมีมูลค่ารวมสูงถึง 8,148.43 ล้านบาท ดังแสดงในตารางที่ 3.11 อัตราความสำเร็จของการเลือกหลักทรัพย์โดยใช้ Logistic Regression, Random Forest และ XGBoost ในช่วงเดือนสิงหาคมและกันยายน 2021 อยู่ที่ 56.25%, 92.50% และ 93.75% ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### โควิด-19 (COVID-19) (5)

วันที่ 28 พฤศจิกายน 2021 ตลาดหลักทรัพย์แห่งประเทศไทย ลดลง 37.85 จุด ปิดที่ 1610.61 จุด จากไวรัสโคโรนาสายพันธุ์โอไมครอน 2019 แนวโน้มลดลงอย่างต่อเนื่อง ดัชนีร่วงอีก 20.92 จุดในวันที่ 29 พฤศจิกายน 2021 ปิดวันที่ 1589.69 จุด ลดลงอีก 21.00 จุดในวันที่ 30 พฤศจิกายน 2021 จบที่ 1568.69 จุด ตารางที่ 3.11 แสดงให้เห็นว่าอัตราความแม่นยำของ Logistic Regression (51.11%), Random Forest (90.00%) และ XGBoost (90.00%) ตามลำดับ ต่ำที่สุดและสูงสุดสำหรับการทำนายหลักทรัพย์ที่ทำกำไรในช่วงเดือนพฤศจิกายนและธันวาคมปี 2021

จากวิกฤตการณ์ต่าง ๆ ที่เกิดขึ้นในประเทศไทยระหว่างปี 2018 - 2021 วิกฤตแบ่งออกเป็น 6 ประเภท ได้แก่ QE, Controversy, Foreign Investors, MSCI, COVID-19 และ Protest ประสิทธิภาพโดยรวมของ Logistic Regression มีอัตราความแม่นยำโดยรวม 51.62% Random Forest มีอัตราความแม่นยำโดยรวม 84.52% (ประสิทธิภาพสูงสุดในช่วงวิกฤตการเงิน: QE, Controversy, Foreign Investors, MSCI, Covid-19 (5) และ Protest) สำหรับ XGBoost ประสิทธิภาพโดยรวมสะท้อนถึงความสำเร็จ 82.37% (ประสิทธิภาพสูงสุดในช่วงวิกฤต: โควิด-19 (1-5)) สุดท้ายสำหรับการประเมินทางการเงินย้อนหลังในช่วงวิกฤต Random Forest อยู่ในอันดับแรกในแง่ของ % ความสำเร็จ 84.52% ในขณะที่ XGBoost อยู่ในอันดับแรกในช่วงการระบาดของ COVID-19 ด้วย 87.77%

จากประสิทธิภาพข้างต้น Random Forest มีประสิทธิภาพโดยรวมในเปอร์เซ็นต์การเลือกหลักทรัพย์ที่ประสบความสำเร็จดีกว่า 2 แบบจำลอง (Logistic Regression และ XGBoost) เนื่องจากอัลกอริทึมมีความเสถียรมาก และทำงานได้ดีกับทั้งตัวแปรต่อเนื่องและตัวแปรที่ไม่ต่อเนื่อง แม้ว่าข้อมูลใหม่จะถูกเพิ่มเข้าไปในชุดข้อมูล อัลกอริทึมโดยรวมจะไม่ได้รับผลกระทบอย่างมีนัยสำคัญ เนื่องจากไม่น่าเป็นไปได้ที่ข้อมูลใหม่จะส่งผลกระทบต่อแบบจำลองต้นไม้มทั้งหมด (Breiman, 2001) อย่างไรก็ตาม ในช่วง Covid-19 นั้นตลาดมีการปรับตัวลงอย่างรุนแรง XGBoost มีเปอร์เซ็นต์การเลือกหลักทรัพย์ที่ประสบความสำเร็จที่สูงกว่า เนื่องมาจากลักษณะข้อมูลช่วง Covid-19 นั้นส่งผลต่อตลาดหุ้นให้มีการปรับตัวลงอย่างรุนแรง ซึ่งการปรับตัวแบบนี้เป็น คลาส (Class) ส่วนน้อย เป็นไปได้ว่า Random Forest จะมีการให้ความสำคัญต่อ class นี้น้อย แต่ XGBoost เป็นทางเลือกที่ยอดเยี่ยมสำหรับชุดข้อมูลที่ไม่สมดุล ด้วยเหตุนี้ XGBoost จึงมีเปอร์เซ็นต์การเลือกหลักทรัพย์ที่ประสบความสำเร็จที่สูงกว่า ในขณะที่ Logistic Regression มีประสิทธิภาพที่ค่อนข้างต่ำ เมื่อเทียบกับ Random Forest และ XGBoost เนื่องมาจาก Logistic Regression ได้รับการปรับให้เหมาะสมสำหรับตัวแปรที่ไม่ต่อเนื่อง (Robles et al., 2008) ข้อมูลเมื่อเปรียบเทียบกับ Random Forest และ XGBoost แต่ลักษณะข้อมูลของ Keyword & Indicators & MACD มีความซับซ้อน ซึ่งมีทั้งลักษณะข้อมูลแบบตัวแปรต่อเนื่องและไม่ต่อเนื่อง

## บทที่ 4

### วิธีการดำเนินงานวิจัยการจัดพอร์ตหลักทรัพย์ที่เหมาะสม

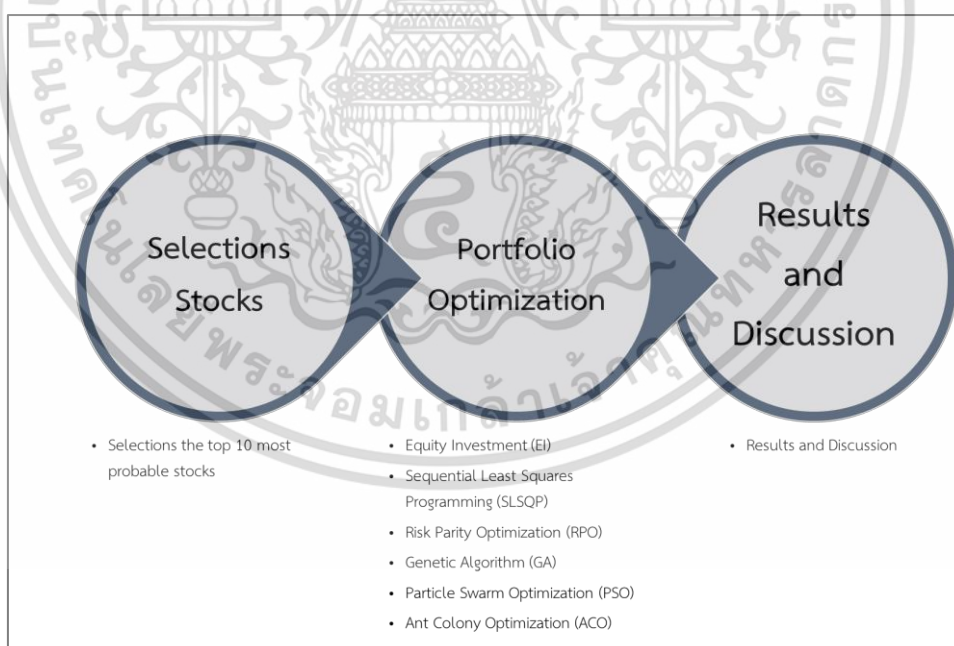
บทนี้จะกล่าวถึงวิธีการดำเนินงานวิจัยในครั้งนี้ ซึ่งมีวัตถุประสงค์เพื่อการจัดพอร์ตหลักทรัพย์ที่เหมาะสม โดยเนื้อหาในบทนี้ประกอบด้วยหัวข้อหลัก ๆ คือ แผนผังขั้นตอนการดำเนินงาน (Process Diagram) ข้อมูล (Dataset) การเตรียมข้อมูล (Data Preprocessing) แบบจำลอง (Modelling) และผลลัพธ์และการอภิปราย (Results and Discussion)

#### 4.1 แผนผังขั้นตอนการดำเนินงาน (Process Diagram)

##### ขั้นตอนการดำเนินงาน

การดำเนินงานวิจัยในครั้งนี้ได้มีกระบวนการในการทำงานวิจัยหลายกระบวนการแสดงในรูปที่ 4.1 โดยมีขั้นตอนในการดำเนินงาน คือ

1. การเลือกหลักทรัพย์ (Selections Stocks)
2. การจัดพอร์ตหลักทรัพย์ที่เหมาะสม (Portfolio Optimization)
3. ผลลัพธ์และการอภิปราย (Results and Discussion)



รูปที่ 4.1 แผนผังขั้นตอนการดำเนินงาน (Process Diagram)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4.2 การเลือกหลักทรัพย์ (Selections Stocks)

การเลือกหลักทรัพย์ด้วยแบบจำลองการสุ่มป่าไม้ (Random Forest) เนื่องจากมีประสิทธิภาพโดยรวมในเปอร์เซ็นต์การเลือกหลักทรัพย์ที่ประสบความสำเร็จดีกว่า 2 แบบจำลอง (Logistic Regression และ XGBoost) กับข้อมูลชุดที่ (7) Keyword & Indicators & MACD เนื่องจากประสิทธิภาพมีคะแนน F1 และคะแนน ROC-AUC สูงสุดในทั้ง 7 ชุด โดยแบบจำลองจะคัดเลือกหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 โดยมีรายละเอียดแสดงในตารางที่ 4.1

ตารางที่ 4.1 ตัวอย่างหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021

Date	symbol_1	symbol_2	symbol_3	symbol_4	symbol_5	symbol_6	symbol_7	symbol_8	symbol_9	symbol_10
2021-01-17	STA	RS	HANA	DELTA	TISCO	ESSO	JMT	TU	KCE	EA
2021-01-24	STA	KCE	RS	JMT	CBG	DELTA	EA	PTTEP	SPRC	GLOBAL
2021-01-31	EPG	RS	CBG	EA	BPP	KCE	THG	DELTA	JMT	AMATA
2021-02-07	BPP	AWC	CENTEL	ERW	MINT	JMT	RS	COM7	PLANB	KBANK
2021-02-14	AEONTS	EPG	ERW	BANPU	MINT	PSL	RS	BEC	CENTEL	DTAC
2021-02-21	STA	PSL	EPG	CPF	JMT	ERW	BEC	MINT	RATCH	CENTEL
2021-02-28	BJC	STA	PSL	GUNKUL	SPRC	TOP	CPALL	CK	MAJOR	CENTEL
2021-03-07	BEC	STEC	SAWAD	KTC	PSL	AOT	MAJOR	AMATA	JMT	GUNKUL
2021-03-14	GUNKUL	AWC	AOT	BH	BJC	BEC	CPALL	TOP	LH	AP
2021-03-21	GUNKUL	HANA	PSL	STEC	COM7	CENTEL	SAWAD	JMT	LH	CKP
2021-03-28	BANPU	AP	KTC	AEONTS	EGCO	PSL	SAWAD	JMT	ERW	RS

จากตารางที่ 4.1 หลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 จะเป็นหลักทรัพย์พื้นฐานสำหรับการจัดพอร์ตหลักทรัพย์ เพื่อเปรียบเทียบประสิทธิภาพการหาอัตราส่วนที่เหมาะสมต่อการจัดพอร์ตที่เหมาะสม (Portfolio Optimization) ระหว่าง การลงทุนแบบเท่าเทียม (Equity Investment - EI), Sequential Least Squares Programming algorithm (SLSQP), Risk Parity Optimization (RPO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), และ Particle Swarm Optimization (PSO)

## 4.3 การจัดพอร์ตหลักทรัพย์ที่เหมาะสม (Portfolio Optimization)

### 4.3.1 การลงทุนแบบเท่าเทียม (Equity Investment - EI)

การลงทุนแบบเท่าเทียม (Equity Investment - EI) คือการให้อัตราส่วนในการลงทุนหลักทรัพย์สำหรับการจัดพอร์ตเท่ากัน โดยอัตราส่วนคือ 10% ทุกสัปดาห์ โดยมีรายละเอียดแสดงในตารางที่ 4.2

ตารางที่ 4.2 ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 เมื่อใช้การลงทุนแบบเท่าเทียม (Equity Investment)

Date	weights_1	weights_2	weights_3	weights_4	weights_5	weights_6	weights_7	weights_8	weights_9	weights_10
2021-01-17	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-01-24	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-01-31	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-02-07	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-02-14	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-02-21	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-02-28	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-03-07	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-03-14	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-03-21	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%
2021-03-28	10%	10%	10%	10%	10%	10%	10%	10%	10%	10%

จากตารางที่ 4.2 โดย symbol\_1 และ weights\_1 คือ อัตราส่วนการลงทุนของ symbol\_1 (หลักทรัพย์ที่ 1) ซึ่งมีความสัมพันธ์แบบเดียวกันจนถึง weights\_10

#### 4.3.2 Sequential Least Squares Programming algorithm (SLSQP)

การกำหนดพารามิเตอร์ของ Sequential Least Squares Programming algorithm (SLSQP) มีรายละเอียดแสดงในตารางที่ 4.3

ตารางที่ 4.3 พารามิเตอร์ของ SLSQP

ชื่อพารามิเตอร์	ความหมาย	ค่า
Objective function	ฟังก์ชันวัตถุประสงค์	Sharpe ratio
Risk free rate	อัตราดอกเบี้ยปราศจากความเสี่ยง	2%

จากตารางที่ 4.3 นั้นนำไปสู่การหาค่าอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรก โดยมีรายละเอียดแสดงในตารางที่ 4.4

**ตารางที่ 4.4** ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละ สัปดาห์ ในช่วงปี 2021 (เมื่อจัดพอร์ตแบบ SLSQP)

Date	weights_1	weights_2	weights_3	weights_4	weights_5	weights_6	weights_7	weights_8	weights_9	weights_10
2021-01-17	4.85%	0.00%	0.00%	95.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-01-24	9.71%	0.00%	0.00%	0.00%	0.00%	90.29%	0.00%	0.00%	0.00%	0.00%
2021-01-31	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
2021-02-07	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
2021-02-14	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
2021-02-21	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-02-28	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-03-07	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-03-14	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
2021-03-21	0.00%	0.00%	52.51%	0.00%	0.00%	0.00%	0.00%	47.49%	0.00%	0.00%
2021-03-28	0.00%	0.00%	0.00%	0.00%	0.00%	41.04%	0.00%	58.96%	0.00%	0.00%

**4.3.3 การปรับปรุงความเสี่ยงแบบเทียบค่าความเสี่ยง (Risk Parity Optimization - RPO)**

โดย Risk Parity Optimization – RPO มีหน้าที่เป็นฟังก์ชันวัตถุประสงค์ ดังนั้นได้ใช้ผสมกับ SLSQP algorithm การกำหนดพารามิเตอร์ของ SLSQP algorithm โดยมีรายละเอียดแสดงในตารางที่ 4.5 และขอเรียกวิธีดังกล่าวว่า “Risk Parity Optimization or RPO”

**ตารางที่ 4.5** พารามิเตอร์ของ RPO

ชื่อพารามิเตอร์	ความหมาย	ค่า
Objective function	ฟังก์ชันวัตถุประสงค์	RPO
Risk free rate	อัตราดอกเบี้ยปราศจากความเสียหาย	2%
Method	กระบวนการ	SLSQP

จากตารางที่ 4.5 นั้นนำไปสู่การหาค่าอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรก โดยมีรายละเอียดแสดงในตารางที่ 4.6

**ตารางที่ 4.6** ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 (เมื่อจัดพอร์ตแบบ RPO)

Date	weights_1	weights_2	weights_3	weights_4	weights_5	weights_6	weights_7	weights_8	weights_9	weights_10
2021-01-17	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
2021-01-24	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-01-31	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-02-07	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%
2021-02-14	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%
2021-02-21	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%
2021-02-28	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-03-07	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%
2021-03-14	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%
2021-03-21	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-03-28	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%

#### 4.3.4 ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm - GA)

การกำหนดพารามิเตอร์ของ ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm - GA) โดยมีรายละเอียดแสดงในตารางที่ 4.7

**ตารางที่ 4.7** พารามิเตอร์ของ Genetic Algorithm

ชื่อพารามิเตอร์	ความหมาย	ค่า
Objective function	ฟังก์ชันวัตถุประสงค์	Sharpe ratio
Risk free rate	อัตราดอกเบี้ยปราศจากความเสี่ยง	2%
pop_size	ขนาดของประชากร	100
num_generations	จำนวนรุ่นหรือการวนซ้ำที่อัลกอริทึมจะทำงาน	100
mutation_prob	โอกาสหรือความน่าจะเป็นของการกลายพันธุ์ที่เกิดขึ้นระหว่างกระบวนการสืบพันธุ์	0.01

จากตารางที่ 4.7 นั้นนำไปสู่การหาค่าอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรก โดยมีรายละเอียดแสดงในตารางที่ 4.8

**ตารางที่ 4.8** ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 (เมื่อจัดพอร์แบบ Genetic Algorithm)

Date	weights_1	weights_2	weights_3	weights_4	weights_5	weights_6	weights_7	weights_8	weights_9	weights_10
2021-01-17	10.89%	15.67%	15.32%	1.83%	6.86%	13.95%	13.47%	12.03%	3.72%	6.27%
2021-01-24	7.66%	1.71%	16.89%	8.16%	26.46%	0.36%	8.93%	3.15%	19.30%	7.39%
2021-01-31	13.60%	14.22%	8.36%	10.65%	10.96%	3.93%	15.65%	1.24%	13.11%	8.27%
2021-02-07	21.62%	8.81%	10.38%	13.66%	9.99%	7.23%	9.95%	10.23%	3.06%	5.06%
2021-02-14	12.44%	7.87%	20.06%	7.94%	20.33%	3.95%	0.84%	2.40%	3.15%	21.01%
2021-02-21	2.85%	1.66%	1.71%	24.11%	5.83%	18.83%	10.45%	12.46%	14.42%	7.69%
2021-02-28	19.41%	0.36%	8.52%	4.14%	17.05%	2.54%	17.93%	15.42%	11.24%	3.38%
2021-03-07	3.26%	14.44%	9.74%	9.58%	4.07%	13.82%	9.46%	14.58%	4.84%	16.22%
2021-03-14	12.80%	1.75%	18.30%	16.88%	4.06%	2.03%	17.24%	4.04%	13.76%	9.13%
2021-03-21	15.29%	9.29%	3.22%	7.28%	10.98%	6.64%	3.63%	14.37%	23.69%	5.61%
2021-03-28	13.69%	4.17%	5.00%	4.14%	24.65%	6.80%	5.09%	14.51%	15.00%	6.95%

#### 4.3.5 อาณาจักรมด (Ant Colony Optimization - ACO)

การกำหนดพารามิเตอร์ของ อาณาจักรมด (Ant Colony Optimization - ACO) มีรายละเอียดแสดงในตารางที่ 4.9

**ตารางที่ 4.9** พารามิเตอร์ของ Ant Colony Optimization

ชื่อพารามิเตอร์	ความหมาย	ค่า
Objective function	ฟังก์ชันวัตถุประสงค์	Sharpe ratio
Risk free rate	อัตราดอกเบี้ยปราศจากความเสี่ยง	2%
colony_size	ขนาดอาณานิคม	50
max_iter	จำนวนรอบในการเรียนรู้ของวิธีอาณาจักรมด	50
evaporation rate	อัตราการระเหย	5%

จากตารางที่ 4.9 นั้นนำไปสู่การหาค่าอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรก โดยมีรายละเอียดแสดงในตารางที่ 4.10

**ตารางที่ 4.10** ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 (เมื่อจัดพอร์แบบ Ant Colony Optimization)

Date	weights_1	weights_2	weights_3	weights_4	weights_5	weights_6	weights_7	weights_8	weights_9	weights_10
2021-01-17	0.00%	2.00%	2.00%	82.00%	2.00%	2.00%	4.00%	2.00%	2.00%	2.00%
2021-01-24	18.00%	2.00%	2.00%	2.00%	2.00%	64.00%	2.00%	4.00%	2.00%	2.00%
2021-01-31	0.00%	2.00%	4.00%	2.00%	6.00%	2.00%	4.00%	60.00%	18.00%	2.00%
2021-02-07	0.00%	2.00%	2.00%	2.00%	6.00%	4.00%	72.00%	8.00%	2.00%	2.00%
2021-02-14	2.00%	2.00%	10.00%	2.00%	2.00%	72.00%	2.00%	2.00%	2.00%	4.00%
2021-02-21	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-02-28	2.00%	52.00%	2.00%	10.00%	4.00%	2.00%	12.00%	10.00%	2.00%	4.00%
2021-03-07	94.00%	0.00%	0.00%	0.00%	6.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-03-14	0.00%	2.00%	2.00%	2.00%	2.00%	46.00%	2.00%	2.00%	2.00%	40.00%
2021-03-21	0.00%	2.00%	14.00%	2.00%	70.00%	4.00%	2.00%	2.00%	2.00%	2.00%
2021-03-28	0.00%	4.00%	8.00%	2.00%	2.00%	74.00%	2.00%	2.00%	4.00%	2.00%

#### 4.3.6 วิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization -

PSO)

การกำหนดพารามิเตอร์ของ วิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization - PSO) มีรายละเอียดแสดงในตารางที่ 4.11

**ตารางที่ 4.11** พารามิเตอร์ของ Particle Swarm Optimization

ชื่อพารามิเตอร์	ความหมาย	ค่า
Objective function	ฟังก์ชันวัตถุประสงค์	Sharpe ratio
Risk free rate	อัตราดอกเบี้ยปราศจากความเสี่ง	2%
n_particles	จำนวนอนุภาค	100
n_iterations	จำนวนรอบในการเรียนรู้	100

จากตารางที่ 4.11 นั้นนำไปสู่การหาค่าอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรก โดยมีรายละเอียดแสดงในตารางที่ 4.12

**ตารางที่ 4.12** ตัวอย่างอัตราส่วนในการลงทุนหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรกสำหรับแต่ละสัปดาห์ ในช่วงปี 2021 (เมื่อจัดพอร์ตแบบ Particle Swarm Optimization)

Date	weights_1	weights_2	weights_3	weights_4	weights_5	weights_6	weights_7	weights_8	weights_9	weights_10
2021-01-17	42.85%	0.00%	0.00%	57.15%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2021-01-24	10.14%	0.00%	0.00%	0.00%	0.00%	89.86%	0.00%	0.00%	0.00%	0.00%
2021-01-31	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%
2021-02-07	7.88%	7.77%	15.63%	2.18%	7.11%	23.51%	4.34%	18.17%	9.42%	4.01%
2021-02-14	20.14%	19.23%	0.65%	11.15%	0.00%	16.61%	19.59%	6.14%	1.11%	5.37%
2021-02-21	19.45%	17.15%	8.12%	0.32%	13.43%	2.98%	9.18%	2.17%	9.71%	17.49%
2021-02-28	3.90%	20.55%	23.07%	7.70%	13.70%	0.71%	5.58%	6.96%	11.53%	6.29%
2021-03-07	17.26%	3.89%	13.23%	9.44%	17.91%	4.45%	0.61%	3.45%	20.65%	9.10%
2021-03-14	3.78%	18.88%	15.35%	0.00%	1.33%	24.65%	6.43%	11.76%	12.53%	5.28%
2021-03-21	10.36%	10.40%	16.82%	5.27%	18.54%	6.56%	6.40%	19.31%	4.01%	2.32%
2021-03-28	5.74%	0.00%	0.00%	0.00%	0.00%	47.10%	0.00%	47.10%	0.00%	0.06%

#### 4.4 ผลลัพธ์และการอภิปราย (Results and Discussion)

การวัดประสิทธิภาพด้วยการจำลองตลาดหลักทรัพย์ตามการซื้อขายรายสัปดาห์ กำหนดซื้อที่ราคาปิดของวันทำการสุดท้ายของสัปดาห์ก่อนหน้าและขายที่ราคาปิดของวันสุดท้ายของสัปดาห์ เมื่อจำลองในตลาดหลักทรัพย์ไทย สัปดาห์แรก และการลงทุนในสัปดาห์ต่อไป จะเพิ่มขึ้นหรือลดลงขึ้นอยู่กับผลการลงทุนของสัปดาห์ก่อน โดยแบบจำลองการสุ่มป่าไม้ (Random Forest) กับข้อมูลชุดที่ (7) Keyword & Indicators & MACD จะคัดเลือกหลักทรัพย์ที่นำลงทุนสูงสุด 10 อันดับแรก นอกจากนี้ ได้แบ่งการซื้อขายในแต่ละปีและทุกปีที่เริ่มต้นด้วยเงิน 10,000 บาทในปี 2021 และอัตราส่วนในการลงทุนหลักทรัพย์จะถูกปรับตามแต่ละสัปดาห์ด้วยแต่ละ Algorithm โดยประสิทธิภาพถูกกำหนดโดยเปอร์เซ็นต์ผลตอบแทนต่อปี (ผลตอบแทนรวมในแต่ละสัปดาห์เป็นเวลา 1 ปี) และ Sharpe Ratio ผลลัพธ์ของการจัดพอร์ตลงทุนโดย Equity Investment, SLSQP, Risk Parity Optimization, Genetic Algorithm, Ant Colony Optimization และ Particle Swarm Optimization ถูกนำมาเปรียบเทียบในรูปที่ 4.13 ซึ่งให้รายละเอียดเกี่ยวกับประสิทธิภาพการจัดพอร์ตหลักทรัพย์ที่เหมาะสมและผลตอบแทนรวมประจำปี 2021

ตารางที่ 4.13 ประสิทธิภาพการจัดพอร์ตหลักทรัพย์ที่เหมาะสมและผลตอบแทนรวมประจำปี 2021

Algorithm	Annualized Return	Annualized Return (%)	Risk free rate	Standard Deviation	Sharpe Ratio
Equity Investment	113K	1132%	2%	2%	483
SLSQP	196K	1960%	2%	8%	232
Risk Parity	130K	1300%	2%	6%	224
Optimization					
Genetic Algorithm	115K	1149%	2%	2%	466
Ant Colony	178K	1777%	2%	7%	251
Optimization					
Particle Swarm	101K	1008%	2%	4%	267
Optimization					

Equity Investment และ Genetic Algorithm เป็นวิธีที่ได้ Sharpe Ratio สูงสุดอยู่ที่ 483 และ 466 สูงมากกว่า Algorithm อื่น ๆ กว่า 2 เท่า โดยค่า Sharpe Ratio ยิ่งสูงหมายความว่าผลตอบแทนสูงแต่ความผันผวนต่ำ ซึ่งสอดคล้องกับค่า Standard Deviation ที่อยู่ที่ 2% ทั้งคู่ และต่ำกว่า Algorithm อื่น ๆ 2-4 เท่า

SLSQP และ Ant Colony Optimization เป็นวิธีที่ได้ผลตอบแทนสูงสุดอยู่ที่ 196,000 บาท และ 178,000 บาท ตามลำดับ จากเงินลงทุน 10,000 บาท ซึ่งเป็นผลตอบแทน 19 และ 18 เท่าของเงินทุน และค่า Sharpe Ratio อยู่ที่ 232 และ 251 ตามลำดับ โดยที่ค่า Sharpe Ratio ของทั้งสอง Algorithm ต่ำ เมื่อเทียบกับ Equity Investment และ Genetic Algorithm สาเหตุนี้มาจากความผันผวนของผลตอบแทนที่สูงอยู่ที่ 8% และ 7% ตามลำดับ

Particle Swarm Optimization เป็นวิธีที่ได้ค่า Sharpe ratio เป็นอันดับ 3 ของ algorithm ทั้งหมดในการทดลองนี้อยู่ที่ 267 โดยที่ผลตอบแทนอยู่ที่ 101,000 บาท หรือ 10 เท่าของเงินลงทุน แต่มีความผันผวนของผลตอบแทนอยู่ที่ 4% ซึ่งมากกว่า Equity Investment และ Genetic Algorithm อยู่ 2 เท่า

Risk Parity Optimization เป็นวิธีที่ได้ผลตอบแทนเป็นอันดับ 3 ของ Algorithm ทั้งหมดในการทดลองนี้อยู่ที่ 130,000 บาท หรือ 13 เท่าของเงินทุน และมีค่า Sharpe ratio อยู่ที่ 224 ซึ่งต่ำที่สุดเมื่อเทียบกับ algorithm อื่น ๆ ในการทดลองนี้ สาเหตุมาจากความผันผวนของผลตอบแทนอยู่ที่ 6%

จากข้อมูลข้างต้น สามารถจัดอันดับตาม Sharpe ratio และ ผลตอบแทน ได้ดังนี้ จัดอันดับด้วยค่า Sharpe ratio: Equity Investment, Genetic Algorithm, Particle Swarm Optimization, Ant Colony Optimization, SLSQP และ Risk Parity Optimization และจัดอันดับด้วยผลตอบแทน:

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SLSQP, Ant Colony Optimization, Risk Parity Optimization, Equity Investment, Genetic Algorithm และ Particle Swarm Optimization

จากการจำลองการจัดพอร์ตหลักทรัพย์ลงทุนในปี 2021 นั้น ผู้วิจัยแนะนำ 4 Algorithm หลัก ๆ คือ Equity Investment และ Genetic Algorithm สำหรับการลงทุนที่เน้นความเสี่ยงต่ำ และ SLSQP และ Ant Colony Optimization สำหรับการลงทุนที่ยอมรับความเสี่ยงได้สูง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## บทที่ 5

### สรุปและอภิปรายผล

บทนี้จะกล่าวถึงสรุปและอภิปรายผลวิจัยในครั้งนี้ โดยเนื้อหาในบทนี้ประกอบด้วยหัวข้อหลักคือ สรุปและอภิปรายผล (Conclusions) และ ข้อจำกัดและการทำงานในอนาคต (Limitation and Future Work)

#### 5.1 สรุปและอภิปรายผล (Conclusions)

เพื่อขยายและจัดการกับความท้าทายที่พบในงานวิจัยที่มีอยู่เกี่ยวกับการจัดพอร์ตการลงทุนด้วยการเลือกหลักทรัพย์ บทความนี้มีวัตถุประสงค์เพื่อ (1) ศึกษาและระบุปัจจัยที่เกี่ยวข้องในการคัดเลือกหลักทรัพย์สำหรับพอร์ตการลงทุน (2) พัฒนาแนวทางการจัดพอร์ตการลงทุนสำหรับนักลงทุนรายย่อย (3) ประเมินความสามารถในการคาดการณ์ของแบบจำลองการคาดการณ์การเคลื่อนไหวของหลักทรัพย์โดยใช้คำสำคัญที่ใกล้เคียงจากชุดข้อมูลทั้งเจ็ด [(1) ตัวชี้วัดเทคนิค (technical indicators), (2) ตัวชี้วัด Moving Average Convergence Divergence (MACD indicators), (3) คำค้นใน Google Trends (Keyword), (4) Indicators & MACD, (5) Keyword & Indicators, (6) Keyword & MACD, และ (7) Keyword & Indicators & MACD)] (4) เปรียบเทียบประสิทธิภาพของแบบจำลองระหว่าง Logistic Regression, Random Forest, และ Extreme Gradient Boosting (XGBoost) และ (5) เปรียบเทียบประสิทธิภาพการหาอัตราส่วนที่เหมาะสมต่อจัดพอร์ตที่เหมาะสม (Portfolio Optimization) ระหว่าง การลงทุนแบบเท่าเทียม, Sequential Least Squares Programming algorithm (SLSQP), Risk Parity Optimization (RPO), Genetic Algorithm (GA), Ant Colony Optimization (ACO), และ Particle Swarm Optimization (PSO) ด้วยการใช้การเรียนรู้ของเครื่อง (Machine Learning) และตัวชี้วัดเทคนิค (technical indicators) กับคำค้นใน Google Trends (Keyword) แบบจำลองการคาดการณ์หลักทรัพย์ และการหาอัตราส่วนที่เหมาะสมต่อจัดพอร์ต ที่เสนอจึงถูกสร้างขึ้น นำประโยชน์ของการเรียนรู้ของเครื่อง (Machine Learning) มาสู่การเลือกหลักทรัพย์ และอัตราส่วนที่เหมาะสม งานวิจัยนี้สามารถสรุปได้หลายประการ

ประการแรก ความแม่นยำในการทำนายของแบบจำลอง Logistic Regression, Random Forest และ XGBoost ได้รับการประเมินเบื้องต้นโดยใช้ชุดข้อมูลเจ็ดชุดที่กล่าวถึงข้างต้นและเปรียบเทียบกันทั้งสามแบบจำลองมีค่า ROC-AUC สูงสุดเท่ากันในชุดข้อมูล Keyword & Indicators & MACD สามารถใช้ชุดข้อมูลร่วมกันเพื่อทำความเข้าใจแนวโน้มการค้นหาทางอินเทอร์เน็ตและอนุกรมเวลาทางการเงินได้ดียิ่งขึ้น ซึ่งจะนำไปสู่การวิเคราะห์หลักทรัพย์ที่แม่นยำยิ่งขึ้น

ประการที่สอง วิธีการเรียนรู้ของเครื่อง (Machine Learning) สามารถใช้เป็นแบบจำลองเปรียบเทียบเพื่อประเมินประสิทธิภาพการทำนายด้วยพารามิเตอร์หลัก: ROC curves, การทดสอบเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การลงทุนย้อนหลังประจำปี และ การทดสอบย้อนหลังวิกฤตการเงิน สำหรับ ROC curves ไม่มีความแตกต่างที่ชัดเจนระหว่างประสิทธิภาพของ Machine Learning ทั้ง 3 แบบจำลอง ในขณะที่ประสิทธิภาพของ XGBoost ดีที่สุดในเปอร์เซ็นต์ผลตอบแทนต่อปี (Annualized Return) และ Random Forest มีเปอร์เซ็นต์การเลือกหลักทรัพย์ที่ประสบความสำเร็จ (Success Prediction Rate) สูงสุด โดย XGBoost มีผลตอบแทนเฉลี่ยสูงสุดต่อปี ในสถานการณ์วิกฤต Random Forests มีอัตราความสำเร็จเฉลี่ยสูงสุด โดยเฉพาะอย่างยิ่งในช่วงวิกฤต COVID-19 XGBoost ได้รับการพิสูจน์แล้วว่า มีเปอร์เซ็นต์การเลือกหลักทรัพย์ที่ประสบความสำเร็จโดยเฉลี่ยสูงสุด โดยสรุป การศึกษานี้สรุปได้ว่าการประเมินการทดสอบการลงทุนย้อนหลังประจำปี มีความเหมาะสมในการเปรียบเทียบประสิทธิภาพของวิธี Machine Learning มากกว่า ROC curves การทดสอบการลงทุนย้อนหลังประจำปี สามารถรวมเหตุการณ์จริงในตลาดหลักทรัพย์เข้ากับแบบจำลองในการเปรียบเทียบได้

ประการที่สาม การหาอัตราส่วนที่เหมาะสมในการจัดพอร์ตเป็นกระบวนการที่ซับซ้อนและขึ้นอยู่กับหลายปัจจัย เช่น วัตถุประสงค์การลงทุน ระดับความเสี่ยงที่ยอมรับ ราคาของหลักทรัพย์ และข้อมูลตลาดอื่น ๆ โดยการหาอัตราส่วนที่เหมาะสมต่อการจัดพอร์ตเป็นกระบวนการที่ต้องดำเนินการด้วยความรอบคอบและการวิเคราะห์อย่างถี่ถ้วน โดย Equity Investment และ Genetic Algorithm แสดงให้เห็นว่าอัตราส่วนในการลงทุนของหลักทรัพย์นั้นกระจายตัวเป็นอย่างมาก ซึ่งการกระจายการลงทุนนี้ส่งผลให้ความผันผวนในการลงทุนสูง ในขณะที่ SLSQP และ Ant Colony Optimization แสดงให้เห็นว่าอัตราส่วนในการลงทุนของหลักทรัพย์นั้นเน้นไปที่หลักทรัพย์ใดหลักทรัพย์หนึ่ง ซึ่งส่งผลให้ความผันผวนในการลงทุนต่ำ และผลตอบแทนก็สูงตามกัน ดังนั้นความสำเร็จของการหาอัตราส่วนที่เหมาะสมขึ้นอยู่กับอัลกอริทึมที่ใช้และการปรับตัวแปรต้องพิจารณาอย่างรอบคอบเพื่อให้ได้ผลลัพธ์ที่เหมาะสมสำหรับการจัดพอร์ต

สุดท้าย ผลลัพธ์จากขั้นตอนการทำนายพบว่า Random Forest และ XGBoost เกือบจะเหมือนกันแต่ก็ยังคงแตกต่างกัน ในช่วงที่ตลาดผันผวนปานกลาง โมเดล Random Forest มีประสิทธิภาพดีกว่า XGBoost ในแง่ของอัตราความสำเร็จเฉลี่ยและผลตอบแทนเฉลี่ยต่อปี อย่างไรก็ตาม XGBoost ทำงานได้ดีขึ้นอย่างมากในช่วงที่ตลาดมีความผันผวนสูงมาก เช่น การระบาดของ COVID-19 อัลกอริทึมมีความสามารถในการระบุหลักทรัพย์ที่มีศักยภาพในการเติบโตของราคาสูงในอนาคต แบบจำลองนี้เหนือกว่าแบบจำลอง Random Forest ทั้งสองด้าน (อัตราความสำเร็จเฉลี่ยและผลตอบแทนเฉลี่ยต่อปี)

## 5.2 ข้อจำกัดและการทำงานในอนาคต (Limitation and Future Work)

แม้ว่าการศึกษานี้จะให้ข้อมูลเชิงลึกที่เป็นประโยชน์ แต่ก็มีข้อจำกัดบางประการ ชั้นแรก เราจะวิเคราะห์เฉพาะข้อมูลหลักทรัพย์ไทยเท่านั้น ดังนั้น ผลจากการศึกษานี้อาจไม่ถูกนำไปสรุปเป็นภาพรวมสำหรับเศรษฐกิจและประเทศอื่นๆ ประการที่สอง ใช้ชุดข้อมูล Keyword & Indicators & MACD เพื่อพิจารณาว่า keyword ใดเหมาะสมกับกรอบเวลาที่กำหนด ดังนั้นจึงจำเป็นต้องมี

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ การเผยแพร่โดยไม่ได้รับอนุญาตเป็นการผิดกฎหมาย  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะต้องอัปเดต keyword ใหม่และตรวจสอบความสัมพันธ์ระหว่างการเคลื่อนไหวของหลักทรัพย์และ keyword ในช่วงเวลาต่างๆ ประการที่สาม การศึกษานี้ใช้เฉพาะค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน ( $r$ ) เพื่อหาความสัมพันธ์ระหว่างหลักทรัพย์และ keyword ในอนาคต จะเป็นการดีกว่าที่จะขยายการศึกษา ทำการวิเคราะห์ความเชื่อมั่น (Sentiment Analysis) เพื่อกำหนดความสัมพันธ์เฉพาะของ keyword ที่มีอิทธิพลต่อหลักทรัพย์แต่ละตัว ประการที่สี่ การปรับพารามิเตอร์ที่เหมาะสมสำหรับ Algorithm Optimization เพื่อหาอัตราส่วนในการลงทุนที่ดีที่สุดเพื่อสร้างพอร์ตโฟลิโอที่ดีที่สุด พอร์ตโฟลิโออาจทำงานได้ดีขึ้นและดีขึ้นโดยรวม ประการสุดท้าย เนื่องจากการศึกษานี้ไม่ได้ตรวจสอบจุดตัดการซื้อ/ขายที่แนะนำสำหรับหลักทรัพย์แต่ละตัวในแต่ละสัปดาห์ จุดตัดการซื้อ/ขายที่แนะนำจึงสามารถเพิ่มผลกำไรได้อย่างมีประสิทธิภาพมากขึ้น

สรุปแล้ว การวิจัยในอนาคตควรใช้การวิเคราะห์ความเชื่อมั่น (Sentiment Analysis) สำหรับ keyword รายหลักทรัพย์ ตรวจสอบการปรับอัลกอริทึมให้เหมาะสมสำหรับการเลือกพอร์ตโฟลิโอ และควรคำนึงถึงจุดตัดซื้อ/ขายที่แนะนำในแต่ละสัปดาห์



## เอกสารอ้างอิง

- Alfonso, G., & Ramirez, D. R. (2020). A Nonlinear Technical Indicator Selection Approach for Stock Markets. Application to the Chinese Stock Market. *Mathematics*, 8(8). <https://doi.org/10.3390/math8081301>
- Ananthakumar, U., & Sarkar, R. (2017). *Application of Logistic Regression in Assessing Stock Performances 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*,
- Anghel, G. D. I. (2015). Stock Market Efficiency and the MACD. Evidence from Countries around the World. *Procedia Economics and Finance*, 32, 1414 - 1431. [https://doi.org/10.1016/s2212-5671\(15\)01518-x](https://doi.org/10.1016/s2212-5671(15)01518-x)
- Antonio Agudelo Aguirre, A., Alfredo Rojas Medina, R., & Darío Duque Méndez, N. (2020). Machine learning applied in the stock market through the Moving Average Convergence Divergence (MACD) indicator. *Investment Management and Financial Innovations*, 17(4), 44-60. [https://doi.org/10.21511/imfi.17\(4\).2020.05](https://doi.org/10.21511/imfi.17(4).2020.05)
- Atkins, A., Niranjana, M., & Gerding, E. (2018). Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science*, 4(2), 120-137. <https://doi.org/10.1016/j.jfds.2018.02.002>
- Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552 - 567. <https://doi.org/10.1016/j.najef.2018.06.013>
- Bhargavi, R. S. G., Anitha, R. (2017). Relative Strength Index for Developing Effective Trading Strategies in Constructing Optimal Portfolio. *International Journal of Applied Engineering Research* 12, 8926-8936.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5 - 32. <https://doi.org/10.1023/a:1010933404324>
- Bustos, O., & Pomares-Quimbaya, A. (2020). Stock market movement forecast: A Systematic review. *Expert Systems with Applications*, 156. <https://doi.org/10.1016/j.eswa.2020.113464>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง (ต่อ)

- Chaves, D., Hsu, J., Li, F., & Shakernia, O. (2011). Risk Parity Portfolio vs. Other Asset Allocation Heuristic Portfolios. *The Journal of Investing*, 20(1), 108-118. <https://doi.org/10.3905/joi.2011.20.1.108>
- Chen, Q., Zhang, W., & Lou, Y. (2020). Forecasting Stock Prices Using a Hybrid Deep Learning Model Integrating Attention Mechanism, Multi-Layer Perceptron, and Bidirectional Long-Short Term Memory Neural Network. *IEEE Access*, 8, 117365-117376. <https://doi.org/10.1109/access.2020.3004284>
- Chen, R. R., Huang, W. K., & Yeh, S. K. (2021). Particle swarm optimization approach to portfolio construction. *Intelligent Systems in Accounting, Finance and Management*, 28(3), 182-194. <https://doi.org/10.1002/isaf.1498>
- Dash, R., & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science*, 2(1), 42-57. <https://doi.org/10.1016/j.jfds.2016.03.002>
- Elizabeth, R. D., David, M. D., & Daniel, L. C.-P. (1988). Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *International Biometric Society*, 44, 837-845.
- Gambeta, V., & Kwon, R. (2020). Risk Return Trade-Off in Relaxed Risk Parity Portfolio Optimization. *Journal of Risk and Financial Management*, 13(10). <https://doi.org/10.3390/jrfm13100237>
- Gatta, F., Iorio, C., Chiaro, D., Giampaolo, F., & Cuomo, S. (2023). Statistical arbitrage in the stock markets by the means of multiple time horizons clustering. *Neural Computing and Applications*, 35(16), 11713-11731. <https://doi.org/10.1007/s00521-023-08313-6>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42. <https://doi.org/10.1007/s10994-006-6226-1>
- Ghatasheh, N. (2014). Business Analytics using Random Forest Trees for Credit Risk Prediction: A Comparison Study. *International Journal of Advanced Science and Technology*, 72, 19-30. <https://doi.org/10.14257/ijast.2014.72.02>
- Gunjan, A., & Bhattacharyya, S. (2022). A brief review of portfolio optimization techniques. *Artificial Intelligence Review*, 56(5), 3847-3886. <https://doi.org/10.1007/s10462-022-10273-7>

เอกสารนี้เป็นเอกสารที่เผยแพร่โดยวารสารวิชาการเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง (ต่อ)

- Huang, M. Y., Rojas, R. R., & Convery, P. D. (2019). Forecasting stock market movements using Google Trend searches. *Empirical Economics*, 59(6), 2821-2839. <https://doi.org/10.1007/s00181-019-01725-1>
- Le, T.-T.-H., Oktian, Y. E., & Kim, H. (2022). XGBoost for Imbalanced Multiclass Classification-Based Industrial Internet of Things Intrusion Detection Systems. *Sustainability*, 14(14). <https://doi.org/10.3390/su14148707>
- Li, Q., Wang, T., Li, P., Liu, L., Gong, Q., & Chen, Y. (2014). The effect of news and public mood on stock movements. *Information Sciences*, 278, 826-840. <https://doi.org/10.1016/j.ins.2014.03.096>
- Mitra Thakur, G. S., Bhattacharyya, R., & Sarkar, S. (2018). Stock portfolio selection using Dempster-Shafer evidence theory. *Journal of King Saud University - Computer and Information Sciences*, 30(2), 223-235. <https://doi.org/10.1016/j.jksuci.2016.07.001>
- Nishimura, Y., & Acoba, J. D. (2022). Impact of Breast Cancer Awareness Month on Public Interest in the United States between 2012 and 2021: A Google Trends Analysis. *Cancers (Basel)*, 14(10). <https://doi.org/10.3390/cancers14102534>
- Papadamou, S., Koulis, A., Kyriakopoulos, C., & Fassas, A. P. (2022). Cannabis Stocks Returns: The Role of Liquidity and Investors' Attention via Google Metrics. *International Journal of Financial Studies*, 10(1). <https://doi.org/10.3390/ijfs10010007>
- Perry, M. B. (2011). The Weighted Moving Average Technique. In *Wiley Encyclopedia of Operations Research and Management Science*. <https://doi.org/10.1002/9780470400531.eorms0964>
- Poutachidou, N., & Papadamou, S. (2021). The Effect of Quantitative Easing through Google Metrics on US Stock Indices. *International Journal of Financial Studies*, 9(4). <https://doi.org/10.3390/ijfs9040056>
- Praekhaow, P. (2010). *Determination of Trading Points using the Moving Average Methods* International Conference for a Sustainable Greater Mekong Subregion, Bangkok, Thailand.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง (ต่อ)

- Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Sci Rep*, 3, 1684. <https://doi.org/10.1038/srep01684>
- Robles, V., Bielza, C., Larranaga, P., González, S., & Ohno-Machado, L. (2008). Optimizing logistic regression coefficients for discrimination and calibration using estimation of distribution algorithms. *Top*, 16, 345–366. <https://doi.org/10.1007/s11750-008-0054-3>
- Sadorsky, P. (2021). A Random Forests Approach to Predicting Clean Energy Stock Prices. *Journal of Risk and Financial Management*, 14(2). <https://doi.org/10.3390/jrfm14020048>
- Smart, H. (2018). หุ้นสุดเหวี่ยงปิดลบ 8 จุด - ต่างชาติขาย 1.8 พันล้าน. <https://hoonsmart.com/archives/39374>
- Sycinska-Dziarnowska, M., Szyszka-Sommerfeld, L., Wozniak, K., Lindauer, S. J., & Spagnuolo, G. (2022). Predicting Interest in Orthodontic Aligners: A Google Trends Data Analysis. *Int J Environ Res Public Health*, 19(5). <https://doi.org/10.3390/ijerph19053105>
- Teixeira, L. A., & de Oliveira, A. L. I. (2010). A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Systems with Applications*, 37(10), 6885–6890. <https://doi.org/10.1016/j.eswa.2010.03.033>
- Today, P. (2018). หุ้นไทย 27 ธ.ค. 61 ปิดลบ 8.56 จุด. <https://www.posttoday.com/finance-stock/stock/575197>
- Tudor, C. (2022). The Impact of the COVID-19 Pandemic on the Global Web and Video Conferencing SaaS Market. *Electronics*, 11(16). <https://doi.org/10.3390/electronics11162633>
- Vaidya, R. (2018). Stochastic and Momentum Analysis of Nepalese Stock Market. *The Journal of Nepalese Business Studies*, XI.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## เอกสารอ้างอิง (ต่อ)

- Wang, K., Wan, J., Li, G., & Sun, H. (2022). A Hybrid Algorithm-Level Ensemble Model for Imbalanced Credit Default Prediction in the Energy Industry. *Energies*, 15(14). <https://doi.org/10.3390/en15145206>
- Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, 55(4), 919-926. <https://doi.org/10.1016/j.dss.2012.12.028>



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทความตีพิมพ์วารสารวิชาการนานาชาติ International Journal of Financial Studies  
เรื่อง Stock Movement Prediction Using Machine Learning Based on Technical Indicators  
and Google Trend Searches in Thailand



International Journal of  
*Financial Studies*



Article

## Stock Movement Prediction Using Machine Learning Based on Technical Indicators and Google Trend Searches in Thailand

Kittipob Saetia<sup>1</sup> and Jiraphat Yokrattanasak<sup>2,3,\*</sup>

<sup>1</sup> KMITL Digital Analytics and Intelligence Center, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

<sup>2</sup> Department of Mathematics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

<sup>3</sup> Business Innovation and Investment Laboratory (B2I-Lab), School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

\* Correspondence: jiraphat.yo@kmitl.ac.th

**Abstract:** Machine learning for stock market prediction has recently been popular for identifying stock selection strategies and providing market insights. In this study, we adopted machine learning algorithms to analyze technical indicators, and Google Trends search terms based on the Thai stock market. This study uses three datasets, which are technical indicators, Google Trends search terms, and a combination of the two. The objectives were to study and identify the factors in stock selection, develop and evaluate portfolio selection models using keyword proxies from the three datasets mentioned, and compare the performance of the selected algorithms. In the prediction process, we discovered that the combination of technical indicators and Google Trends search terms while applying Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost) exhibited the highest ROC curves. For success prediction rate and annualized return, Random Forest and XGBoost were almost similar but still different. While XGBoost performs well during a period of market critical conditions (COVID-19), Random Forest performs marginally better than XGBoost during normal market conditions in terms of average success rate.

**Citation:** Saetia, Kittipob, and Jiraphat Yokrattanasak. 2023. Stock Movement Prediction Using Machine Learning Based on Technical Indicators and Google Trend Searches in Thailand.

*International Journal of Financial Studies* 11: 5. <https://doi.org/10.3390/ijfs11010005>

Academic Editor: Muhammad Ali Nasir

Received: 24 October 2022

Revised: 15 December 2022

Accepted: 19 December 2022

Published: 22 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** stocks; Google Trends; machine learning

### 1. Introduction

An inefficient market is observed when the price of a security at a given point in time does not represent the whole worth of the asset in the stock market. This might be due to investor decision-making behaviors that cannot be predicted using existing data and indicators. It would be beneficial if analysts and investors could grasp the link between variables that cause different market phenomena such as trade volume, online search trends, and investor behaviors. The study and comprehension of these variables will be essential for modeling to more accurately predict the trading behaviors of investors and market movements.

In the context of the investment market, irrational investor behavior can take various forms. One of the examples is investors who are exiting the market out of fear of a potential market downturn. Future asset price fluctuations and the outlook for the market as a whole are influenced by behavior patterns. To date, prior studies have investigated issues related to modeling and stock prediction by understanding human behaviors and their impacts on the stock markets. Economic and technical indicators, market information, headline news, and online search terms were incorporated into these research studies (Antonio Agudelo Aguirre et al. 2020; Atkins et al. 2018; Chen et al. 2020; Dash and Dash 2016; Huang et al. 2019; Li et al. 2014; Papadamou et al. 2022; Poutachidou and Papadamou 2021; Teixeira and de Oliveira 2010; Yu et al. 2013). Most of the mentioned

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

indicators were proven to be helpful in predicting the total price of shares or securities. However, relatively few efforts are made to forecast the price per security or price per share, especially in low- and middle-income country settings.

In a related study, Teixeira and de Oliveira (2010) combined technical indicators, such as Simple Moving Average (SMA), Relative Strength Index (RSI), Stochastic Oscillator K, Stochastic Oscillator D, and Bollinger bands together with the K-Nearest Neighbor Algorithm (KNN) to create a prediction model. This study concluded that the method outperformed the buy-and-hold strategy for 12 out of the 15 Brazil equities taken into account in the studies when we compare the outcomes to those. As a result, this study agreed that it is possible to use this method to forecast actual short-term stock trends.

Another study by Dash and Dash (2016) in India using technical indicators, but combined with machine learning, was tested in the S&P 500 and BSE SENSEX contexts. Among the many modeling methodologies, the Artificial Neural Network (ANN) generated the largest returns, with rises of 34.42% and 42.58% in the BSE SENSEX and S&P 500, respectively.

Scholars also tried to leverage unstructured qualitative data to predict stock prices and returns in capital markets. Li et al. (2014) adapted the headline news, blogs, and financial discussion board coupled with the Support Vector Regression (SVR). The study concluded that the impact of the media on companies varied by the nature of the information. The fundamentals of an article, such as company-specific news, can enhance investors' knowledge. Public sentiment can cause volatility in the mood of investors and interferes with investor decision-making. On the other hand, Atkins et al. (2018) combined two types of information, news from Reuters USA and financial data on Yahoo Finance incorporated into the Latent Dirichlet Allocation. This study concluded that news data influence volatility forecasts better than stock market closing prices forecasts.

The review of stock market forecasting processes conducted by Bustos and Pomares-Quimbaya (2020) shed light on the overall process of forecasting stock market movements using machine learning (ML), deep learning, text mining, and clustering techniques to create an investment model. These techniques, used as investors' strategy planning and decision-making tools, tend to perform better than traditional trading strategies. Evidently, Huang et al. (2019) investigated the effectiveness of stock prediction models for the S&P 500 Index with Google Trends, Support Vector Machine (SVM), Ridge Regression, Lasso Regression, and Elastic-Net Regression. With 63.75% accuracy, the top-performing model was Ridge Regression with selection factor modifying VAR model. Papadamou et al. (2022) examined the connections between investor sentiment, as represented by Google Trends, and stock market return, volatility, and liquidity in the setting of cannabis industry stocks. Extending the three-factor Fama-French model, returns on cannabis stocks and liquidity are statistically positively correlated. Augmented investor interest increases returns. On the U.S. stock exchange, another study by Poutachidou and Papadamou (2021) concluded that there is a positive correlation between returns on stock market indices and increased investor attention on the U.S. Quantitative Easing policy, as measured by Google Metrics, which suggests that investor attention on QE seems to reduce volatility in the stock market and increase stock returns.

With that, there are gaps in academic research, as only few papers investigate the effectiveness of the application of machine learning approaches to not only forecast stock performance but also automate portfolio selection for higher returns and better investment decisions. Moreover, selecting the "right" securities to invest requires specific skill sets and experience as a number of criteria must be examined and a large amount of data must be analyzed while selecting securities. Hence, our focus is to leverage the use of machine learning techniques, namely Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost), in selecting investment securities based on three datasets: (1) technical indicators, (2) Google Trends search terms, and (3) the combination of the aforementioned. We sought to (1) study and identify the relevant factors in selecting securities for the investment portfolio, (2) develop a portfolio formation approach for

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

individual investors, (3) evaluate the predictive capabilities of stock movement forecasting models using keywords proxied from the three datasets above, and (4) compare the model performance between Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost).

The selected technical indicators indicate three characteristics. First is trend direction, namely (1) Simple Moving Average (SMA) used under normal market conditions, (2) Weighted Moving Average (WMA) used under unusual volatile market conditions, (3) Exponential Moving Average (EMA) suitable for analysis that need to prioritize the latest data. Second is reliability, namely (1) Moving Average Convergence Divergence (MACD), which provides information on both trend and momentum of the stock prices, and (2) Relative Strength Index (RSI), used to determine the probability of a trend reversal. Third is momentum, represented by Stochastic Oscillators K and D to compare the closing price with a range of prices for a given period of time (Bhargavi et al. 2017; Bustos and Pomares-Quimbaya 2020; Perry 2011; Praekhaow 2010; Vaidya 2018).

The Google Trends search dataset determines the popularity of keywords online such as product names, personal names, or possibly website names. Google Trends is created for users who rely on or benefit from trends, whether they are marketers or owners of online stores or even those who wish to start their own blog or vlog but are unsure of what to sell or how to determine the most frequently searched terms each day or month. Additionally, it aids in the development of marketing strategies and the gradual development of a business plan. In finance, Huang et al. (2019) discovered that the numerous directional movements of the S&P 500 vary in search volume based on the individual phrases searched for and, by extension, the sense of the word.

For forecasting models, the three machine learning models are applied in this study: Logistic Regression, Random Forest, and Extreme Gradient Boost (XGBoost). These approaches are used as comparison models to evaluate the prediction accuracy between the mentioned three datasets. The key measures used to compare the performance of the models include Receiver Operating Characteristic curves (ROC curves) and annualized and crisis backtest financial evaluations.

This study consists of (1) an introduction to explain the background and rationale of this study, (2) the methodology of research discussing the process and relevant theories used in this study, (3) the results and discussion to show the reflection from forecasting model performance and evaluations, and ends with (4) conclusions, limitations, and future workflow.

## 2. Methodology

Based on a machine learning algorithm, we constructed a generic common system, which aims to compare the effectiveness of stock forecasting models proxied by technical indicators, keywords from Google Trends, and a combination of both. According to Figure 1, this system is divided into four modules: dataset, data preprocessing, modelling, and evaluation. The details are as follows.

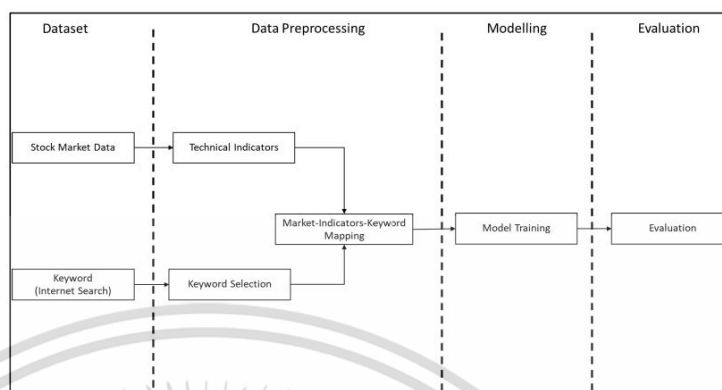


Figure 1. Generic common system components diagram.

### 2.1. Study Design

We would like to clarify that this study examines the price movement of the top 100 indices of the Stock Exchange of Thailand (SET100). WHA corporation is the example index in which we randomly pick from SET100.

For the stock price prediction, this study focuses on (1) the stock market data or technical indicators and (2) related keywords based on internet search.

As described above, the datasets come from two sources in the Thai stock market from 1 January 2017 to 31 December 2021:

1. Yahoo Finance (stock market data)—weekly stock prices and trading volumes for stocks in SET100.
2. Google Trends (keywords; internet search)—weekly search data for 484 specific terms which are commonly mentioned by the public on the internet. The selection of internet search terms are also based on the research by Preis et al. (2013) and the websites of Finnomena, SET, Krungsri, Encyclopedia, and stock2morrow.

Then, we processed the mentioned data to create technical indicators and ranked the top 20 keywords per each index which have both positive and negative correlations. The results from this data processing were used in the model creation.

### 2.2. Dataset

The data used in this study are structured data types, which are organized in a tabular format with definable columns. This layout enables easy access to information and simple or uncomplicated searching, without requiring additional work. We collected the data from two sources: the stock market and specific keyword searches.

#### 2.2.1. Stock Market Data

Regarding stock market data, this information is public information available for download from the Stock Exchange's website, which includes the following:

- The opening price is the first price of any listed stock at the start of a trading day.
- The high and low values represent the stock's highest and lowest prices on that particular day. Generally, traders utilize these statistics to determine the volatility of a stock.
- The closing price is the price of the stock at the close of the trading day.
- The adjusted close price is regarded as the genuine price of that stock, as it reflects the stock's worth after dividends are distributed.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Stock prices are influenced by many factors and are often considered as one of the indicators to study market behaviors. As a result, using securities prices to study technical indicators improves the efficiency through which we can comprehend market behavior. The Stock Exchange of Thailand (SET100) data are collected weekly from Yahoo Finance. Examples of the daily historical stock prices and volumes collected from SET100 are shown in Table 1 for “WHA Corporation Public Company Limited” (WHA) stock.

**Table 1.** Examples of securities information of WHA Corporation Public Company Limited (WHA).

Date	High	Low	Open	Close	Volume	Adj Close	Symbol
4 January 2017	3	2.96	2.96	2.98	25,413,400	2.378223	WHA
5 January 2017	3.04	2.98	3	3.02	82,795,000	2.410144	WHA
6 January 2017	3.04	3	3.02	3.02	48,678,800	2.410144	WHA
9 January 2017	3.1	3.02	3.02	3.08	1.52 × 10 <sup>8</sup>	2.458028	WHA
10 January 2017	3.12	3.04	3.1	3.04	90,063,300	2.426106	WHA
11 January 2017	3.1	3.06	3.08	3.06	74,300,900	2.442067	WHA
12 January 2017	3.2	3.1	3.1	3.16	3.55 × 10 <sup>8</sup>	2.521873	WHA
16 January 2017	3.26	3.18	3.26	3.24	90,585,100	2.585719	WHA
17 January 2017	3.26	3.18	3.24	3.2	83,648,500	2.553796	WHA
18 January 2017	3.24	3.18	3.22	3.2	53,264,600	2.553796	WHA

### 2.2.2. Keywords (Internet Search)

The term “keywords” in this research refers to the terms used by investors to communicate in the securities investing sector, as well as the terms used by internet users to search for information on securities. This study consists of 484 keywords. There are eight categories according to research (Preis et al. 2013) and sources such as the Fintnomena, SET, Krungsri, Encyclopedia website, and stock2morrow; these are basic investment terms, industry groups, stock names, trading methods, global search, popular words, idiom, and yearly search terms. The examples and definitions of keywords are exhibited in Table 2.

**Table 2.** Keyword examples.

Keywords		
Type	Keyword	Definition
Basic Investment Term	P/E	Price-to-Earnings (P/E) Ratio: The ratio for valuing a company that measures its current share price relative to its earnings per share (EPS)
Basic Investment Term	P/BV	Price to Book Value Ratio (P/BV): The market’s valuation of a company relative to its book value
Basic Investment Term	EPS	Earnings Per Share (EPS): Calculated as a company’s profit divided by the outstanding shares of its common stock
Industry Group	Agribusiness	Agribusiness
Industry Group	Food and Beverage	Food and beverage
Industry Group	Insurance	Insurance
Stock Name	ADVANC	Advanced Info Service PCL (ADVANC.BKK)
Stock Name	BBL	Bangkok Bank PCL (BBL.BKK)
Stock Name	CPN	Central Pattana PCL (CPN.BKK)
Trading Method	Technical	A trading strategy that primarily relies on technical indicators

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Trading Method	Day Trade	A trading strategy that is often informed by technical analysis of price movements and requires a high degree of self-discipline and objectivity
Trading Method	Swing Trade	A trading strategy that focuses on taking smaller gains in short term trends and cutting losses quicker
Global Search	Economics	Economics
Global Search	Politics	Politics
Global Search	Conflict	Conflict
Popular Word	กอง (Kong)	Mutual fund
Popular Word	ปอ (Pop)	Broker
Popular Word	ฝรั่ง (Rang)	Foreign investor
Idiom	กำไร (Lamyai)	Profit
Idiom	ซื้อควาย (Sue Khwai)	Buy stock(s) right before the stock's price goes down
Idiom	ขายหมู (Khai Mu)	Sale stock(s) right before the stock's price moves up
Yearly Search Term	คนละครึ่ง (Khon La Khrueng)	Thailand's government COVID-19 financial relief campaign
Yearly Search Term	โควิด-19 (COVID-19)	An infectious disease caused by the SARS-CoV-2 virus
Yearly Search Term	ชิมช้อปใช้ (Chim Chop Chai)	Thailand's government COVID-19 financial relief campaign

Google Trends is a website of Google that was created to help determine the popularity of keywords online such as product names, personal names, or possibly website names. Users can view the popularity of these keywords by place, whether global, national, or provincial, and can also view daily popular trends (Huang et al. 2019; Nishimura and Acoba 2022; Papadamou et al. 2022; Poutachidou and Papadamou 2021; Sycinska-Dziarnowska et al. 2022; Tudor 2022). It is designed to even examine the keyword's popularity over time to determine potential development prospects. Google Trends is created for users who rely on or benefit from trends, whether they are marketers or owners of online stores or even those who wish to start their own blog or vlog but are unsure of what to sell, or how to determine the most frequently searched terms each day or month. Additionally, it aids in the development of marketing strategies and the gradual development of a business plan. In finance, Huang et al. (2019) discovered that the numerous directional movements of the S&P 500 vary in search volume based on the individual phrases searched for and, by extension, the sense of the word.

Google Trends normalizes its search data in order to make word comparisons easier. The following process normalizes the search results according to the time and location of the search query. Each data point is split by the total number of search locations and the time period over which the relative popularity is being compared. Otherwise, the location with the greatest amount of traffic is always ranked first. The result number is then scaled from 0 to 100 as a percentage of all searches across topics, areas exhibiting the same search interest for a term (Huang et al. 2019; Nishimura and Acoba 2022; Papadamou et al. 2022;

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Poutachidou and Papadamou 2021; Sycinska-Dziarnowska et al. 2022; Tudor 2022). Table 3 illustrates the examples of popular words that the public used to search on Google ranked in Google Trends during 2017–2021 in Thailand.

**Table 3.** Google Trends search terms for each word in Thailand.

Date	Debt	Color	Stocks	Restaurant	Portfolio	Inflation	Housing	Dow Jones
1 January 2017	3	21	17	60	30	17	11	3
8 January 2017	22	31	32	51	29	24	7	11
15 January 2017	9	39	25	43	36	15	9	6
22 January 2017	37	41	42	43	38	24	32	9
29 January 2017	12	29	20	50	41	38	17	8
5 February 2017	34	38	21	51	25	55	30	8

### 2.3. Data Preprocessing

The data processing is divided into two sections: (1) technical indicators and (2) keyword selection. The technical indicators in the stock market refer to mathematical formulas based on the changes and direction of the market, such as prices and volumes. For keyword selection, the authors chose to study the search terms used by investors in the stock market, as well as the terms used by internet users to search for relevant information on securities.

#### 2.3.1. Technical Indicators

In general, technical indicators are studied and plotted on charts to assist in providing directional information about a security's price. When the indices are used on a graph, they are displayed as lines, and the values are displayed between those lines alongside the price in order to determine the next likely direction of movement.

Technical indicators can be used to infer the behavior or trends of a time series and can be used to forecast the price of securities (Alfonso and Ramirez 2020; Anghel 2015; Basak et al. 2019; Bhargavi et al. 2017; Bustos and Pomares-Quimbaya 2020; Dash and Dash 2016; Perry 2011; Praekhaow 2010; Vaidya 2018). This study demonstrates how techniques can perform a summary rather than the full time series of securities prices, which simplifies machine learning. The technical indicators used in the forecasting model are as shown in Table 4.

**Table 4.** Technical indicators and formulas.

Technical Indicators	Formula
Simple Moving Average (SMA)	$\frac{C_t + C_{t-1} + C_{t-2} + \dots + C_{t-n+1}}{n}$
Weighted Moving Average (WMA)	$\frac{(n)(C_t) + (n-1)(C_{t-1}) + \dots + C_{t-n+1}}{(n) + (n-1) + \dots + 1}$
Exponential Moving Average (EMA)	$\alpha C_t + (1-\alpha)C_{t-1} + (1-\alpha)^2 C_{t-2} + \dots + (1-\alpha)^{t-n+2} C_{t-n+2} + (1-\alpha)^{t-n+1} EMA_{t-n+1}$
Moving Average Convergence Divergence (MACD)	$EMA12_t - EMA26_t$
Relative Strength Index (RSI)	$100 - \frac{100}{1 + \left( \frac{\sum_{i=0}^{n-1} UP_{t-i}/n}{\sum_{i=0}^{n-1} DW_{t-i}/n} \right)}$
Stochastic Oscillator K (K)	$\frac{C_t - L_{t-n}}{H_{t-n} - L_{t-n}} \times 100$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Stochastic Oscillator D (D)	$\frac{\sum_{t=0}^{n-1} K_t}{n}$
-----------------------------	----------------------------------

Note:  $C_t$  is the closing price,  $L_t$  is the low price,  $H_t$  is the high price at time  $t$ ,  $\alpha$  is a smoothing factor,  $L_t$  and  $H_t$  mean lowest low and highest high in the last  $t$  days, respectively.  $UP_t$  means upward price change, while  $DW_t$  is the downward price change at time  $t$ .

**The Moving Average (MA)** is a technical indicator that indicates the average trend over a specified time period by smoothing and filtering various abnormal signals and moving averages formed by using average closing prices over a specified time period.

#### Simple Moving Average (SMA)

This mean is appropriate in normal-volatility situations in a time series to help data analysts better interpret the patterns and contexts of stock market conditions (Perry 2011).

#### Weighted Moving Average (WMA)

This mean is appropriate in high-volatility situations in a time series to help data analysts better interpret the patterns and contexts of historical trends (Perry 2011).

#### Exponential Moving Average (EMA)

This average prioritizes the latest data. Therefore, it responds to price changes faster than a simple moving average (Bustos and Pomares-Quimbaya 2020; Praekhaow 2010).

#### Moving Average Convergence Divergence (MACD)

MACD indicates a trend that has the idea of drawing two moving averages at the same time and then analyzing the nature of the two moving averages. The special interesting point of MACD is that it is an indicator that can provide two pieces of information: the simultaneous view is the trend direction of the stock price (trend) and the momentum of the share price (momentum). The standard MACD is the 12-period EMA subtracted by the 26-period EMA (Alfonso and Ramirez 2020; Anghel 2015; Bustos and Pomares-Quimbaya 2020; Dash and Dash 2016).

#### The Relative Strength Index (RSI)

The Relative Strength Index (RSI) is a technical indicator that indicates whether an asset is strong or weak in relation to its recent closing price. Additionally, it is used to determine the probability of a trend reversal (Bhargavi et al. 2017).

#### The Stochastic Oscillator

The Stochastic Oscillator is a momentum technical indicator that is used to compare closing prices over a specified time period to a range of prices. This oscillator is extremely sensitive to market price changes. The indicator's volatility can be smoothed somewhat by changing the time interval being measured. The most frequently used stochastic oscillators are Stochastic Oscillator K (K) and Stochastic Oscillator D (D) (Vaidya 2018). The K line compares the lowest low and the highest high of a given period to define a price range, then displays the last closing price as a percentage of this range. The D line is a moving average of K.

#### 2.3.2. Keyword Selection

This is the selection of factors that affect each security. The factors mentioned are internet search keywords and the Pearson Correlation Coefficient ( $r$ ) in Equation (1). The top 20 keywords, both positive and negative correlations, are selected for each security. The top 20 keywords, both positive and negative, of WHA Corporation Public Company Limited are shown in Table 5.

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \quad (1)$$

where

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$r_{xy}$  : Pearson Correlation Coefficient between variables x and y.

$\Sigma x$  : The sum of the measured data from variable x.

$\Sigma y$  : The sum of the measured data from variable y.

$\Sigma xy$  : The sum of product of the variables x and y.

$\Sigma x^2$  : The sum of the squares of variable x.

$\Sigma y^2$  : The sum of the squares of variable y.

n : Number of data.

Table 5. Top 20 keywords of WHA Corporation Public Company Limited.

Symbol	Positive	Correlation	Negative	Correlation
WHA	wha	0.2322	restaurant	-0.2028
WHA	sta	0.2262	major	-0.1999
WHA	tcap	0.2161	holiday	-0.1818
WHA	sawad	0.2066	fun	-0.1771
WHA	settrade	0.1885	แขก (kaek)	-0.1693
WHA	dow jones	0.1878	hybride	-0.1671
WHA	banpu	0.1851	food& beverage	-0.1642
WHA	tisco	0.1840	bec	-0.1567
WHA	toa	0.1832	ประชัย (bpra chai)	-0.1511
WHA	aot	0.1824	forex	-0.1430
WHA	bcp	0.1822	short selling	-0.1411
WHA	hmpro	0.1766	water	-0.1362
WHA	ปอด (bpot)	0.1753	mbk	-0.1340
WHA	lh	0.1724	thani	-0.1339
WHA	amata	0.1712	ปะหนัง (bor nang)	-0.1331
WHA	bdms	0.1689	travel	-0.1249
WHA	cpf	0.1679	top	-0.1224
WHA	bb1	0.1672	ตกรถ (dtok rot)	-0.1218
WHA	บริการ (bor-ri-gaan)	0.1645	markets	-0.1214
WHA	cpn	0.1615	bts	-0.1198

Note: แขก (kaek): Indorama Ventures PCL (IVL) stock, ประชัย (bpra chai): TPI Polene PCL (TPIPL) stock, ปะหนัง (bor nang): Interhides PCL (IHL) stock, ตกรถ (dtok rot): about to buy the stock(s) when the price is considerably low but then it goes up, ปอด (bpot): stock portfolio, บริการ (bor-ri-gaan): industry group service.

#### 2.4. Modeling

Modeling is the relationship of data in various forms. Modeling in this study includes the Classification Model, which is a supervised learning model to create a model that must have variables used to measure the target as a starting point to learn. The goals of the classification will be grouped or discrete such as yes/no, A/B/C, etc. Therefore, in evaluating the results obtained from the classification model, accuracy can be measured, e.g., with a confusion matrix.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

The following types of modeling were used to predict the selection of securities in this study. We chose Logistic Regression, Random Forest, and XGBoost because (1) the three models were commonly used and cited in many relevant studies (Ananthakumar and Sarkar 2017; Basak et al. 2019; Geurts et al. 2006; Ghatasheh 2014; Sadorsky 2021) and (2) they can manage the imbalance data as also discussed (Le et al. 2022; Wang et al. 2022).

#### 2.4.1. Logistic Regression

Logistic Regression is classified as a supervised learning model that is used to predict the probability of occurrence of an event of interest from a dataset of appropriate independent variables and algorithms based on mathematical equations. If breaking a Logistic Regression follows the type of independent variables, it can be divided into two types: Simple Logistic Regression, which has only one independent variable, and Multiple Logistic Regression, which has many independent variables. It is widely used in applications such as predicting the likelihood of a customer incurring bad debt to a banking company or predicting the likelihood of a customer migrating to a telephone network. This technique has also been used in applied marketing to discover market segmentation in order to maximize the chance of campaign offers for each market category (Ananthakumar and Sarkar 2017; Robles et al. 2008). Figure 2 depicts the approach for developing a Logistic Regression model, with further information provided below.

- Use a dataset to create a simple linear regression or multiple linear regression depending on the independent variables used in the type of work performed.
- Bring the regression equation to the Sigmoid function to adjust the value to be in the range 0–1 because the regression equation can have values greater than 1 or less than 0. It should be between 0 and 1 only, so this function has been implemented.
- By passing the sigmoid function, the probability of the event of interest is obtained.

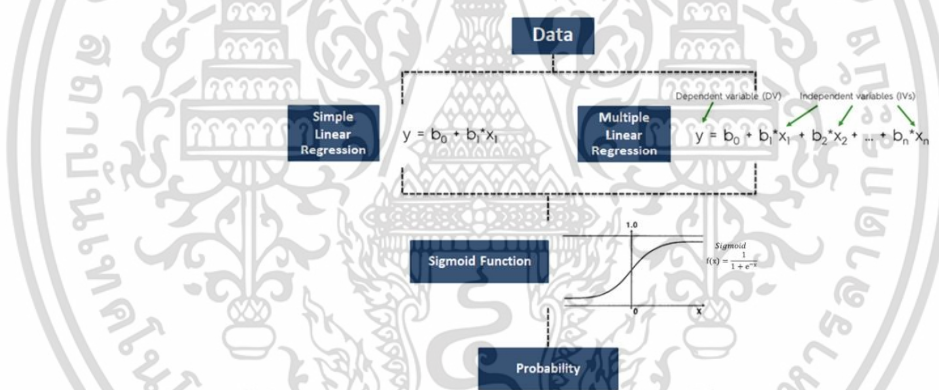


Figure 2. Logistic Regression modeling process diagram.

#### 2.4.2. Random Forest

Random Forest is a tree-based model composed of several decision trees. It is a predictive model based on the Wisdom of the Crowd technique. This is also referred to as the Ensemble Technique and it is a type of Bootstrap Aggregation that is used to sample data and then build a predictive model. In Random Forest, the random sampling with replacement method is used so that the size of the new dataset is the same as the original dataset. This modeling provides high accuracy and low variance, which is better than a single decision tree but takes more complex calculations (Basak et al. 2019; Sadorsky 2021). There is one study that applied this modeling to predict credit risk prediction and compared it

with more than 10 other models (Ghatasheh 2014). The study found that Random Forest generated the highest accuracy, sensitivity, and F-Measure. The main modeling processes are shown in Figure 3 and are described below:

- In Random Forest,  $n$  number of random records are taken from the dataset having  $k$  number of records (Bootstrapped Dataset).
- Individual decision trees are constructed for each sample.
- Each decision tree will generate an output.
- The final output is considered based on majority voting.

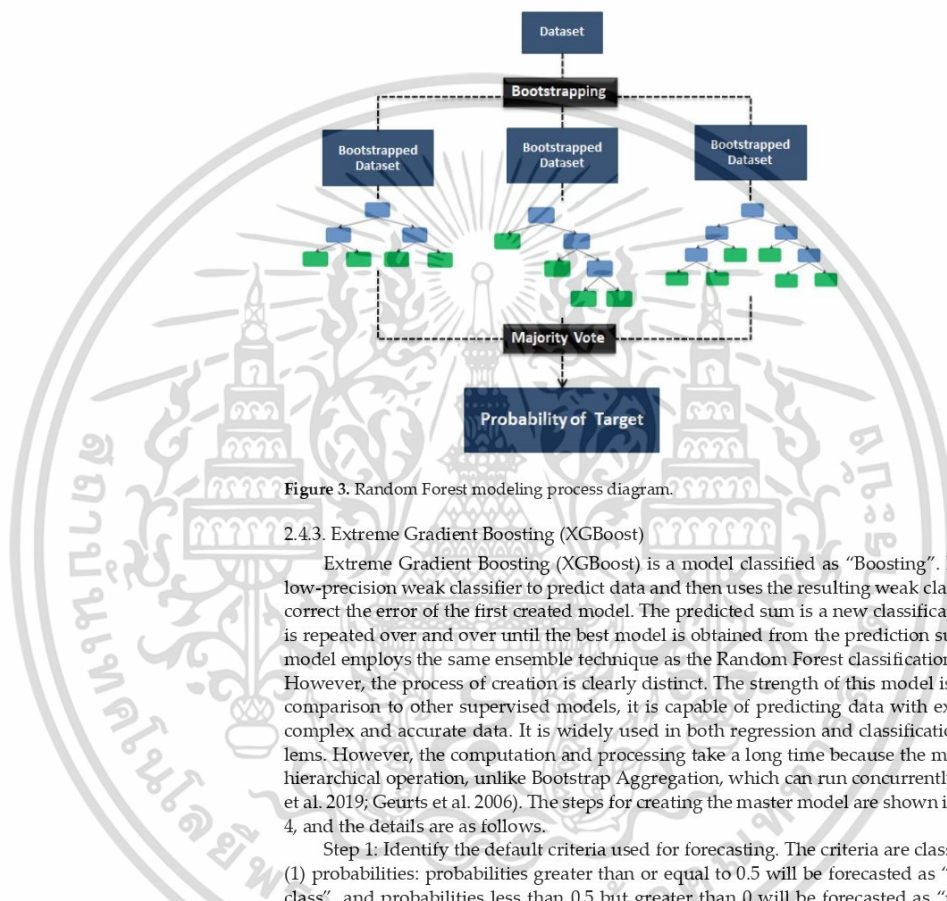


Figure 3. Random Forest modeling process diagram.

#### 2.4.3. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a model classified as “Boosting”. It uses a low-precision weak classifier to predict data and then uses the resulting weak classifier to correct the error of the first created model. The predicted sum is a new classification that is repeated over and over until the best model is obtained from the prediction sum. This model employs the same ensemble technique as the Random Forest classification model. However, the process of creation is clearly distinct. The strength of this model is that, in comparison to other supervised models, it is capable of predicting data with extremely complex and accurate data. It is widely used in both regression and classification problems. However, the computation and processing take a long time because the model is a hierarchical operation, unlike Bootstrap Aggregation, which can run concurrently (Basak et al. 2019; Geurts et al. 2006). The steps for creating the master model are shown in Figure 4, and the details are as follows.

Step 1: Identify the default criteria used for forecasting. The criteria are classified by (1) probabilities: probabilities greater than or equal to 0.5 will be forecasted as “positive class”, and probabilities less than 0.5 but greater than 0 will be forecasted as “negative class”; and by (2) types of targets.

Step 2: Calculate the residual from the actual conversion of the forecasted variable by setting the positive class to 1 and negative class to 0. Then, subtract the true value of the variable from the criteria obtained in Step 1.

Step 3: Create the decision tree from all independent variables; this decision tree is used to predict the residual from Step 2.

Step 4: Calculate the residual sum of all leaf nodes from the formula below.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{Residual}_{\text{New}} = \frac{\sum \text{Residual}}{\sum (\text{Previous Probability} \times (1 - \text{Previous Probability})) + \lambda}$$

Step 5: Combine the criteria obtained in Step 1 through the log odds function with the threshold values created in Step 3 via the decision tree and utilize the learning rates as a parameter to help tune the decision tree so that it does not overfit the data. When the learning rates are between 0 and 1, we obtain a weak classifier.

Step 6: Recalculate the residual value from the weak classifier in Step 5 with all the data calculated. Then, take the value of each data point through the Sigmoid function, which will derive the probability of each of the data.

Step 7: Repeat Step 2 until the residual value is unchanged or reaches the specified number of decision trees.

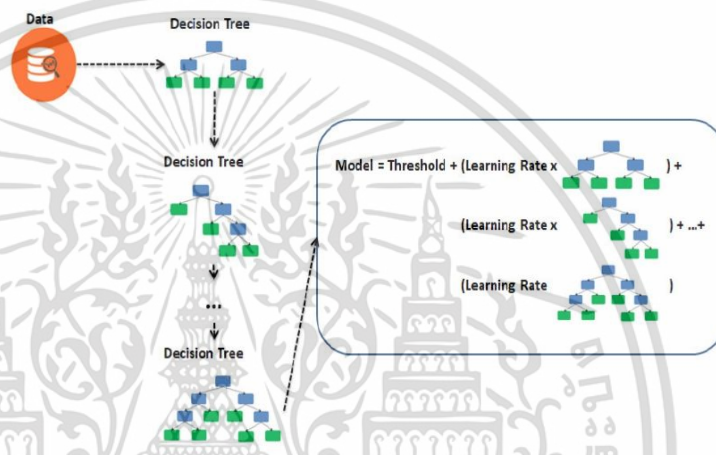


Figure 4. Extreme Gradient Boosting (XGBoost) modeling process diagram.

### 2.5. Evaluation

Evaluation is an important tool for measuring the effectiveness of this research. The tools used in the evaluation must be consistent with the measurements of performance. In this study, we opted to use the Receiver Operating Characteristic curves as follows.

#### Receiver Operating Characteristic curves (ROC curves)

ROC curves are frequently used to assess the effectiveness of binary classification algorithms. They display the performance of a classifier graphically, rather than as a single value, as most other metrics do (Ananthakumar and Sarkar 2017; Basak et al. 2019; Bustos and Pomares-Quimbaya 2020; Huang et al. 2019; Trifonova et al. 2014). The dependent variable (Y) in this case is a qualitative variable, which can be divided into two cases:  $Y = 1$  when the interesting event or test result is positive, and  $Y = 0$  when the incident is ignored, or the test result is negative.

The cut-off point is the point used to classify events into interesting and uninteresting events. Comparisons between the forecast values and observations can be divided into 4 cases as shown in Table 6.

Table 6. Confusion matrix.

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positive (TP)	False Positive (FP)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Predicted Negative (0)	False Negative (FN)	True Negative (TN)
------------------------	---------------------	--------------------

TP, TN, FP, and FN in the table is represented by the frequency values:

- A True Positive (TP) response is when what is predicted matches what is actually happening. In the case of a prediction that is “true”, what happened is “true”.
- A True Negative answer (TN) is when what the prediction matches what happened. In the event that the prediction is “not true”, what happened is “not true”.
- A False Positive (FP) is a prediction that does not match what happened, that is, a prediction is “true”, but what happens is “not true”.
- False Negative (FN) is a prediction that does not match what actually happened, which is a prediction that something is “not true”, but what happens is “true”.

The True Positive Rate (TPR) is a measure of when what is predicted matches what is actually happening with all true events (True Positive (TP) + False Negative (FN)), as shown in Equation (2).

$$TPR = \frac{TP}{TP+FN} \quad (2)$$

The False Positive Rate (FPR) is a measure of when what is predicted does not match what is happening with all false events (True Negative (TN) + False Positive (FP)), as shown in Equation (3).

$$FPR = \frac{FP}{TN+FP} \quad (3)$$

Precision is the weighted average of positive and negative precision, while recall is the weighted average of positive recall, as shown in Equations (4)–(7).

$$\text{Precision}_{\text{positive}} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Precision}_{\text{negative}} = \frac{TN}{TN+FN} \quad (5)$$

$$\text{Recall}_{\text{positive}} = \frac{TP}{TP+FN} \quad (6)$$

$$\text{Recall}_{\text{negative}} = \frac{TN}{TN+FP} \quad (7)$$

Accuracy and F1-score are estimated using Equations (8)–(9).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

$$F1\text{-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

ROC curves are plotted between TPR (Sensitivity) and FPR (1-Specificity) for all possible cut-off values. The y-axis represents TPR, while the x-axis represents FPR. At each cut-off value, it separates the forecast results into two groups: the event where  $P(Y = 1)$  is greater than or equal to the cut-off point, and the event where  $P(Y = 1)$  is less than the cut-off point, where  $P(Y = 1)$  is the probability that the interesting event or test result is positive. The area under the ROC curve is an index used to indicate the accuracy or reliability of a model. The model with the greatest area under the ROC curve is the most accurate model (Elizabeth et al. 1988)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3. Results and Discussion

#### 3.1. Performance on Dataset

The comparative data are grouped into three categories: (1) Technical Indicators, including SMA, WMA, EMA, MACD, RSI, K, and D; (2) Google Trends keywords; and (3) Technical Indicators and Keywords. Additionally, machine learning data are divided into three sets: Train and Test, which is 1 January 2017 through 30 May 2021 with an 80:20 ratio, and Unknown, which is 1 June 2021 to 31 December 2021. The unknown dataset is used for the purpose of knowing the performance on unseen data.

Table 7 shows that during the data test process, the combination of Keywords and Indicators outperforms other datasets in terms of accuracy, precision, recall, F1-score, and ROC-AUC score. For unknown data, Random Forest and XGBoost models with Keywords and Indicators also outperform other datasets. Specifically, Keywords and Indicators coupled with Logistic Regression perform better when comparing F1-score and ROC-AUC scores.

Table 7. Performance on dataset.

Model	Type Dataset	Test						Unknown			
		Accuracy	Precision	Recall	F1-Score	AUC	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	Indicators	0.9699	0.9589	0.9699	0.9587	0.8457	0.9818	0.9739	0.9818	0.9745	0.8251
	Keywords	0.9727	0.9664	0.9727	0.9653	0.8999	0.9840	0.9843	0.9840	0.9779	0.8880
	Keywords and Indicators	0.9813	0.9794	0.9813	0.9790	0.9767	0.9834	0.9827	0.9834	0.9830	0.9733
Random Forest	Indicators	0.9694	0.9555	0.9694	0.9553	0.9348	0.9818	0.9638	0.9818	0.9727	0.9324
	Keywords	0.9727	0.9734	0.9727	0.9619	0.9120	0.9814	0.9717	0.9814	0.9737	0.9045
	Keywords and Indicators	0.9734	0.9708	0.9734	0.9642	0.9645	0.9831	0.9833	0.9831	0.9758	0.9623
XGBoost	Indicators	0.9732	0.9674	0.9732	0.9681	0.9577	0.9818	0.9767	0.9818	0.9782	0.9499
	Keywords	0.9743	0.9694	0.9743	0.9682	0.9339	0.9850	0.9840	0.9850	0.9802	0.9456
	Keywords and Indicators	0.9823	0.9810	0.9823	0.9796	0.9787	0.9879	0.9871	0.9879	0.9855	0.9810

Overall, when comparing the predictive performance of the three datasets, the Keywords and Indicators dataset has the highest F1-score and ROC-AUC score across the three models. We imply that the datasets can reveal online search trends and financial time series that can lead to stock analysis. Therefore, Table 8 uses the Keywords and Indicators dataset for further analysis.

Table 8. Performance for crisis.

% Success	Model			Win Model
	Logistic Regression	Random Forest	XGBoost	
Crisis				
QE	55.00%	85.00%	83.75%	Random Forest
Controversy	51.11%	78.89%	65.56%	Random Forest
Foreign Investors	45.00%	71.67%	63.33%	Random Forest
MSCI	57.78%	86.67%	82.22%	Random Forest
COVID-19 (1)	51.54%	73.85%	79.23%	XGBoost
Protestation	65.56%	94.44%	90.00%	Random Forest
COVID-19 (2)	42.86%	90.00%	91.43%	XGBoost
COVID-19 (3)	40.00%	82.22%	84.44%	XGBoost
COVID-19 (4)	56.25%	92.50%	93.75%	XGBoost
COVID-19 (5)	51.11%	90.00%	90.00%	Random Forest
AVERAGE	51.62%	84.52%	82.37%	Random Forest

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

### 3.2. Performance

We conducted a stock market simulation based on weekly trading. For the strategy, we determined buying at the closing price of the previous week's final trading day and selling at the week's final day's closing price. When simulating in the Thai stock market, the first week and the investment in the following weeks will increase or decrease depending on the investment results of the previous week, where the model selects the top 10 most probable stocks. Additionally, we split trading for each year and every year beginning with 10,000 Thai baht from 2017 to 2021. Performance is determined by the percentage of successful stock picks (percentage of right weekly stock picks) and the percent annualized return (total return each week over 1 year). The results of Logistic Regression, Random Forest, and XGBoost are compared in Figures 5–8, which detail the performance of stock selection and total return on an annual basis from 2017 to 2021.

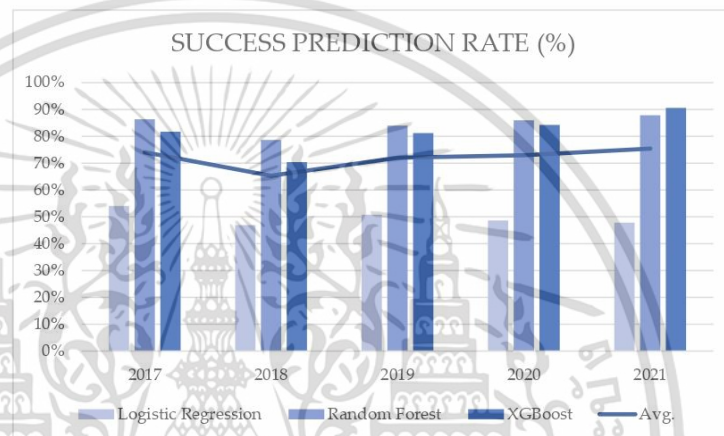


Figure 5. Stock selection performance (success prediction rate).

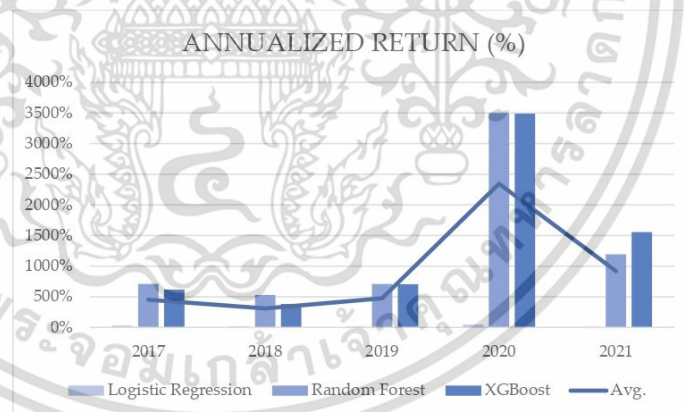


Figure 6. Stock selection performance (annualized return).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

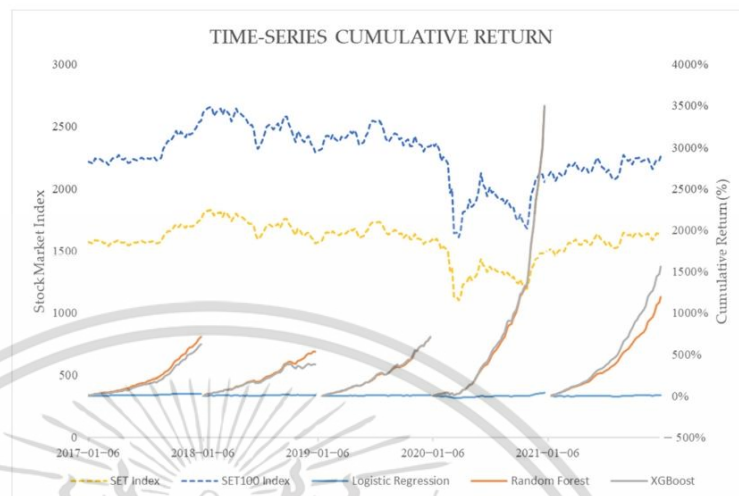


Figure 7. Time series cumulative return.

The following are the annual results from 2017 to 2021 (Figures 5 and 6). The average success rate for the Logistic Regression model is 49.65% and the average annualized return is 23.25%, with the highest success prediction rate in 2017 being 54.04% and the highest annualized return in 2020 being 45.45%. The Random Forest model has an average success rate of 84.56% and an average annualized return of 1328.43%, with the best success prediction rate in 2021 being 87.84% and the best annualized return in 2020 being 3500.66%. The XGBoost model has an average success rate of 81.63% and an average annualized return of 1350.38%, with a success rate of 90.59% for the best year 2021 and 3489.08% for the best year 2020. From Figure 7, the Logistic Regression, Random Forest, and XGBoost portfolios show dominance over the benchmarks SET index and SET100 index in the portfolio across the simulation periods (2017 to 2021). The dominance relationship becomes more significant starting during the COVID-19 outbreak in 2020. In conclusion, performance comparisons with the annualized backtest financial evaluation reveal the distinction between the three models. Clearly, the use of Random Forest has the highest average percentage of success, whereas XGBoost has the highest average annualized return.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	2017	2018	2019	2020	2021
Jan		Quantitative Easing (QE)			Covid-19 (3)
Feb				Covid-19 (1)	
Mar					
Apr					
May			MSCI		
Jun		Controversy			
Jul					
Aug					Covid-19 (4)
Sep					
Oct				Protestation	
Nov					Covid-19 (5)
Dec		Foreign Investors		Covid-19 (2)	

Figure 8. Crisis in Thailand during the years 2017–2021.

In 2020, it was the year with the most return because the Thai stock market was extremely volatile due to various COVID-19 events and crises. Every year, there are events that have an impact on the stock market, both positive and negative. Thus, Table 8 demonstrates the model's performance for several crises in Thailand.

### 3.3. Performance for Crisis

During the various crises in Thailand, the performance of models differs between periods and models. From 2017 to 2021, according to Figure 8, there are six types of crises: Quantitative Easing (QE), Controversy, Foreign Investors, MSCI, Protestation, and COVID-19. For COVID-19 crises, it can be classified into five states of COVID-19, as each state is pretty unique in terms of model performances in the Thai stock market. Hence, we explain the states of COVID-19 outbreaks and draw attention to their impacts during each phase. Table 8 displays the performance of stock selection during various times of crisis.

#### 3.3.1. Quantitative Easing (QE)

Quantitative Easing (QE) is a form of monetary policy borne by each country's central bank. In January 2018, the Stock Exchange of Thailand index closed at its all-time high of 1838.96 points, which was a significant factor in the decline of gold prices (Chai et al. 2021). This was due to the impact of QE measures in the United States and Japan and the reduction in interest rates by the European Central Bank (ECB), which encouraged investors to invest in stock markets in search of higher yields. The results of Logistic Regression, Random Forest, and XGBoost from January to February 2018, as shown in Table 8, revealed successful stock pick rates of 55.0%, 85.0%, and 83.75%, respectively.

#### 3.3.2. Controversy

The U.S. and China's trade retaliation is a global trade war. Particularly, trade countermeasures between the United States and China have caused the SET index to decline for the majority of the week, although there are factors supporting energy stocks due to the rise in oil prices on the global market. However, this is insufficient to offset the pressure exerted by foreign investors' net selling of Thai stocks from the beginning to the middle of the week. However, the Thai stock market still closed in the red in contrast to regional stock markets. In June and July 2018, the successful stock selection rates for Logistic

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Regression, Random Forest, and XGBoost were 51.11%, 78.89%, and 65.55%, respectively, as shown in Table 8.

### 3.3.3. Foreign Investors

The SET market decreased dramatically to 1548.37 points, the lowest in 19 months on 27 December 2018. The major reason is that the foreign investors sold a net 1832.50 million baht (Smart 2018; Today 2018). As shown in Table 8, the results of Logistic Regression, Random Forest, and XGBoost for the period December 2018 to January 2019 showed stock selection performance rates of 45.00%, 71.67%, and 63.33%, respectively.

### 3.3.4. MSCI

MSCI Index is Morgan Stanley Capital International's benchmark. It was developed as a metric for foreign investors to use in selecting stocks and returns. On 28 May 2019, the Thai stock market closed at 1632.04 points, a gain of 7.20 points, with a trading value of 204,855.67 million baht, a record high. This was due to the reweighting of the MSCI Thailand index, which boosted the weight of some Thai equities in the index's investment. According to Table 8, the stock selection rates predicted by Logistic Regression, Random Forest, and XGBoost for the period of May–June 2019 were 57.78%, 86.67%, and 82.22%, respectively.

### 3.3.5. COVID-19 (1)

The COVID-19 crisis has had various effects on the global economy, with the lockdown measures having the most economic impact. When there are lockdowns, it makes it harder for people and businesses to buy things in public places such as department stores. It began affecting the Thailand Stock Exchange on 26 February 2020. The Thai Stock Exchange fell sharply by 72.69 points, or -5.05%, in a single day, closing at 1366.41 points. The SET ended trading on 9 March at 1255.94 points, which was down 108.63 points or -7.96%. The SET for 12 March closed at 1114.91 points, a decrease of 134.98 points, or a loss of 10.80%. The stock market closed on 16 March at 1046.08 points, with a loss of 82.83 points, or 7.34%. As indicated in Table 8, the performance of stock selection from February to April 2020 using Logistic Regression, Random Forest, and XGBoost were 51.54%, 73.85%, and 79.22%, respectively.

### 3.3.6. Protestation

Protests in Bangkok, Thailand, on 30 October 2020 led to the declaration of a state of emergency in the capital city. Because of this, the SET index closed the trading day at 1214.95 points. Using Logistic Regression, Random Forest, and XGBoost, the success prediction rates for October and November 2020 were as follows (Table 8): 65.56%, 94.44%, and 90%, respectively.

### 3.3.7. COVID-19 (2)

The COVID-19 pandemic began in the province of Samut Sakhon in Thailand at the same time that the country instituted quarantine measures to stop the virus from spreading further. With investors concerned about the closure of Samut Sakhon, the Stock Exchange of Thailand Index decreased by more than 5% on 21 December 2020. The market closed with a loss of 80.60 points, or -5.44%, at 1401.78 points. According to Table 8, the stock selection success rates predicted by Logistic Regression, Random Forest, and XGBoost for the period December 2020 to January 2021 were 42.86%, 90%, and 91.4%, respectively.

### 3.3.8. COVID-19 (3)

As a consequence of the relaxation of lockdown measures, the Stock Exchange of Thailand index ended on 11 June 2021 at 1636.56 points, a gain of 11.29 points, the highest

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

level in 1 year, 4 months, and 18 days. However, with the easing of the lockdown measures, the epidemic became more severe, and the stock market continued to decline. According to Table 8, the performance of successful stock picks using Logistic Regression, Random Forest, and XGBoost for the period of June to July 2021 was 40%, 82.22%, and 84.44%, respectively.

### 3.3.9. COVID-19 (4)

As the number of COVID-19 cases in the country continued to exert more pressure, the Stock Exchange of Thailand closed at 1521.72 points on 6 August 2021, a fall of 5.94 points and the lowest close in five months. This number is still far over 20,000 every day. The market is well aware of the issue, yet it still poses risks. Investors are wary as a result. According to a tally compiled by the Stock Exchange of Thailand (SET) on 6 August 2021, institutional investors sold 135.90 million baht worth of shares, brokers sold 363.93 million baht worth of shares, foreign investors sold 2078.31 million baht worth of shares, and retail investors bought 2578.14 million baht worth of shares. Foreign investors' sales over the first six days of August totaled as much as 8148.43 million baht. As can be shown in Table 8, the success rates of stock picks made using Logistic Regression, Random Forest, and XGBoost during the months of August and September 2021 were 56.25%, 92.50%, and 93.75%, respectively.

### 3.3.10. COVID-19 (5)

On 28 November 2021, the Stock Exchange of Thailand dropped 37.85 points, closing at 1610.61 as a result of the Omicron strain of the 2019 coronavirus. Continuing its downward trend, the index shed another 20.92 points on 29 November 2021, ending the day at 1589.69. It dropped another 21.00 points on 30 November 2021, finishing at 1568.69. Table 8 displays that the accuracy rates of Logistic Regression (51.11%), Random Forest (90.00%), and XGBoost (90.00%) were the lowest and highest for predicting profitable stocks during November and December of 2021, respectively.

Based on the various crises that occurred in Thailand between 2017 and 2021, the crisis is divided into six categories: QE, Controversy, Foreign Investors, MSCI, COVID-19, and Protest. The overall performance of Logistic Regression is proxied by %success, which in this experiment was 51.62% on average. Random Forest has an overall accuracy rate of 84.52% (the highest performance during the financial crisis: QE, Controversy, Foreign Investors, MSCI, COVID-19 (5), and Protest). For XGBoost, the overall performance reflects 82.37% success (the highest performance during the crisis: COVID-19 (1–5)). Lastly, for backtest financial evaluation during the crisis, Random Forest ranked first in terms of %success, with 84.52%, while XGBoost ranked the first during COVID-19 outbreaks with 87.77%.

Based on the above performance, Random Forest had a higher stock selection rate than the other two models (Logistic Regression and XGBoost) because its algorithm was very stable, and it works well with both continuous and discrete variables. Even if a new data point is added to the dataset, the overall algorithm is not significantly impacted because it is unlikely that the new data will affect all trees (Breiman 2001). However, during COVID-19 when the market fell sharply, XGBoost had a higher stock selection rate. The stock market experienced a sharp decline during the COVID-19 period due to the nature of the data. It is possible that Random Forest will assign less weight to this class, but XGBoost is an excellent alternative for unbalanced datasets. Consequently, XGBoost has a higher stock selection rate, whereas Logistic Regression has a relatively low performance. Logistic Regression is optimized for discrete variables (Robles et al. 2008). Compared to Random Forest and XGBoost, the Keywords and Indicators dataset characteristics are complex, having both continuous and discrete variables data characteristics.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## 4. Conclusions

### 4.1. Conclusion

To address the challenges found in the existing literature on portfolio formation with stock selection, we aimed to (1) study and identify relevant factors in selecting securities for the investment portfolio, (2) develop a portfolio formation approach for individual investors, (3) evaluate the predictive capabilities of stock movement forecasting models using keywords proxied from the three datasets (technical indicators, Google Trends search terms, and the combination of the aforementioned), and (4) compare the model performance between Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost). With the use of the machine learning (ML) models and technical indicators combined with Google Trends search terms, the proposed prediction model was constructed, bringing the benefits of machine learning to stock selection. Several conclusions can be drawn from this research.

First, the predictive accuracy of Logistic Regression, Random Forest, and XGBoost models was initially evaluated using the three datasets mentioned above and compared. All three models have the same maximum ROC-AUC Score in the dataset, which combines technical indicators and Google Trends search terms. The combination of datasets can be applied to better understand the internet search trends as well as financial time series, which in turn can lead to more precise stock analysis.

Second, the three ML approaches were used as comparison models to assess the predictive performance with the key parameters: ROC curves, annualized backtests, and financial crisis backtests. For ROC curves, there was no clear distinction between the performances of the three ML models, while XGBoost performed the best. Backtest evaluation (annualized) made the contrast clearer, Random Forest had the highest average success rate, and XGBoost had the highest average annualized returns. In the crisis situation, Random Forest had the highest average success rate. During the COVID-19 crisis, XGBoost proved to have the highest average success rate. In summary, this study concluded that the backtest evaluation was more suitable to compare the performance of each ML method than ROC curves. The backtest evaluation was able to incorporate real events in the stock markets into the comparison model.

Finally, the results from the prediction procedure revealed that Random Forest and XGBoost were nearly identical but still different. During moderately volatile markets, the Random Forest model outperforms the XGBoost model in terms of both average success rate and average annualized return. Nevertheless, the XGBoost model performs significantly better during periods of extreme market volatility, such as the COVID-19 pandemic. The algorithm has the capability to identify stocks with a high future price growth potential. This model is superior to the Random Forest model in both aspects (average success rate and average annualized return).

### 4.2. Limitation and Future Work

Despite the fact that this study provides useful insights, it has some limitations. First, we only analyze Thai Stock data. Hence, the result from this study might not be generalizable for other economies and countries. Second, we utilized the Keywords and Indicators dataset to determine which keyword is appropriate for a specific time frame. Consequently, it is essential to update new keywords and ensure the correlation between stock movements and keywords in different time periods. Third, this study only used the Pearson Correlation Coefficient ( $r$ ) to determine the relationship between stocks and keywords. In the future, it would be better to expand the study and perform a sentiment analysis determining the specific correlation of keywords that influence individual stocks. Fourth, there is a limit on the number of shares that may be imported to the portfolio with the same amount every time. This is because this research only chose companies with the best chance of price growth for a total portfolio of 10 stocks every weekly transaction per round. In fact, the number of shares that will give the best return in each round

unnecessarily equals the number of equities in the portfolio. If optimization algorithms are used to construct an optimal portfolio, the portfolio may perform better and be better overall. Finally, since this study did not examine the recommended buy/sell cut-off points for each stock each week, a suggested buy/sell cut-off point could increase profits more effectively.

All in all, future research should apply sentiment analysis for keywords and individual stocks, examine algorithm optimization for portfolio selection, and take into account the suggested buy/sell cut-off points for each week.

**Author Contributions:** Conceptualization, K.S. and J.Y.; Methodology, K.S. and J.Y.; Software, K.S.; Validation, K.S. and J.Y.; Formal analysis, K.S. and J.Y.; Investigation, J.Y.; Resources, K.S.; Data curation, K.S.; Writing—original draft preparation, K.S. and J.Y.; Writing—review and editing, J.Y.; Visualization, K.S.; Supervision, J.Y.; Project administration, J.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to acknowledge Praewa Kulatnam for her contribution in proofreading and editing this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Alfonso, Gerardo, and Daniel R. Ramirez. 2020. A Nonlinear Technical Indicator Selection Approach for Stock Markets. Application to the Chinese Stock Market. *Mathematics* 8: 1301. <https://doi.org/10.3390/math8081301>.
- Ananthakumar, Usha, and Ratul Sarkar. 2017. Application of Logistic Regression in Assessing Stock Performances. Paper presented at the 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th International Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Orlando, FL, USA, November 6–10, pp. 1242–47.
- Anghel, Gabriel Dan I. 2015. Stock Market Efficiency and the MACD. Evidence from Countries around the World. *Procedia Economics and Finance* 32: 1414–31. [https://doi.org/10.1016/s2212-5671\(15\)01518-x](https://doi.org/10.1016/s2212-5671(15)01518-x).
- Antonio Agudelo Aguirre, Alberto, Ricardo Alfredo Rojas Medina, and Néstor Darío Duque Méndez. 2020. Machine learning applied in the stock market through the Moving Average Convergence Divergence (MACD) indicator. *Investment Management and Financial Innovations* 17: 44–60. [https://doi.org/10.21511/imfi.17\(4\).2020.05](https://doi.org/10.21511/imfi.17(4).2020.05).
- Atkins, Adam, Mahesan Niranjan, and Enrico Gerding. 2018. Financial news predicts stock market volatility better than close price. *The Journal of Finance and Data Science* 4: 120–37. <https://doi.org/10.1016/j.jfds.2018.02.002>.
- Basak, Suryoday, Saibal Kar, Snehanshu Saha, Luckyson Khaidem, and Sudeepa Roy Dey. 2019. Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance* 47: 552–67. <https://doi.org/10.1016/j.najef.2018.06.013>.
- Bhargavi, R., Srinivas Gumparathi, and R. Anith. 2017. Relative Strength Index for Developing Effective Trading Strategies in Constructing Optimal Portfolio. *International Journal of Applied Engineering Research* 12: 8926–36.
- Breiman, Leo. 2001. Random Forests. *Machine Learning* 45: 5–32. <https://doi.org/10.1023/a:1010933404324>.
- Bustos, Oscar, and Alexandra Pomares-Quimbaya. 2020. Stock market movement forecast: A Systematic review. *Expert Systems with Applications* 156: 113464. <https://doi.org/10.1016/j.eswa.2020.113464>.
- Chai, Jian, Chenyu Zhao, Yi Hu, and Zhe George Zhang. 2021. Structural analysis and forecast of gold price returns. *Journal of Management Science and Engineering* 6: 135–45. <https://doi.org/10.1016/j.jmse.2021.02.011>.
- Chen, Qian, Wenyu Zhang, and Yu Lou. 2020. Forecasting Stock Prices Using a Hybrid Deep Learning Model Integrating Attention Mechanism, Multi-Layer Perceptron, and Bidirectional Long-Short Term Memory Neural Network. *IEEE Access* 8: 117365–76. <https://doi.org/10.1109/access.2020.3004284>.
- Dash, Rajashree, and Pradipta Kishore Dash. 2016. A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science* 2: 42–57. <https://doi.org/10.1016/j.jfds.2016.03.002>.
- Elizabeth, R. DeLong, M. DeLong David, and L. Clarke-Pearson Daniel. 1988. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *International Biometric Society* 44: 837–45.
- Geurts, Pierre, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine Learning* 63: 3–42. <https://doi.org/10.1007/s10994-006-6226-1>.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Ghatasheh, Nazeeh. 2014. Business Analytics using Random Forest Trees for Credit Risk Prediction: A Comparison Study. *International Journal of Advanced Science and Technology* 72: 19–30. <https://doi.org/10.14257/ijast.2014.72.02>.
- Huang, Melody Y., Randall R. Rojas, and Patrick D. Convery. 2019. Forecasting stock market movements using Google Trend searches. *Empirical Economics* 59: 2821–39. <https://doi.org/10.1007/s00181-019-01725-1>.
- Le, Thi-Thu-Huong, Yustus Eko Oktian, and Howon Kim. 2022. XGBoost for Imbalanced Multiclass Classification-Based Industrial Internet of Things Intrusion Detection Systems. *Sustainability* 14: 8707. <https://doi.org/10.3390/su14148707>.
- Li, Qing, Tiejun Wang, Ping Li, Ling Liu, Qixu Gong, and Yuanzhu Chen. 2014. The effect of news and public mood on stock movements. *Information Sciences* 278: 826–40. <https://doi.org/10.1016/j.ins.2014.03.096>.
- Nishimura, Yoshito, and Jared D. Acoba. 2022. Impact of Breast Cancer Awareness Month on Public Interest in the United States between 2012 and 2021: A Google Trends Analysis. *Cancers* 14: 2534. <https://doi.org/10.3390/cancers14102534>.
- Papadamou, Stephanos, Alexandros Koulis, Constantinos Kyriakopoulos, and Athanasios P. Fassas. 2022. Carnabris Stocks Returns: The Role of Liquidity and Investors' Attention via Google Metrics. *International Journal of Financial Studies* 10: 7. <https://doi.org/10.3390/ijfs10010007>.
- Perry, Marcus B. 2011. The Weighted Moving Average Technique. In *Wiley Encyclopedia of Operations Research and Management Science*. Hoboken: John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470400531.eorms0964>.
- Poutachidou, Nikolettta, and Stephanos Papadamou. 2021. The Effect of Quantitative Easing through Google Metrics on US Stock Indices. *International Journal of Financial Studies* 9: 56. <https://doi.org/10.3390/ijfs9040056>.
- Praekhaow, Puchong. 2010. Determination of Trading Points using the Moving Average Methods. Paper presented at the International Conference for a Sustainable Greater Mekong Subregion, Bangkok, Thailand, August 26–27, pp. 519–44.
- Preis, Tobias, Helen Susannah Moat, and H. Eugene Stanley. 2013. Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports* 3: 1684. <https://doi.org/10.1038/srep01684>.
- Robles, Victor Bielza, Concha Larranaga, Santiago Pedro González, and Lucila Ohno-Machado. 2008. Optimizing logistic regression coefficients for discrimination and calibration using estimation of distribution algorithms. *Top* 16: 345–66. <https://doi.org/10.1007/s11750-008-0054-3>.
- Sadorsky, Perry. 2021. A Random Forests Approach to Predicting Clean Energy Stock Prices. *Journal of Risk and Financial Management* 14: 48. <https://doi.org/10.3390/jrfm14020048>.
- Smart, Hoon. 2018. หุ้นสุดเหวี่ยงปิดลบ 8 จุด - ตามขาดขาย 1.8 พันล้าน. Available online: <https://hoonsmart.com/archives/39374> (accessed on 8 April 2021).
- Sycinska-Dziarnowska, Magdalena, Liliana Szyszka-Sommerfeld, Krzysztof Woźniak, Steven J. Lindauer, and Gianrico Spagnuolo. 2022. Predicting Interest in Orthodontic Aligners: A Google Trends Data Analysis. *Journal of Environmental Research and Public Health* 19: 3105. <https://doi.org/10.3390/ijerph19053105>.
- Teixeira, Lamartine Almeida, and Adriano Lorena Inácio de Oliveira. 2010. A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Systems with Applications* 37: 6885–90. <https://doi.org/10.1016/j.eswa.2010.03.033>.
- Today, Post. 2018. หุ้นไทย 275.61 ปิดลบ 8.56 จุด. Available online: <https://www.posttoday.com/finance-stock/stock/575197> (accessed on 28 April 2021).
- Trifonova, Oxana, Lkhov Petr, and Archakov Alexander. 2014. Metabolic profiling of human blood. *Biomed Klim* 60: 281–94. <https://doi.org/10.18097/pbmc20146003281>.
- Tudor, Cristiana. 2022. The Impact of the COVID-19 Pandemic on the Global Web and Video Conferencing SaaS Market. *Electronics* 11: 2633. <https://doi.org/10.3390/electronics11162633>.
- Vaidya, Rashesh. 2018. Stochastic and Momentum Analysis of Nepalese Stock Market. *The Journal of Nepalese Business Studies* XI: 14–22.
- Wang, Kui, Jie Wan, Gang Li, and Hao Sun. 2022. A Hybrid Algorithm-Level Ensemble Model for Imbalanced Credit Default Prediction in the Energy Industry. *Energies* 15: 5206. <https://doi.org/10.3390/en15145206>.
- Yu, Yang, Wenjing Duan, and Qing Cao. 2013. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems* 55: 919–26. <https://doi.org/10.1016/j.dss.2012.12.028>.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

## ประวัติผู้เขียน

ชื่อ	นายกิตติภาพ แซ่เตีย
วัน เดือน ปีเกิด	28 สิงหาคม 2540
ที่อยู่ปัจจุบัน	23/2 หมู่ 6 ตำบล สลักไธ อำเภอ เมือง จังหวัด สุรินทร์ 32000
ประวัติการศึกษา	(2562) วิทยาศาสตรบัณฑิต สาขาคณิตศาสตร์ประยุกต์ เกรดเฉลี่ย 3.33 (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง) (2565) วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการข้อมูลและการวิเคราะห์ (สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง)



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า  
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้