เอกซ์จีบูสต์สำหรับการคัดเลือกหลักทรัพย์ขนาดใหญ่และกลางในประเทศไทยแบบพหุปัจจัย

นายธนดล ประพฤทธิกุล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

XGBoost for Thailand Large-Mid Capitalization Stock Selection Based on Multi-factor

Mr. Thanadon Praphutikul

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2022

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	เอกซ์จีบูสต์สำหรับการคัดเลือกหลักทรัพย์ขนาดใหญ่และกลางในประเทศไทยแบบพหุปัจจัย
โดย	นายธนดล ประพฤทธิกุล
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.ภาสกร อภิรักษ์วรพินิต)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ธนดล ประพฤทธิกุล : เอกซ์จีบูสต์สำหรับการคัดเลือกหลักทรัพย์ขนาดใหญ่และกลางในประเทศไทยแบบพหุปัจจัย. (XGBoost for Thailand Large-Mid Capitalization Stock Selection Based on Multi-factor) อ.ที่ปรึกษาหลัก : รศ. ดร.ญาใจ ลิ่มปิยะกรณ์

วิธีการลงทุนแบบหมุนเวียนหลักทรัพย์ในพอร์ตโฟลิโอซึ่งจะทำการคัดเลือกหลักทรัพย์ที่มีค่าคะแนนสูงที่สุดเมื่อเปรียบเทียบกับหลักทรัพย์ตัวอื่นๆนี้ เป็นวิธีที่ได้รับความนิยมเมื่อเปรียบเทียบกับการเข้าซื้อขายตามสัญญาณแบบดั้งเดิม โดยปัจจุบันเทคโนโลยีการเรียนรู้ของเครื่องถูกนำมาประยุกต์ใช้กับงานต่างๆในชีวิตประจำวัน รวมถึงการลงทุนในตลาดหลักทรัพย์ด้วย ซึ่งการเรียนรู้ของเครื่องนี้สามารถนำมาใช้กับการลงทุนเชิงปริมาณเพื่อเพิ่มสมรรถนะของพอร์ตโฟลิโอสำหรับการลงทุนในตลาดการเงิน งานวิจัยนี้เสนอการใช้เอกซ์จีบูสต์สำหรับการคัดเลือกหลักทรัพย์แบบรายเดือนและรายไตรมาส ซึ่งแบบจำลองทั้งสองนี้ถูกพัฒนาบนชุดข้อมูลหลักทรัพย์ขนาดใหญ่และขนาดกลางในประเทศไทยโดยมีปัจจัยทั้งหมด 27 ปัจจัยแบ่งออกตามกลุ่มต่างๆได้ ดังนี้ ปัจจัยเชิงคุณค่า, ปัจจัยการเติบโตของกิจการ, ปัจจัยแนวโน้มของราคา, ปัจจัยสภาพคล่อง, ปัจจัยคุณภาพ, ปัจจัยเงินปันผล และปัจจัยด้านขนาด เพื่อใช้เป็นตัวแปรต้น สำหรับให้แบบจำลองเรียนรู้และจากผลการศึกษา พบว่า ปัจจัยเชิงเทคนิคเป็นปัจจัยที่มีความสำคัญสำหรับการทำนายการเคลื่อนไหวของราคาสินทรัพย์แบบรายเดือน ซึ่งแตกต่างกับการเคลื่อนไหวของราคาสินทรัพย์แบบรายไตรมาสที่ปัจจัยพื้นฐานจะมีความสำคัญมากกว่าปัจจัยเชิงเทคนิค นอกจากนี้ เมื่อจำลองการสร้างพอร์ตโฟลิโอการหมุนเวียนหลักทรัพย์แบบรายเดือนและรายไตรมาส พบว่า พอร์ตโฟลิโอแบบรายเดือนให้ค่าสถิติในเชิงพอร์ตโฟลิโอที่ดีกว่าซึ่งเป็นเพราะพอร์ตโฟลิโอดังกล่าวมีโอกาสในการคัดเลือกหลักทรัพย์ใหม่ๆเข้ามาในพอร์ตโฟลิโอได้มากกว่า อย่างไรก็ตาม ในด้านค่าสถิติของการซื้อขายกลับพบว่าพอร์ตโฟลิโอแบบรายไตรมาสกลับให้ค่าที่ดีกว่า เนื่องจากระยะเวลาการถือครองหลักทรัพย์ที่นานขึ้นทำให้มีค่าผลตอบแทนคาดหวังต่อการซื้อขายที่สูงขึ้น ทั้งนี้ พอร์ตโฟลิโอทั้งสองยังคงให้ระดับผลตอบแทนต่อความเสี่ยงที่สูงกว่าดัชนีอ้างอิงทั้งหมดประกอบด้วย ดัชนี SET TRI, ดัชนีที่ให้น้ำหนักการลงทุนแบบเท่าเทียม, และดัชนีผลตอบแทนแบบปัจจัยเดียว

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่อนิสิต .....
ปีการศึกษา	2565	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6470190621 : MAJOR COMPUTER SCIENCE

KEYWORD: stock selection, multi-factor, XGBoost, ensemble learning

Thanadon Praphutikul : XGBoost for Thailand Large-Mid Capitalization Stock Selection Based on Multi-factor. Advisor: Assoc. Prof. YACHAI LIMPIYAKORN, Ph.D.

Unlike traditional buy and sell signals, Rotational trading is a popular method of switching positions between various symbols based on their relative score. Today machine learning techniques are used in various real-world applications, including investments in stock markets. Quantitative investment powered by machine learning would enhance the performance of portfolio formation in financial markets. In this research, we propose an approach of applying a highly efficient gradient boosting tree-based ensemble, XGBoost, for multi-factor stock selection. The two models, monthly and quarterly stock selection were trained on Thailand large-mid capitalization data containing twenty-seven factors which belong to several categories such as value, growth, momentum, liquidity, quality, dividend, and size. It is discovered that the technical factor mainly affects the price movement in monthly, whereas the fundamental factor majorly influences the stock changing trends in quarterly. The monthly and quarterly rotational portfolio simulation were then performed to evaluate the investing performance measured by portfolio and trade statistics. The results show that the monthly portfolios outperformed in terms of portfolio statistics, due to more opportunities to select new stocks into the portfolio, while in terms of trade statistics, the quarterly portfolios achieved the better results since the longer holding period would increase trade expectancy. Finally, the both monthly and quarterly rotational portfolio advised by XGBoost also gain the higher risk-adjusted return on investment compared to the SET TRI Index, Equal-Weighted Index that are widely used for stock selection, and Factor Index.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Field of Study:	Computer Science	Student's Signature .....
Academic Year:	2022	Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงเป็นอย่างดีได้ด้วยความอนุเคราะห์จากรองศาสตราจารย์ ดร. ญาใจ ลิ่มปิยะกรณ์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ได้สละเวลาอันมีค่ามาให้ความรู้แนวคิด คำปรึกษา ตลอดจนตรวจสอบ และแก้ไขปัญหาข้อผิดพลาดต่าง ๆ จนทำให้งานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดีผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้ด้วย

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ดร.สุกรีสินธุภิญโญ ประธานกรรมการสอบวิทยานิพนธ์ และอาจารย์ดร.ภาสกร อภิรักษ์วรพินิต กรรมการสอบวิทยานิพนธ์ที่กรุณาเสียสละเวลาอันมีค่า ให้คำแนะนำที่เป็นประโยชน์ในการทำวิทยานิพนธ์ในครั้งนี้

ขอขอบพระคุณบิดา มารดา และญาติพี่น้องทุกคน ที่ได้ให้การสนับสนุน ความรัก ความเป็นห่วง และเป็นกำลังใจที่ดีเสมอมา

ขอขอบคุณเพื่อน ๆ พี่ๆ น้อง ๆ ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ทุกคน ที่คอยช่วยเหลือ ให้คำปรึกษา แลกเปลี่ยนความรู้ความคิดเห็นในด้านต่างๆ ตลอดระยะเวลาที่ผ่านมา

สุดท้ายนี้ขอขอบพระคุณผู้ที่เกี่ยวข้องทุกท่านที่ไม่ได้กล่าวมาข้างต้นที่คอยให้ความช่วยเหลือ ซึ่งทำให้วิทยานิพนธ์สำเร็จลุล่วงไปได้ด้วยดีผู้วิจัยหวังเป็นอย่างยิ่งว่าวิทยานิพนธ์ฉบับนี้จะเป็นประโยชน์ต่อผู้ที่สนใจไม่มากก็น้อย

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ธนดล ประพฤทธิกุล

สารบัญ

หน้า

..... ค

บทคัดย่อภาษาไทย ..... ค

..... ง

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ฌ

สารบัญรูปภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ที่มาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์การวิจัย 2

1.3 ขอบเขตการวิจัย 2

1.4 ขั้นตอนการวิจัย 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ 3

1.6 ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์ 3

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 ทฤษฎีที่เกี่ยวข้อง 4

2.1.1 สมมติฐานตลาดที่มีประสิทธิภาพ (Efficient Market Hypothesis) 4

2.1.2 แบบจำลองพหุปัจจัย (Multi-factor Model) 4

2.1.3 เอกซ์จีบูสต์ (XGBoost) 5

2.2 งานวิจัยที่เกี่ยวข้อง 6

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2.1 Stock Selection based on eXtreme Gradient Boosting 6
2.2.2 Dynamic Weighting Multi Factor Stock Selection Strategy Based on XGBoost Machine Learning 6
2.2.3 A Novel Quantitative Stock Selection Model Based on Support Vector Regression 7
2.2.4 Research on quantitative multi-factor stock selection model based on machine learning 7
บทที่ 3 ระเบียบวิธีการวิจัย 8
3.1 เลือกช่วงระยะเวลา (Window Selection) 8
3.2 คัดเลือกกลุ่มตัวอย่างหลักทรัพย์ (Universe Selection) 9
3.3 การทำความสะอาดข้อมูล (Data Cleaning) 11
3.4 การแปลงข้อมูล (Data Transformation) 12
3.5 แบ่งชุดข้อมูล (Train/Test Split) 13
3.6 สร้างแบบจำลองการคัดเลือกหลักทรัพย์ (Stock Selection Model Construction) 14
3.7 ประเมินสมรรถนะของแบบจำลอง (Model Evaluation) 14
3.8 สร้างพอร์ตโฟลิโอด้วยการจำลองการทดสอบย้อนหลัง (Portfolio Construction with Backtesting Simulation) 16
3.9 กำหนดค่าสถิติหรือดัชนีชี้วัดจากมิติต่างๆ สำหรับการประเมินพอร์ตโฟลิโอ (Specify Portfolio Evaluation Metrics) 20
3.10 การประเมินผลพอร์ตโฟลิโอ (Portfolio Evaluation) 21
บทที่ 4 ผลลัพธ์และการอภิปรายผล 26
4.1 ประเมินสมรรถนะแบบจำลอง 26
4.1.1 ประเมินสมรรถนะแบบจำลองบนชุดข้อมูลตรวจสอบ 26
4.1.2 ปัจจัยที่สำคัญสำหรับการทำนายผลตอบแทนของหลักทรัพย์ 28
4.1.3 สมรรถนะการทำนายหลักทรัพย์บนชุดข้อมูลทดสอบ 30

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2 ประเมินผลลัพธ์ของพอร์ตโฟลิโอการลงทุน 32
4.2.1 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์รายเดือนแบบพหุปัจจัยด้วย XGBoost กับพอร์ตโฟลิโอรายไตรมาส 32
4.2.2 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์รายเดือนและรายไตรมาสแบบพหุปัจจัยด้วย XGBoost กับดัชนีอ้างอิง 36
4.2.3 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบพหุปัจจัยด้วย XGBoost กับพอร์ตโฟลิโอการลงทุนแบบปัจจัยเดียว 38
4.2.4 ศึกษาผลกระทบต้นทุนการซื้อขายที่มีต่อพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบพหุปัจจัยด้วย XGBoost ในเชิงปฏิบัติ 43
บทที่ 5 บทสรุป 47
5.1 สรุปผลการวิจัย 47
5.2 ข้อจำกัดในงานวิจัย 48
5.3 งานวิจัยในอนาคต 48
ภาคผนวก สูตรคำนวณค่าสถิติและดัชนีชี้วัดพอร์ตโฟลิโอ 49
บรรณานุกรม 52
ประวัติผู้เขียน 56

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตารางที่ 1 ค่าสถิติเชิงบรรยายมูลค่าหลักทรัพย์ตามราคาตลาดของกลุ่มตัวอย่าง 9

ตารางที่ 2 มูลค่าหลักทรัพย์ตามราคาตลาดขนาดใหญ่สุด *5* อันดับแรก ณ สิ้นปี ค.ศ. 2019 10

ตารางที่ 3 ตัวอย่างการคำนวณค่าความคลาดเคลื่อนของราคาสำหรับการเข้าซื้อหลักทรัพย์..... 19

ตารางที่ 4 ค่าสถิติของพอร์ตโฟลิโอ Equal-Weighted 23

ตารางที่ 5 ค่าสถิติของพอร์ตโฟลิโอ Low PE และ High PE 25

ตารางที่ 6 ค่าไฮเปอร์พารามิเตอร์ของแบบจำลองการคัดเลือกหลักทรัพย์รายเดือนและรายไตรมาส26

ตารางที่ 7 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือนและแบบรายไตรมาส 33

ตารางที่ 8 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือน, รายไตรมาส ระหว่าง XGBoost, ดัชนี SET TRI และดัชนี Equal-Weighted 38

ตารางที่ 9 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือนและแบบรายไตรมาสกับค่าสถิติเชิงบรรยายของดัชนีผลตอบแทนตามปัจจัยต่างๆ..... 42

ตารางที่ 10 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost หลังต้นทุนการซื้อขายแบบรายเดือน และแบบรายไตรมาส 44

ตารางที่ 11 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost หลังต้นทุนการซื้อขายแบบรายเดือน รายไตรมาส, ดัชนี SET TRI และดัชนี Equal-Weighted 46

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญรูปภาพ

หน้า

ภาพที่ 1 ตัวอย่างขั้นตอนการทำงานของ XGBoost 5

ภาพที่ 2 ผังระเบียบวิธีการวิจัย 8

ภาพที่ 3 การเคลื่อนไหวของดัชนี SET TRI ตั้งแต่ช่วงปี ค.ศ. 2012 ถึงปี ค.ศ. 2021..... 9

ภาพที่ 4 บริษัทที่มีมูลค่าหลักทรัพย์ตามราคาตลาดใหญ่ที่สุด *5* อันดับแรก ณ สิ้นปี ค.ศ. 2019 10

ภาพที่ 5 กราฟราคาและมูลค่าหลักทรัพย์ตามราคาตลาดของบริษัท การบินไทย จำกัด (มหาชน)... 11

ภาพที่ 6 ตัวอย่างการทำความสะอาดข้อมูล 12

ภาพที่ 7 ตัวอย่างการแปลงข้อมูล..... 13

ภาพที่ 8 การแบ่งชุดข้อมูล 13

ภาพที่ *9* Confusion Matrix..... 15

ภาพที่ 10 ตัวอย่าง AUC ROC..... 15

ภาพที่ 11 กลยุทธ์การลงทุนแบบจับจังหวะตลาดโดยอาศัยแนวโน้มของราคา..... 17

ภาพที่ 12 ตัวอย่างการจัดพอร์ตโฟลิโอหมุนเวียนหลักทรัพย์แบบรายเดือน 17

ภาพที่ 13 ตัวอย่างการคำนวณค่าความคลาดเคลื่อนของการซื้อขายแบบความผันผวน 19

ภาพที่ 14 ผลตอบแทนสะสมของดัชนี SET TRI และค่าการถดถอยของพอร์ตโฟลิโอ..... 21

ภาพที่ 15 ผลตอบแทนสะสมของ Equal-Weighted Portfolio และจำนวนหลักทรัพย์ที่ถือครอง . 22

ภาพที่ 16 ผลตอบแทนสะสมของพอร์ตโฟลิโอ Low PE (เส้นเขียว) และ High PE (เส้นแดง)..... 24

ภาพที่ 17 AUC ROC การทำนายผลตอบแทนหลักทรัพย์ในอนาคตหนึ่งเดือนบน Validation set. 27

ภาพที่ 18 AUC ROC การทำนายผลตอบแทนหลักทรัพย์ในอนาคตหนึ่งไตรมาสบน Validation set 28

ภาพที่ 19 ปัจจัยที่มีอิทธิพลต่อการทำนายผลตอบแทนหลักทรัพย์หนึ่งเดือนข้างหน้า..... 29

ภาพที่ 20 ปัจจัยที่มีอิทธิพลต่อการทำนายผลตอบแทนหลักทรัพย์หนึ่งไตรมาสข้างหน้า..... 30

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาพที่ 21 AUC ROC การทำนายผลตอบแทนหลักทรัพย์ในอนาคตหนึ่งเดือนบน Test set 31

ภาพที่ 22 AUC ROC การทำนายผลตอบแทนหลักทรัพย์ในอนาคตหนึ่งไตรมาสบน Test set 31

ภาพที่ 23 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือน (เส้นเขียว) และแบบรายไตรมาส (เส้นน้ำเงิน)..... 33

ภาพที่ 24 การเคลื่อนไหวของราคาบริษัท ศรีตรังแอโกรอินดัสทรี จำกัด (มหาชน)..... 34

ภาพที่ *25* การเคลื่อนไหวของราคาบริษัท เอเชีย เอวิเอชั่น จำกัด (มหาชน)..... 35

ภาพที่ 26 การเคลื่อนไหวของราคาบริษัทเดลต้า อีเลคโทรนิคส์ (ประเทศไทย) จำกัด (มหาชน)..... 35

ภาพที่ 27 การเคลื่อนไหวของราคาบริษัท ทีคิวเอ็ม อัลฟา จำกัด (มหาชน)..... 36

ภาพที่ 28 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือน (เส้นเขียว), รายไตรมาส (เส้นน้ำเงิน), ดัชนี SET TRI (เส้นประสีดำ) และดัชนี Equal-Weighted (เส้นประสีแดง)..... 37

ภาพที่ 29 การเคลื่อนไหวของราคาหลักทรัพย์ PLANB 39

ภาพที่ 30 การเคลื่อนไหวของหลักทรัพย์ TFMAMA 40

ภาพที่ 31 การเคลื่อนไหวของหลักทรัพย์ DELTA 41

ภาพที่ 32 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือน (เส้นเขียว), รายไตรมาส (เส้นน้ำเงิน) และดัชนีผลตอบแทนตามปัจจัยต่างๆ (เส้นประ) 42

ภาพที่ 33 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost หลังต้นทุนการซื้อขายแบบรายเดือน (เส้นเขียว) และแบบรายไตรมาส (เส้นน้ำเงิน) 44

ภาพที่ 34 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost หลังต้นทุนการซื้อขายแบบรายเดือน (เส้นเขียว), แบบรายไตรมาส (เส้นน้ำเงิน), ดัชนี SET TRI (เส้นประสีดำ) และดัชนี Equal-Weighted (เส้นประสีแดง) 45

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

หนึ่งในองค์ประกอบหลักที่สำคัญของการลงทุนคือ การคัดเลือกหลักทรัพย์ ซึ่งราคาของสินทรัพย์มักจะถูกขับเคลื่อนด้วยปัจจัยต่างๆ มากมาย ได้แก่ ปัจจัยด้านการเติบโต (Growth Factor) และปัจจัยโมเมนตัม (Momentum Factor) เป็นต้น ในปี ค.ศ. 1989 งานวิจัยของ Bernard และ Thomas [1] ได้ศึกษาปรากฏการณ์การเคลื่อนไหวของหลักทรัพย์ภายหลังที่งบการเงินประกาศ (Post-Earnings-Announcement Drift) พบว่า พฤติกรรมการเคลื่อนไหวของราคาหลักทรัพย์นั้นมีความสัมพันธ์กับอัตราการเติบโตของกิจการกล่าวคือ กลุ่มของบริษัทที่มีอัตราการเติบโตของกิจการสูง ราคาหลักทรัพย์ก็มักจะปรับตัวเพิ่มสูงขึ้นตามไปด้วย ในทางกลับกัน กลุ่มของบริษัทที่มีอัตราการเติบโตของกิจการต่ำ ราคาหลักทรัพย์ก็มักจะปรับตัวลดลงตามไปด้วย ซึ่งสะท้อนให้เห็นถึงปัจจัยด้านการเติบโตของกิจการที่มีผลต่อการคัดเลือกหลักทรัพย์ ในปี ค.ศ. 1993 Jegadeesh และ Titman [2] ได้ศึกษาการจัดพอร์ตโฟลิโอ (Portfolio) โดยทำการซื้อกลุ่มหลักทรัพย์ผู้ชนะ (Winner Stock) และขายกลุ่มหลักทรัพย์ผู้แพ้ (Loser Stock) พบว่า กลุ่มหลักทรัพย์ผู้ชนะในอดีตมีแนวโน้มที่จะเป็นผู้ชนะต่อไปในอนาคต และกลุ่มหลักทรัพย์ผู้แพ้ในอดีตก็มักจะเป็นผู้แพ้ต่อไปในอนาคตด้วยเช่นกัน ซึ่งสะท้อนให้เห็นถึงปัจจัยโมเมนตัม (Momentum Factor) ที่มีผลต่อการคัดเลือกหลักทรัพย์ นอกจากปัจจัยทั้งสองดังกล่าว ยังมีปัจจัยอื่นๆที่ส่งผลกระทบต่อราคาหลักทรัพย์ รวมทั้งความสัมพันธ์ระหว่างตัวแปรเหล่านี้ก็มีความซับซ้อน ผู้วิจัยจึงมีแนวคิดการประยุกต์ใช้เทคโนโลยีการเรียนรู้ของเครื่อง (Machine Learning) เพื่อช่วยให้การคัดเลือกหลักทรัพย์มีประสิทธิผลยิ่งขึ้น

งานวิจัยนี้เสนอการใช้เอกซ์จีบูสต์ (XGBoost) สำหรับการคัดเลือกหลักทรัพย์ขนาดใหญ่และกลางในประเทศไทย ซึ่งเป็นกลุ่มหลักทรัพย์ที่นักลงทุนให้ความสนใจและนิยมลงทุน เนื่องจากบริษัทเหล่านี้มักเป็นที่รู้จักอย่างกว้างขวาง รวมถึงมีสภาพคล่องในการซื้อขายสูง ในงานวิจัยนี้จะใช้ปัจจัยเชิงเทคนิค (Technical Factor) และปัจจัยงบการเงิน (Fundamental Factor) เป็นตัวแปรตั้งต้นในการสร้างแบบจำลองเพื่อคัดเลือกหลักทรัพย์แบบรายเดือนและรายไตรมาสที่สามารถสร้างอัตราผลตอบแทนในระดับที่สูงกว่าดัชนีอ้างอิงอื่นๆ ประกอบด้วย SETTRI Index, Equal-Weighted Index และ Factor Index / Style Index เพื่อสนับสนุนแนวคิดการประยุกต์ใช้ปัญญาประดิษฐ์

(Artificial Intelligence) สำหรับการคัดเลือกหลักทรัพย์แบบพหุปัจจัยนั้นมีประสิทธิภาพ รวดเร็ว แม่นยำ และมีประสิทธิผลให้ผลตอบแทนสูงกว่าการคัดเลือกหลักทรัพย์จากปัจจัยเพียงปัจจัยเดียว

1.2 วัตถุประสงค์การวิจัย

1. ศึกษาปัจจัยที่มีอิทธิพลต่อการเปลี่ยนแปลงราคาหลักทรัพย์ขนาดใหญ่และกลางในประเทศไทย
2. พัฒนาแบบจำลองการคัดเลือกหลักทรัพย์เพื่อสร้างพอร์ตโฟลิโอที่ให้อัตราผลตอบแทนต่อความเสี่ยงที่สูงกว่าดัชนีอ้างอิงต่างๆ ในภาพรวม

1.3 ขอบเขตการวิจัย

1. ใช้ชุดข้อมูลจากบริษัท สยามควอนท์ จำกัด ประกอบด้วย ข้อมูลราคาหุ้น, ปัจจัยพื้นฐาน และงบการเงินที่ระบุวันประกาศข่าวตามจริง รวมทั้งข้อมูลสถิติการซื้อขายต่างๆ ที่เกิดขึ้นในตลาดหลักทรัพย์แห่งประเทศไทย
2. พัฒนาแบบจำลองสำหรับการจำแนกหลักทรัพย์ด้วยอัลกอริทึม XGBoost โดยใช้ไลบรารีภาษา Python
3. พหุปัจจัยในงานวิจัยนี้ประกอบด้วยอย่างน้อย 20 ปัจจัย ซึ่งจัดอยู่ในหมวดปัจจัยเชิงเทคนิคและปัจจัยด้านงบการเงิน
4. จำลองการสร้างพอร์ตโฟลิโอด้วยกระบวนการทดสอบย้อนหลังด้วยภาษา Amibroker Formula Language (AFL) บนโปรแกรม Amibroker เพื่อเปรียบเทียบประสิทธิผลและความเสถียรของการคัดเลือกหลักทรัพย์ด้วยแบบจำลอง XGBoost กับดัชนีอ้างอิงต่างๆ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

1.4 ขั้นตอนการวิจัย

1. ศึกษาค้นคว้าทฤษฎีและงานวิจัยที่เกี่ยวข้อง
2. เตรียมชุดข้อมูลสำหรับการวิจัย
3. ออกแบบและพัฒนาโมเดลการเรียนรู้สำหรับคัดเลือกหลักทรัพย์
4. จำลองการสร้างพอร์ตโฟลิโอ
5. ประเมินผลงานวิจัย
6. เผยแพร่ผลงานทางวิชาการ
7. สรุปผลและเรียบเรียงวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้องค์ความรู้เกี่ยวกับปัจจัยที่มีอิทธิพลต่อการเปลี่ยนแปลงของราคาหลักทรัพย์ขนาดกลางและขนาดใหญ่ในประเทศไทย
2. ได้แบบจำลองในการคัดเลือกหลักทรัพย์ในประเทศไทยที่มีแนวโน้มจะได้รับอัตราผลตอบแทนต่อความเสี่ยงที่สูงกว่าตลาด ซึ่งสามารถใช้เป็นเครื่องมือสำหรับการตัดสินใจของนักลงทุน

1.6 ผลงานที่ได้รับการตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์ฉบับนี้ได้รับการตีพิมพ์บทความทางวิชาการจำนวน 2 บทความ

1. T. Praphutikul, and Y. Limpiyakorn, “XGBoost for Smart Portfolio Management Based on Multi Factor Stock Selection”, 9th International Conference on Computer Technology Applications (ICCTA 2023), May 10-12, 2023, Vienna, Austria.
2. T. Praphutikul, and Y. Limpiyakorn, “XGBoost-Based Multi-Factor Stock Selection Model for Rotational Trading”, 5th International Conference on Information Technology and Computer Communications (ITCC 2023), June 15-17, 2023, Tianjin, China.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 สมมติฐานตลาดที่มีประสิทธิภาพ (Efficient Market Hypothesis)

ในปี ค.ศ. 1970 Fama [3] ได้แบ่งภาวะของตลาดออกเป็น 3 ระดับ คือ ตลาดที่เข้มแข็ง (Strong), กึ่งเข้มแข็ง (Semi-Strong) และ อ่อน (Weak) ตลาดที่เข้มแข็ง หมายความว่า นักลงทุนไม่สามารถสร้างผลตอบแทนที่สูงกว่าตลาดได้ เนื่องจากราคาของสินทรัพย์ได้สะท้อนถึงข้อมูลข่าวสารทั้งหมดในตลาดเป็นที่เรียบร้อยแล้ว อย่างไรก็ตาม ในความเป็นจริง ราคาหลักทรัพย์อาจจะไม่ได้ตอบสนองต่อข้อมูลข่าวสารใหม่ทุกชนิดอย่างทันท่วงที ตัวอย่างของตลาดที่ไม่เข้มแข็ง ได้แก่ Post-Earnings-Announcement Drift [1] ซึ่งเป็นปรากฏการณ์ที่ภายหลังงบการเงินประกาศแล้ว ราคาของสินทรัพย์จึงค่อยๆที่จะปรับตัวเพิ่มสูงขึ้น รวมถึงการเข้าซื้อหลักทรัพย์ผู้ชนะ (Winner Stock) [2] ซึ่งราคาของสินทรัพย์ก็ยังคงมีการปรับตัวเพิ่มสูงขึ้นอย่างต่อเนื่อง จากปรากฏการณ์ที่กล่าวมาข้างต้นเป็นหลักฐานที่บ่งชี้ว่าตลาดมีความผิดปกติ (Market Anomaly) ทำให้นักลงทุนสามารถสร้างผลตอบแทนที่สูงกว่าตลาดได้ด้วยการบริหารพอร์ตโฟลิโออย่างชาญฉลาด

2.1.2 แบบจำลองพหุปัจจัย (Multi-factor Model)

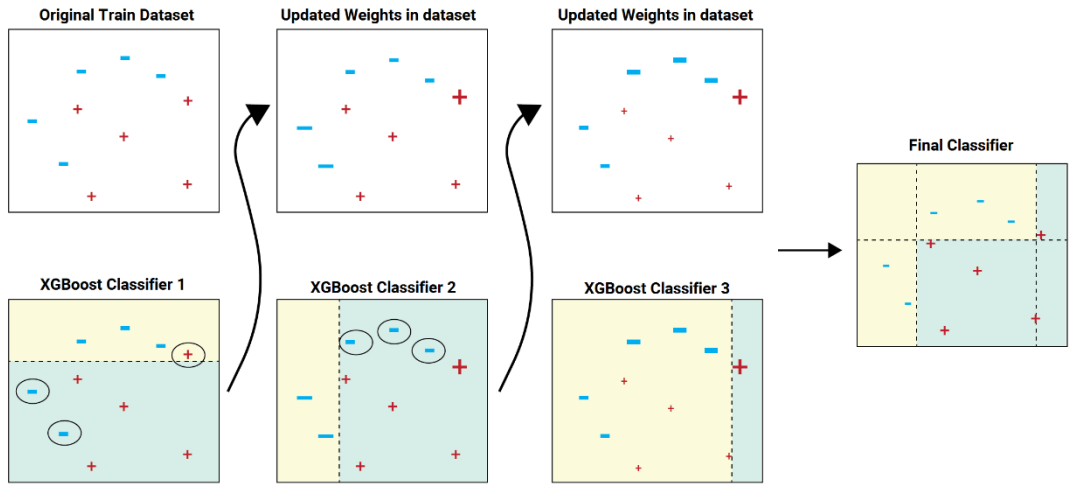
แบบจำลองพหุปัจจัยเป็นแบบจำลองความสัมพันธ์ระหว่างปัจจัยต่างๆ มากกว่าหนึ่งปัจจัยเพื่อวิเคราะห์หรืออธิบายการเคลื่อนไหวของราคาสินทรัพย์ [4] โดยในปี ค.ศ. 1993 Fama และ French [5] ได้นำเสนอแบบจำลอง 3 ปัจจัย (Three-Factor Model) ประกอบด้วย ปัจจัยด้านความเสี่ยงจากตลาด (Market Risk Premium), ปัจจัยด้านขนาด (Size Factor) และปัจจัยมูลค่าทางบัญชี (Book Value Factor) สำหรับการอธิบายผลตอบแทน ต่อมา ในปี ค.ศ. 1997 แนวคิดดังกล่าวได้มีการพัฒนาเป็นแบบจำลอง 4 ปัจจัย (Four-Factor Model) โดย Carhart [6] และพัฒนาเป็นแบบจำลอง 5 ปัจจัย (Five-Factor Model) ในปี ค.ศ. 2015 โดย Fama และ French [7] ซึ่งแสดงให้เห็นถึงความพยายามในการนำปัจจัยต่างๆ มาใช้ร่วมกันทำนาย

อย่างไรก็ตาม แบบจำลองการเงินดั้งเดิมเหล่านี้มักอธิบายความสัมพันธ์ระหว่างตัวแปรต่างๆ ในเชิงเส้นตรง (Linear Expression) แต่ในความเป็นจริง ตัวแปรต่างๆเหล่านี้มักไม่ได้มีความสัมพันธ์เชิงเส้น ในปัจจุบันจึงได้มีการประยุกต์ใช้เทคโนโลยีการเรียนรู้ของเครื่องมาอธิบายความสัมพันธ์เหล่านี้ เช่น ในปี ค.ศ. 2018 Zhang และคณะ [8] ได้สร้างแบบจำลองพหุปัจจัยด้วย Long Short-

Term Memory (LSTM) ซึ่งสามารถค้นพบความสัมพันธ์ในรูปแบบที่ไม่ใช่เส้นตรงได้เป็นอย่างดี ต่อมาในปี ค.ศ. 2021 Zhong และคณะ [9] ได้ศึกษาการสร้างแบบจำลองพหุปัจจัยด้วย Logistic Regression (LR), XGBoost และ Linear Regression พบว่า ทั้ง LR และ XGBoost สามารถสร้างผลตอบแทนได้สูงกว่าแบบจำลองที่สร้างด้วย Linear Regression

2.1.3 เอกซ์จีบูสต์ (XGBoost)

ในปี ค.ศ. 2016 Chen และคณะ [10] ได้คิดค้น XGBoost (eXtreme Gradient Boosting) ซึ่งเป็นโมเดลการเรียนรู้ที่มีรากฐานมาจากต้นไม้ตัดสินใจ (Decision Tree) โดยใช้เทคนิค Ensemble Learning คือ การรวมของโมเดลการเรียนรู้ที่หลากหลายเพื่อช่วยกันหาคำตอบหรือทำนายที่ให้ผลลัพธ์ความคลาดเคลื่อนที่น้อยที่สุด ในการเพิ่มสมรรถนะของแบบจำลอง (model performance) ต้นไม้ตัดสินใจแต่ละต้นจะเรียนรู้จากค่าความผิดพลาดของต้นก่อนหน้าเพื่อเพิ่มความแม่นยำในการทำนาย ดังภาพที่ 1 นอกจากนี้ ในการเรียนรู้ของต้นไม้แต่ละต้นนั้นจะแบ่งชุดข้อมูลออกเป็นข้อมูลย่อยๆ เพื่อยังลดปัญหา Overfitting จากการเรียนรู้ของแบบจำลอง



ภาพที่ 1 ตัวอย่างขั้นตอนการทำงานของ XGBoost

สำหรับการประยุกต์ใช้ XGBoost พบว่า มีการนำไปใช้ในหลากหลายสาขาด้วยกัน เช่น ในปี ค.ศ. 2019 Hsieh และคณะ [11] ได้เสนอการประยุกต์ใช้ XGBoost เพื่อตรวจจับการตอบสนองของร่างกายมนุษย์ต่อความเครียดได้เป็นอย่างดี รวมถึงในปี ค.ศ. 2020 Li และคณะ [12] พบว่า XGBoost มีประสิทธิภาพในการคัดเลือกคุณลักษณะเฉพาะ (Feature Selection) รวมถึงการจำแนกประเภทในการประเมินปัญหาเครดิตบนฐานข้อมูลขนาดใหญ่

2.2 งานวิจัยที่เกี่ยวข้อง

2.2.1 Stock Selection based on eXtreme Gradient Boosting

ในปี ค.ศ. 2019 Zhange และ Chen [13] ได้ทำการศึกษาวิจัยการคัดกรองหลักทรัพย์ในดัชนี CSI300 Index ที่มีคุณลักษณะที่ผลตอบแทนที่สูงกว่าตลาด (Outperformance Stock) ในหนึ่งเดือนข้างหน้า โดยทำการแบ่งกลุ่มของข้อมูลออกเป็น 3 กลุ่มด้วยกันคือ กลุ่มที่ให้อัตราผลตอบแทนส่วนเกินที่สูงที่สุด 30% บน, กลุ่มที่ให้อัตราผลตอบแทนส่วนเกินต่ำที่สุด 30% ล่าง และกลุ่มที่อยู่ตรงกลางระหว่างทั้งสองกลุ่ม จากนั้นใช้อัลกอริทึมจำแนกประเภทประกอบด้วย XGBoost, Logistic Regression (LR) และ Support Vector Machine (SVM) โดยมีปัจจัยทั้งหมด 62 ปัจจัย ประกอบด้วยปัจจัยบัญชี, ปัจจัยคุณค่า, ปัจจัยด้านอารมณ์ และปัจจัยเทคนิคเป็นตัวแปรตั้งต้น

ในการประเมินประสิทธิผลของการลงทุนประเมินจากการสร้างพอร์ตโฟลิโอโดยการคัดเลือกหลักทรัพย์ที่ถูกจำแนกว่าอยู่ในกลุ่มที่ดีที่สุดทั้งหมด 15 ตัว ซึ่งจะทำการกระจายน้ำหนักการลงทุนแบบเท่าเทียมกัน (Equal-Weighted) และจะมีการหมุนเวียนหลักทรัพย์ในทุกๆเดือน โดยพบว่าการคัดเลือกหลักทรัพย์ด้วย XGBoost นั้นให้ผลตอบแทนแบบสะสม (Cumulative Return) ตลอดระยะเวลา 32 เดือนในระดับที่สูงที่สุดเมื่อเปรียบเทียบกับ LR และ SVM โดยมีระดับผลตอบแทนเมื่อเปรียบเทียบกับดัชนี CSI 300 Index นั้นมีค่าเท่ากับ 134% และ 28% ตามลำดับ

2.2.2 Dynamic Weighting Multi Factor Stock Selection Strategy Based on XGBoost Machine Learning

ในปี ค.ศ. 2018 Li และ Zhang [14] ทำการศึกษาการคัดเลือกหลักทรัพย์โดยใช้ XGBoost ปรับน้ำหนักการให้ความสำคัญของแต่ละปัจจัยในแต่ละช่วงระยะเวลา ซึ่งมีปัจจัยทั้งหมด 6 ปัจจัยด้วยกันประกอบด้วย ปัจจัยสภาพคล่อง, ปัจจัยขนาด, ปัจจัยอัตราผลตอบแทนต่อส่วนของผู้ถือหุ้น (ROE Factor), ปัจจัยการเติบโต, ปัจจัยผลตอบแทน และปัจจัยคุณภาพ สำหรับการเรียนรู้และผลลัพธ์ที่ได้จากการทำนายนั้นจะเป็นค่าน้ำหนักความสำคัญของแต่ละปัจจัยซึ่งจะพลวัตไปตามช่วงระยะเวลาต่างๆ

จากนั้นจึงทำการสร้างพอร์ตโฟลิโอพบว่าพอร์ตโฟลิโอที่ให้ความสำคัญของปัจจัยแบบพลวัตรนั้นมีผลตอบแทนโดยเฉลี่ยทบต้อต่อปีเท่ากับ 22.54% เมื่อเปรียบเทียบกับการให้น้ำหนักความสำคัญของปัจจัยแบบคงที่และการลงทุนที่การกระจายน้ำหนักการลงทุนแบบเท่าเทียมกันซึ่งมีค่าเท่ากับ 16.22% และ 14.56% ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2.3 A Novel Quantitative Stock Selection Model Based on Support Vector Regression

ในปี ค.ศ. 2019 Dai และ Zhou [15] ได้ศึกษาวิจัยการคัดเลือกหลักทรัพย์ด้วยการใช้ Support Vector Regression (SVR) บนฐานข้อมูลหลักทรัพย์ในกลุ่ม A-Share ของประเทศจีน โดยคณะผู้วิจัยทำการคัดเลือกปัจจัยที่ให้อัตราส่วนผลตอบแทนต่อความเสี่ยงสูงจากทั้งหมด 500 ปัจจัย ประกอบด้วย 5 ปัจจัยหลัก ได้แก่ ปัจจัยคุณค่า, ปัจจัยคุณภาพ, ปัจจัยด้านขนาด, ปัจจัยด้านกำไร, ปัจจัยการเติบโต เพื่อเป็นตัวแปรตั้งต้นสำหรับการพัฒนาแบบจำลอง และมีผลตอบแทนในอนาคตของหลักทรัพย์เป็นตัวแปรตาม จากนั้นจึงสร้างพอร์ตโฟลิโอโดยเลือกหลักทรัพย์ที่มีผลลัพธ์การทำนายว่ามีผลตอบแทนในอนาคตที่สูงที่สุด 5 ลำดับแรก และจะทำการหมุนเวียนหลักทรัพย์ในทุกๆเดือน พบว่า พอร์ตโฟลิโอมีอัตราผลตอบแทนต่อความเสี่ยง โดยวัดจากอัตราส่วน Sharpe Ratio ซึ่งมีค่าสูงกว่าการลงทุนแบบเท่าเทียมกัน (Equal-Weighted) ซึ่งมีค่าเท่ากับ 0.56 และ 0.26 ตามลำดับ ทั้งนี้สำหรับผลตอบแทนโดยเฉลี่ยทบต้นต่อปี (Compound Annual Growth Rate/CAGR) นั้นก็สูงกว่า Equal Weighted โดยมีค่าเท่ากับ 13.62% และ 5.68% ตามลำดับ

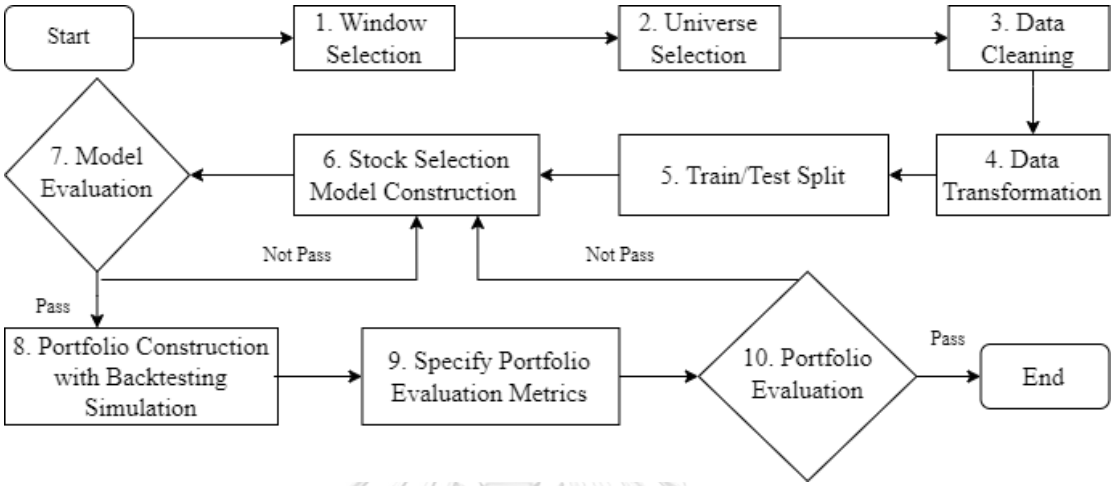
2.2.4 Research on quantitative multi-factor stock selection model based on machine learning

ในปี ค.ศ. 2021 Yang [16] ได้ศึกษาวิจัยการคัดเลือกหลักทรัพย์ที่อยู่ในดัชนี CSI 500 Index ซึ่งถือเป็นตัวแทนของหลักทรัพย์ที่มูลค่าตามราคาตลาดขนาดกลาง โดยทำการคัดเลือกปัจจัยจากทั้งหมด 44 ปัจจัย ประกอบด้วย ปัจจัยการเติบโต, ปัจจัยคุณค่า, ปัจจัยด้านขนาด, ปัจจัยคุณภาพ, ปัจจัยเชิงสภาวะตลาด รวมถึงปัจจัยเชิงเทคนิคอื่นๆ ด้วยการทำ Factor Investing เพื่อค้นหาปัจจัยที่สามารถสร้างผลตอบแทนได้สูงกว่าตลาด นอกจากนี้ ผู้วิจัยได้ทำการวิเคราะห์และตัดทิ้งปัจจัยที่มีความคล้ายคลึงกันมาก โดยวัดจากค่าสหสัมพันธ์ระหว่างปัจจัยด้วยกัน เมื่อได้ปัจจัยที่เป็นตัวแปรตั้งต้นแล้ว จึงทำการสร้างแบบจำลองด้วย AdaBoost ซึ่งผลลัพธ์ที่ได้จะเป็นคะแนน (Scoring) ของหลักทรัพย์นั้นๆ และประเมินผลลัพธ์จากการลงทุนด้วยกระบวนการทดสอบย้อนหลัง (Backtesting) ด้วยการคัดเลือกหุ้นที่มีค่าความสำคัญโดยวัดจากคะแนนสูงที่สุด 10 อันดับเข้ามาในพอร์ตโฟลิโอ พบว่า การคัดเลือกหลักทรัพย์ด้วย AdaBoost นั้น สามารถสร้างผลตอบแทนในระดับ 551.75% ตั้งแต่ปี ค.ศ. 2015 ถึง 2017 รวมระยะเวลาการทดสอบทั้งสิ้น 3 ปีและมีอัตราส่วนผลตอบแทนต่อความเสี่ยงที่สูงถึง 0.97

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 3
ระเบียบวิธีการวิจัย

ผังขั้นตอนการวิจัยพอสังเขป แสดงดังภาพที่ 2 โดยมีรายละเอียดดังนี้



ภาพที่ 2 ผังระเบียบวิธีการวิจัย

3.1 เลือกช่วงระยะเวลา (Window Selection)

งานวิจัยนี้กำหนดช่วงระยะเวลาของการศึกษาทั้งหมด 10 ปีตั้งแต่ปี ค.ศ. 2012 ถึงปี ค.ศ. 2021 ดังภาพที่ 3 ประกอบด้วย

1. ช่วงพัฒนาแบบจำลอง (Training Model) สำหรับพัฒนาและการประเมินสมรรถนะแบบจำลอง (Model Validation) การคัดเลือกหลักทรัพย์ด้วย XGBoost โดยกำหนดไว้ที่ 80% ของปริมาณข้อมูลทั้งหมด คือตั้งแต่ ปี ค.ศ. 2012 ถึง 2019 รวมระยะเวลาทั้งสิ้น 8 ปี
2. ช่วงทดสอบแบบจำลอง (Testing Model) สำหรับการสร้างพอร์ตโฟลิโอ โดยกำหนดไว้ที่ 20% ของปริมาณข้อมูลทั้งหมด คือตั้งแต่ ปี ค.ศ. 2020 ถึง 2021 รวมระยะเวลาทั้งสิ้น 2 ปี ซึ่งในช่วงระยะเวลาดังกล่าวจะครอบคลุมถึงช่วงที่ดีและแย่ของตลาด เช่น ภาวะวิกฤตเศรษฐกิจจากการแพร่ระบาดของไวรัสโควิด 19 ช่วงต้นปี ค.ศ. 2020 และช่วงที่เศรษฐกิจจะเริ่มมีการฟื้นตัวตอนปลายปี ค.ศ. 2020 เป็นต้น

ภาพที่ 3 การเคลื่อนไหวของดัชนี SET TRI ตั้งแต่ช่วงปี ค.ศ. 2012 ถึงปี ค.ศ. 2021

3.2 คัดเลือกกลุ่มตัวอย่างหลักทรัพย์ (Universe Selection)

คัดเลือกกลุ่มตัวอย่างที่มีมูลค่าหลักทรัพย์ตามราคาตลาด (Market Capitalization) หรือเรียกย่อๆ ว่า Market Cap สูงที่สุด 200 อันดับแรก ณ สิ้นปี ค.ศ. 2019 เป็นกลุ่มตัวอย่างที่เป็นตัวแทนของหลักทรัพย์ขนาดใหญ่และขนาดกลางของประเทศไทยสำหรับการศึกษาในงานวิจัยนี้ ตารางที่ 1 สรุปค่าเฉลี่ยและค่ามัธยฐานของมูลค่าหลักทรัพย์ตามราคาตลาดจำนวนทั้งหมด 200 ตัว มีค่าเท่ากับ 78,861.23 และ 27,598.57 ล้านบาท โดยมีค่าสูงสุดและต่ำสุดเท่ากับ 1,256,771.81 และ 8,109 ล้านบาท ตามลำดับ และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 151,076.33 ล้านบาท

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 1 ค่าสถิติเชิงบรรยายมูลค่าหลักทรัพย์ตามราคาตลาดของกลุ่มตัวอย่าง

Market Capitalization (as of 30/12/2019)	
#Total	200
Average	78,861.23
Median	27,598.57
Minimum	8,109.00
Maximum	1,256,771.81
St. Dev.	151,076.33

ภาพที่ 4 แสดงบริษัทที่มีมูลค่าหลักทรัพย์ตามราคาตลาด (Market Capitalization) สูงที่สุด 5 อันดับแรก ประกอบด้วย บริษัท ปตท. จำกัด (มหาชน), บริษัท ท่าอากาศยานไทย จำกัด (มหาชน), บริษัท ซีพี ออลล์ จำกัด (มหาชน), บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน) และ บริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) โดยมีมูลค่าระหว่าง 494,263.17 ถึง 1,256,771.81 ล้านบาท รายละอียดดังแสดงในตารางที่ 2

ภาพที่ 4 บริษัทที่มีมูลค่าหลักทรัพย์ตามราคาตลาดใหญ่ที่สุด *5* อันดับแรก ณ สิ้นปี ค.ศ. 2019

ตารางที่ 2 มูลค่าหลักทรัพย์ตามราคาตลาดขนาดใหญ่สุด *5* อันดับแรก ณ สิ้นปี ค.ศ. 2019

บริษัท	มูลค่าหลักทรัพย์ตามราคาตลาด (ล้านบาท)
บริษัท ปตท. จำกัด (มหาชน) หรือ PTT	1,256,771.81
บริษัท ท่าอากาศยานไทย จำกัด (มหาชน) หรือ AOT	1,060,713.2
บริษัท ซีพี ออลล์ จำกัด (มหาชน) หรือ CPALL	649,029.09
บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน) หรือ ADVANC	633,287.02
บริษัท ปตท.สำรวจและผลิตปิโตรเลียม จำกัด (มหาชน) หรือ PTTEP	494,263.17

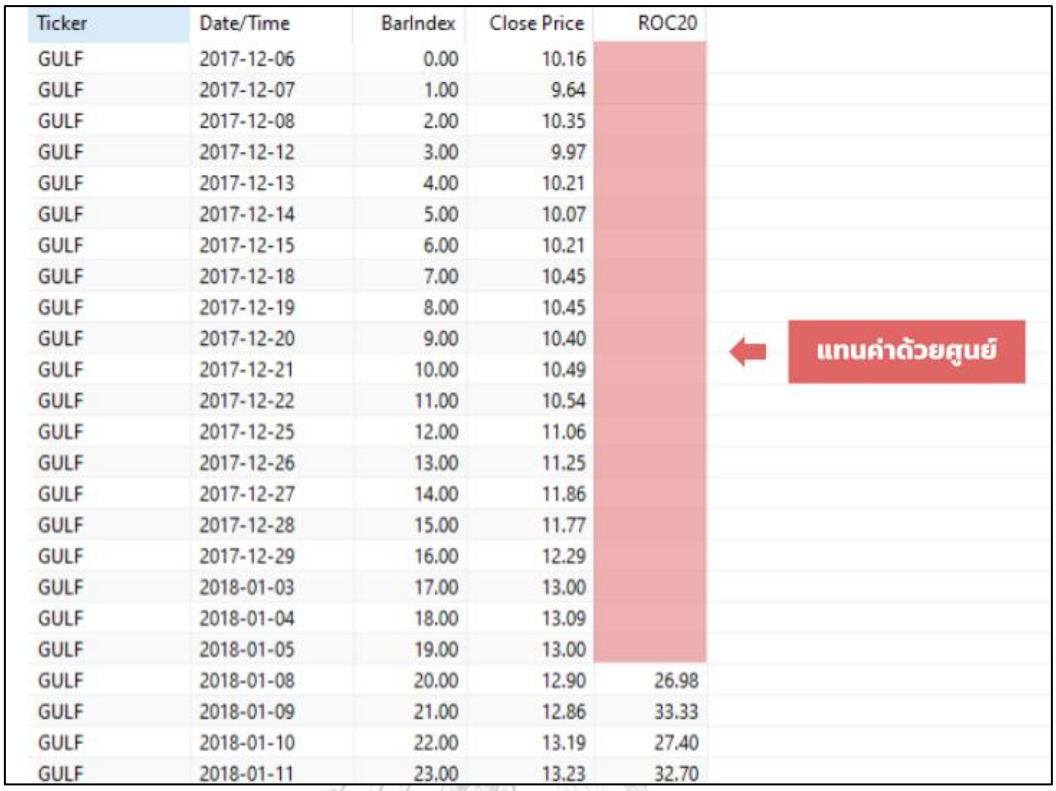
ทั้งนี้ ในกรณีที่หลักทรัพย์ถูกเพิกถอนออกจากตลาดหลักทรัพย์แห่งประเทศไทยภายหลังปี ค.ศ. 2019 เช่น บริษัท การบินไทย จำกัด (มหาชน) ซึ่งมี Market Cap เท่ากับ 14,951.99 ล้านบาท (ภาพที่ 5) และเป็นหนึ่งในกลุ่มตัวอย่างของงานวิจัยนี้ ถึงแม้ว่าบริษัทดังกล่าวถูกระงับการซื้อขาย (Suspending Stock) ตั้งแต่วันที่ 2021-05-17 เพื่อเข้าสู่การฟื้นฟูกิจการตามมติความเห็นชอบของศาลล้มละลายกลาง ในงานวิจัยนี้ กลุ่มหลักทรัพย์ที่ถูกเพิกถอนออกจากตลาดเหล่านี้จะไม่ถูกคัดออกจากกลุ่มตัวอย่างเพื่อเป็นการป้องกันอคติการแอบดูข้อมูลอนาคต (Lookahead bias) และอคติการคัดเลือกเฉพาะผู้ที่อยู่รอด (Survivorship bias) ในช่วงทดสอบแบบจำลอง อ้างอิงจากงานวิจัย Baquero และคณะ [17] ที่ค้นพบว่า อคติดังกล่าวจะทำให้ผลลัพธ์ของการวิจัยนั้นดีเกินจริง

ภาพที่ 5 กราฟราคาและมูลค่าหลักทรัพย์ตามราคาตลาดของบริษัท การบินไทย จำกัด (มหาชน)

3.3 การทำความสะอาดข้อมูล (Data Cleaning)

ข้อมูลจะถูกจัดในลักษณะภาคตัวขวาง (Cross-Section) สำหรับกรณีที่ค่าคุณลักษณะเป็นค่าว่าง (Null) เนื่องจากข้อมูลไม่เพียงพอ จะแทนค่าว่างด้วยค่าศูนย์ ตัวอย่างเช่น จากภาพที่ 6 ซึ่งเป็นตัวอย่างการคำนวณอัตราผลตอบแทนย้อนหลัง 1 เดือน (20 วันทำการ) ซึ่งจำเป็นจะต้องใช้ราคาในปัจจุบันเปรียบเทียบกับราคาในอดีต 20 วันก่อนหน้า พิจารณาบริษัท กัลฟ์ เอ็นเนอร์จี ดีเวลลอปเมนท์ จำกัด (มหาชน) หรือ GULF ที่พึ่งเข้าจดทะเบียนในตลาดหลักทรัพย์ และเข้าซื้อขายในตลาดหลักทรัพย์แห่งประเทศไทยวันแรก คือ 6 ธันวาคม ค.ศ. 2017 ทำให้ในช่วง 20 วันทำการแรก คือ ตั้งแต่วันแรกจนถึงวันที่ 5 มกราคม ค.ศ. 2018 ไม่สามารถคำนวณผลตอบแทนในอดีตย้อนหลังหนึ่งเดือน (ROC20) ได้ ส่งผลให้ค่าดังกล่าวมีลักษณะเป็นค่าว่าง ผู้วิจัยจึงทำการแทนค่าด้วยศูนย์ เป็นต้น

สำหรับกรณีที่คุณลักษณะของหลักทรัพย์มีค่าที่สูงหรือต่ำกว่าปกติ (Outlier) จะคงค่าดังกล่าวไว้ เนื่องจากลักษณะข้อมูลดังกล่าวถือเป็นเรื่องปกติที่เกิดขึ้นได้ในตลาดหลักทรัพย์ อ้างอิงจากปรากฏการณ์ กลุ่มความผันผวน (Volatility clustering) ในงานวิจัยปี ค.ศ. 1963 Mandelbrot [18] ซึ่งพบว่า แนวโน้มการเปลี่ยนแปลงของราคาหลักทรัพย์มักมีขนาดเหมือนเดิมในวันถัดๆไป นอกจากนี้ ในปี ค.ศ. 2020 Chong และคณะ [19] พบว่า ปริมาณการซื้อขายที่สูงจนผิดปกติ (Unusual trading volumes) สามารถทำนายอัตราผลตอบแทนที่ผิดปกติ (Abnormal returns) ในช่วงที่มีการประกาศงบการเงินได้



Ticker	Date/Time	BarIndex	Close Price	ROC20
GULF	2017-12-06	0.00	10.16	
GULF	2017-12-07	1.00	9.64	
GULF	2017-12-08	2.00	10.35	
GULF	2017-12-12	3.00	9.97	
GULF	2017-12-13	4.00	10.21	
GULF	2017-12-14	5.00	10.07	
GULF	2017-12-15	6.00	10.21	
GULF	2017-12-18	7.00	10.45	
GULF	2017-12-19	8.00	10.45	
GULF	2017-12-20	9.00	10.40	
GULF	2017-12-21	10.00	10.49	
GULF	2017-12-22	11.00	10.54	
GULF	2017-12-25	12.00	11.06	
GULF	2017-12-26	13.00	11.25	
GULF	2017-12-27	14.00	11.86	
GULF	2017-12-28	15.00	11.77	
GULF	2017-12-29	16.00	12.29	
GULF	2018-01-03	17.00	13.00	
GULF	2018-01-04	18.00	13.09	
GULF	2018-01-05	19.00	13.00	
GULF	2018-01-08	20.00	12.90	26.98
GULF	2018-01-09	21.00	12.86	33.33
GULF	2018-01-10	22.00	13.19	27.40
GULF	2018-01-11	23.00	13.23	32.70

ภาพที่ 6 ตัวอย่างการทำความสะอาดข้อมูล

3.4 การแปลงข้อมูล (Data Transformation)

รูปแบบการป้อนข้อมูลสำหรับการเรียนรู้แบบจำลองจะอยู่ในรูปแบบภาคตัดขวางของแต่ละวัน ประกอบด้วยรายชื่อของหลักทรัพย์, วันที่, ปัจจัยทั้งยี่สิบเจ็ดปัจจัย และกลุ่มของผลตอบแทน ซึ่งจะถูกดำเนินการ ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1. แปลงข้อมูลตัวแปรตั้งต้นหรือปัจจัย โดยในแต่ละปัจจัยค่าถูกแปลงเป็นตัวเลขอันดับแล้วแปลงเป็นกลุ่มเดไซล์ (Decile) โดยมีค่าตั้งแต่ 0 ถึง 9 หรือแบ่งออกเป็นสิบกลุ่มตามลำดับเปอร์เซ็นต์ไทล์ดังภาพที่ 7 ประกอบไปด้วยข้อมูลกลุ่มตัวอย่างหลักทรัพย์ ณ วันที่ 2020-01-02 ทั้งหมด 10 ตัวจากจำนวนหลักทรัพย์ทั้งหมด 100 ตัว ของวันที่ 2020-01-02 ซึ่งข้อมูล ณ วันดังกล่าวถูกเรียงลำดับตามค่าของปัจจัย PE (Price-to-Earnings Ratio) จากมากไปน้อย หลังจากนั้นผู้วิจัยจึงจัดกลุ่มค่าของปัจจัยดังกล่าวด้วยการแปลงให้เป็นค่า 0-9 ขึ้นอยู่กับลำดับเปอร์เซ็นไทล์ตั้งแต่ 0-9 เช่น จากจำนวนหลักทรัพย์ตัวอย่างทั้งหมดคือ 100 รายการนี้ กรณีหลักทรัพย์ที่มีปัจจัยอยู่ในลำดับที่ 1-

10 จะถูกจัดหมวดหมู่ในกลุ่มแรกด้วยค่า 0 ซึ่งสะท้อนถึงหลักทรัพย์สิบเปอร์เซ็นไทล์แรกที่มีค่าของปัจจัยนั้นๆสูงที่สุด เป็นต้น

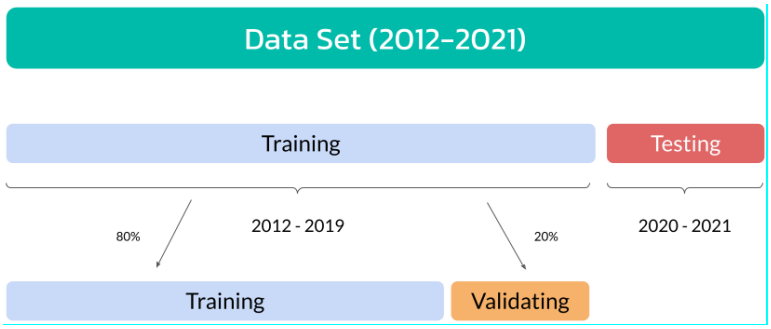
2. แบ่งตัวแปรตามออกเป็น 3 กลุ่มประกอบด้วย กลุ่มที่ดีที่สุด 30% (Top30%), กลุ่มตรงกลาง 40% (Middle 40%) และกลุ่มล่างสุด 30% (Bottom 30%) โดยพิจารณาจากผลตอบแทนในอนาคตหนึ่งเดือน (20 วันทำการซื้อขาย) และผลตอบแทนในอนาคตสามเดือนหรือรายไตรมาส (60 วันทำการซื้อขาย)

Symbol	Date/Time	PE	RankPE	Decile RankPE	Decile RankPBV	...	20D Returns	20D Returns Class	60D Returns	60D Returns Class
AWC	2020-01-02	215.74	1	0	9	...	-5.26	1	-12.74	2
CKP	2020-01-02	107.52	2	0	4	...	-18	0	-40	1
CIMBT	2020-01-02	91.64	3	0	9	...	26.92	2	-21.02	2
GULF	2020-01-02	79.5	4	0	0	...	15.68	2	-46.21	0
GPSC	2020-01-02	71.48	5	0	0	...	2.61	2	-3.05	2
VGI	2020-01-02	68.46	6	0	1	...	-10.66	0	-29.35	1
BGRIM	2020-01-02	67.34	7	0	1	...	21.23	2	-37.23	1
ACE	2020-01-02	63.65	8	0	9	...	-30.23	0	-37.5	1
TRUE	2020-01-02	63.27	9	0	6	...	-16.96	0	-40.4	0
THG	2020-01-02	57.11	10	0	2	...	0	1	-32.47	1
...							...			

ภาพที่ 7 ตัวอย่างการแปลงข้อมูล

3.5 แบ่งชุดข้อมูล (Train/Test Split)

แบ่งข้อมูลที่เก็บรวบรวมทั้งหมดตั้งแต่ปี ค.ศ. 2012 ถึง 2021 ออกเป็น 2 ชุด: 1) ชุดข้อมูลสอน (Training set) สำหรับการพัฒนาแบบจำลอง ด้วยอัตราส่วน 80% (ค.ศ. 2012 – 2019) และ 2) ชุดข้อมูลทดสอบ (Test set) สำหรับการทดสอบแบบจำลอง ด้วยอัตราส่วน 20% (ค.ศ. 2020 – 2021) สำหรับชุดข้อมูลสอนจะถูกแบ่งออกเป็น 2 ส่วน ด้วยอัตราส่วน 80:20 โดยส่วน Validation set 20% จะถูกใช้สำหรับการปรับแต่งค่าไฮเปอร์พารามิเตอร์ระหว่างการเรียนรู้ของแบบจำลอง และการประเมินสมรรถนะแบบจำลองบนชุดข้อมูลฝึกสอน ดังแสดงในภาพที่ 8



ภาพที่ 8 การแบ่งชุดข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.6 สร้างแบบจำลองการคัดเลือกหลักทรัพย์ (Stock Selection Model Construction)

สร้างแบบจำลองการคัดเลือกหลักทรัพย์แบบรายเดือนและรายไตรมาสด้วย XGBoost โดยใช้ปัจจัยเชิงเทคนิคและปัจจัยด้านงบการเงินรวมทั้งสิ้น 27 ตัวแปรเป็นตัวแปรตั้งต้น และผลลัพธ์ของการทำนายคือ ค่าความน่าจะเป็นของหลักทรัพย์ที่จะตกอยู่ในกลุ่มต่างๆ คือ กลุ่มที่จะให้ผลตอบแทนมากที่สุดในอนาคตข้างหน้า (Top30%), กลุ่มที่จะให้ผลตอบแทนในระดับกลางในอนาคตข้างหน้า (Middle40%), และกลุ่มที่จะให้ผลตอบแทนน้อยที่สุดในอนาคตข้างหน้า (Bottom30%)

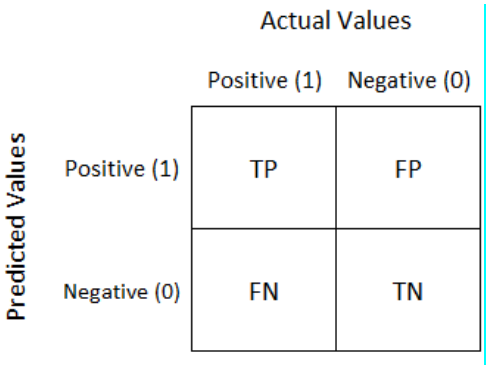
การสร้างแบบจำลองการคัดเลือกหลักทรัพย์ด้วย XGBoost ผู้วิจัยใช้ไลบรารี XGBoost บนภาษา Python 3.9.12 ภายใต้สภาพแวดล้อม Anaconda เครื่องคอมพิวเตอร์ที่ใช้ มีหน่วยการประมวลผลกลาง (CPU) คือ Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz และมีหน่วยความจำ (Memory) ทั้งหมด16 กิกะไบต์ ระบบปฏิบัติการไมโครซอฟท์ วินโดวส์ 10 โฮม (Windows 10 Home) รุ่น 64 บิต

3.7 ประเมินสมรรถนะของแบบจำลอง (Model Evaluation)

คุณภาพของแบบจำลองวัดด้วย ROC Curve (Receiver Operating Characteristics Curve) และ AUC (Area under the ROC Curve) โดยเส้นโค้ง ROC แสดงคุณภาพการทำนายของแบบจำลองที่หาค่าสมดุลระหว่างค่าความไว (Sensitivity) กับค่าความจำเพาะ (Specificity) วัดจากความสัมพันธ์ระหว่างค่า Sensitivity หรือ อัตราผลบวกจริง (True Positive Rate—TPR) บนแกน Y กับค่า 1 - Specificity หรือ อัตราผลบวกลวง (False Positive Rate—FPR) บนแกน X โดยเส้นกราฟ ROC ที่ดีกว่าจะให้พื้นที่ใต้กราฟที่มากกว่า ภาพที่ 9 แสดง Confusion Matrix ซึ่งอธิบายความหมายของค่าตัววัด: True Positive (TP), FP (False Positive), FN (False Negative), TN (True Negative) ซึ่งเป็นตัววัดฐาน (Base measure) ที่ใช้คำนวณตัววัดอนุพันธ์ (Derived measure) ต่อไป ได้แก่ ค่า Precision, Recall หรือ Sensitivity, Speificity สมการที่ 1 แสดงสูตรการคำนวณค่า Sensitivity (TPR) และสมการที่ 2 แสดงสูตรการคำนวณค่า 1 – Specificity (FPR)

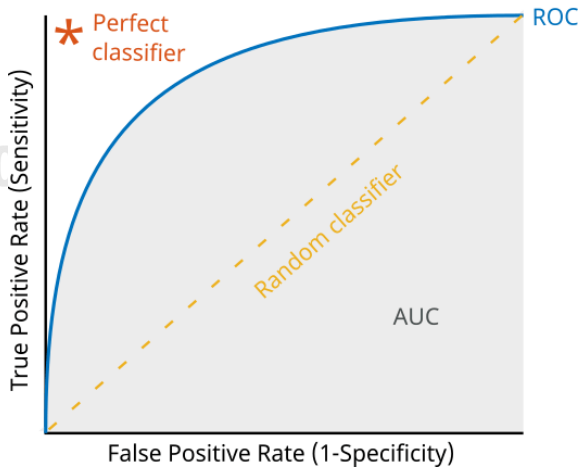
$$True\ Positive\ Rate\ (TPR) = \frac{Ture\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \quad (1)$$

$$False\ Positive\ Rate\ (FPR) = \frac{False\ Positive\ (FP)}{False\ Positive\ (FP) + Ture\ Negative\ (TN)} \quad (2)$$



ภาพที่ *9* Confusion Matrix

ภาพที่ 10 แสดงตัวอย่าง AUC-ROC หากแบบจำลองมีสมรรถนะที่ดีที่สุดคือ มีค่า True Positive Rate เท่ากับ 1 กล่าวคือ แบบจำลองมีสมรรถนะการทำนายกลุ่มตัวอย่างได้ถูกต้องทั้งหมด หรือไม่มีตัวอย่างบวกใดเลยที่แบบจำลองทำนายเป็นกลุ่มตัวอย่างลบ และมีค่า False Positive Rate เท่ากับ 0 คือ สำหรับกลุ่มตัวอย่างที่เป็นกลุ่มลบทั้งหมด ไม่มีตัวอย่างใดเลยที่แบบจำลองทำนายเป็นกลุ่มบวก ทำให้ลักษณะเส้นโค้ง ROC จะอยู่ที่มุมด้านซ้ายบนของกราฟ และมี AUC-ROC เท่ากับ 1 ในทางกลับกัน หากค่าพื้นที่ใต้กราฟมีค่าที่ต่ำแสดงถึงแบบจำลองมีสมรรถนะการทำนายที่ต่ำ ทั้งนี้ สมรรถนะแบบจำลองที่ยอมรับได้จะต้องทีเส้นโค้งที่อยู่เหนือระดับการทำนายแบบสุ่ม (Random Classifier) หรือมีค่าพื้นที่ใต้เส้นโค้งที่มากกว่า 0.5



ภาพที่ 10 ตัวอย่าง AUC ROC

3.8 สร้างพอร์ตโฟลิโอด้วยการจำลองการทดสอบย้อนหลัง (Portfolio Construction with Backtesting Simulation)

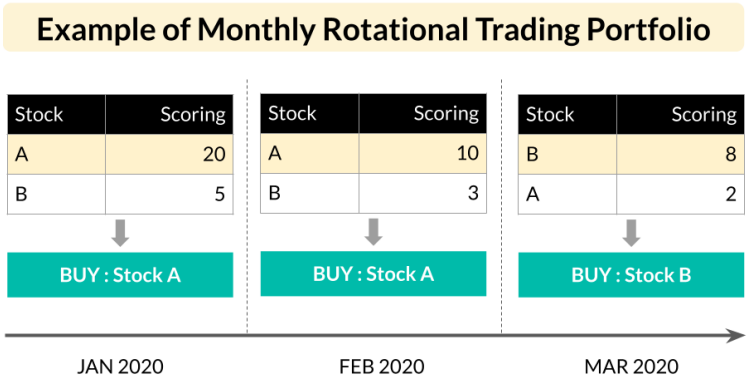
สร้างพอร์ตโฟลิโอการลงทุนเพื่อตรวจสอบประสิทธิผลของกลยุทธ์การเลือกหุ้นที่แนะนำโดยแบบจำลอง XGBoost ด้วยกระบวนการทดสอบย้อนหลัง (Backtesting) ในช่วงปี ค.ศ. 2020 ถึง 2021 โดยทำการสร้างพอร์ตโฟลิโอแบบหมุนเวียนหลักทรัพย์ (Rotational Trading) ซึ่งเป็นวิธีที่ได้รับความนิยมสำหรับการลงทุน เนื่องจากวิธีการดังกล่าวจะทำการคัดเลือกหลักทรัพย์ที่มีค่าคะแนนสูงที่สุดเมื่อเปรียบเทียบกับหลักทรัพย์ตัวอื่นๆเข้ามาในพอร์ตโฟลิโอ โดยจะหมุนเวียนหลักทรัพย์ตามช่วงระยะเวลาที่กำหนดไว้ เพื่อให้พอร์ตโฟลิโอประกอบด้วยกลุ่มของหลักทรัพย์ที่มีค่าสมรรถนะที่ดีสุดตลอดเวลา โดยกลยุทธ์การลงทุนแบบหมุนเวียนหลักทรัพย์นี้ มีความแตกต่างกับกลยุทธ์การลงทุนแบบจับจังหวะตลาด (Market Timing) เนื่องจากการจับจังหวะตลาดเป็นการคาดการณ์เพื่อค้นหาจังหวะสำหรับเข้าลงทุน โดยหากนักลงทุนคาดการณ์ว่าในอนาคตราคาของสินทรัพย์จะมีการปรับตัวเพิ่มขึ้น นักลงทุนก็จะเข้าซื้อหลักทรัพย์เพื่อสร้างผลตอบแทนจากการลงทุน ในทางกลับกัน หากนักลงทุนคาดการณ์ว่าในอนาคตราคาของสินทรัพย์จะมีการปรับตัวลดลง นักลงทุนจะทำการขายหลักทรัพย์ออก และจะทำการถือครองเงินสดแทนเพื่อรอจังหวะสำหรับการเข้าลงทุนครั้งถัดไป

สำหรับตัวอย่างกลยุทธ์การลงทุนแบบจับจังหวะของตลาด เช่น การเข้าซื้อขายหลักทรัพย์ตามแนวโน้มของราคาโดยพิจารณาจากการตัดกันของเส้นค่าเฉลี่ย (Moving Average Crossover) ดังภาพที่ 11 ซึ่งแสดงกราฟแท่งเทียน (Candle Stick) ของราคาบริษัท ปตท. จำกัด (มหาชน) โดยมีสัญลักษณ์ย่อคือ “PTT” พร้อมทั้งเส้นค่าเฉลี่ย 20 วันทำการ (เส้นประสีดำ) เป็นตัวแทนของเส้นค่าเฉลี่ยระยะสั้น และเส้นค่าเฉลี่ย 50 วันทำการ (เส้นประสีแดง) เป็นตัวแทนของเส้นค่าเฉลี่ยระยะกลาง โดยเมื่อเส้นค่าเฉลี่ยระยะสั้นตัดขึ้นเหนือเส้นค่าเฉลี่ยระยะกลางเป็นการบ่งชี้ว่าราคาหลักทรัพย์เริ่มมีการปรับตัวเพิ่มสูงขึ้นหรือเป็นสัญญาณให้เข้าลงทุน ในทางกลับกัน หากเส้นค่าเฉลี่ยระยะสั้นตัดลงต่ำกว่าเส้นค่าเฉลี่ยระยะกลางเป็นการบ่งชี้ว่าราคาของหลักทรัพย์เริ่มมีการปรับตัวลดลงหรือเป็นสัญญาณให้ขายหลักทรัพย์ โดยในช่วงที่ราคาของสินทรัพย์มีแนวโน้มปรับตัวลดลง กลยุทธ์จะถือครองเงินสดไว้แทนการถือครองสินทรัพย์ เป็นต้น

ภาพที่ 11 กลยุทธ์การลงทุนแบบจับจังหวะตลาดโดยอาศัยแนวโน้มของราคา

สังเกตได้ว่า กลยุทธ์การจัดพอร์ตโฟลิโอแบบหมุนเวียนหลักทรัพย์นี้ มีความเรียบง่าย (Simplified) รวมถึงมีความซับซ้อนที่น้อยกว่าการจับจังหวะตลาด โดยกลยุทธ์จะถือครองหลักทรัพย์ตลอดช่วงระยะเวลาของการลงทุน และมีกลไกหลักคือ การค้นหาหลักทรัพย์ที่มีสมรรถนะที่ดีที่สุดเข้ามาในพอร์ตโฟลิโอ เช่น การจัดพอร์ตโฟลิโอการหมุนเวียนหลักทรัพย์แบบรายเดือน จะคัดเลือกหลักทรัพย์ที่มีค่าสมรรถนะที่ดีที่สุดเพียงหนึ่งตัวมาลงทุนในแต่ละเดือน ดังแสดงในภาพที่ 12 คือ ในตอนต้นเดือนมกราคม ค.ศ. 2020 พบว่า หลักทรัพย์ A มีค่าสมรรถนะที่ดีกว่าหลักทรัพย์ B ดังนั้น จึงซื้อหลักทรัพย์ A และถือครองหลักทรัพย์ดังกล่าวไปจนสิ้นเดือนแล้วจึงขายหลักทรัพย์ดังกล่าวออกจากพอร์ตโฟลิโอ โดยกระบวนการดังกล่าวจะทำซ้ำเพื่อหมุนเวียนหลักทรัพย์ในทุกๆเดือน เป็นต้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



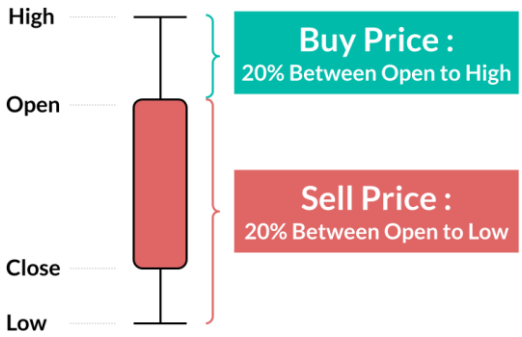
ภาพที่ 12 ตัวอย่างการจัดพอร์ตโฟลิโอหมุนเวียนหลักทรัพย์แบบรายเดือน

ในงานวิจัยนี้ ผู้วิจัยมุ่งศึกษาที่การสร้างพอร์ตโฟลิโอการหมุนเวียนหลักทรัพย์แบบรายเดือน (Monthly Rotational Trading Portfolio) และแบบรายไตรมาส (Quarterly Rotational Trading Portfolio) เนื่องจากช่วงระยะเวลาหนึ่งเดือนและหนึ่งไตรมาสนี้ เป็นช่วงระยะเวลาที่นักลงทุนสถาบันรวมถึงนักลงทุนทั่วๆไปนิยมปรับน้ำหนักการลงทุนของหลักทรัพย์ในพอร์ตโฟลิโอ ซึ่งกฎการคัดเลือกหลักทรัพย์เข้าสู่พอร์ตโฟลิโอในงานวิจัยนี้ คือ คัดเลือกหลักทรัพย์ที่มีค่าความน่าจะเป็นที่ถูกทำนายว่าเป็นกลุ่มที่ให้ผลตอบแทนสูงที่สุดในอนาคตข้างหน้าเป็นจำนวนทั้งหมด 10% ของจำนวนหลักทรัพย์ทั้งหมดที่มีอยู่ ณ ขณะนั้น (Top 10% Percentile) โดยมีการกำหนดขนาดการลงทุนของหลักทรัพย์ทุกตัวอย่างเท่าเทียมกัน (Equal-Weighted) เช่น ในกรณีที่มีหลักทรัพย์ทั้งหมด 200 ตัว จะทำการลงทุนในหลักทรัพย์ที่มีความน่าจะเป็นที่ให้ผลตอบแทนสูงที่สุดทั้งหมด 20 ตัวแรก โดยให้น้ำหนักการลงทุนของหลักทรัพย์แต่ละตัวเท่ากับ 5% ของพอร์ตโฟลิโอ เป็นต้น และกำหนดราคาสำหรับการซื้อขายคือ ราคาเปิดของวัน (At-The-Open Price)

อย่างไรก็ตาม สำหรับการลงทุนจริงนั้นจะมีต้นทุนของการซื้อขายหลักทรัพยซึ่งประกอบไปด้วยค่านายหน้า (Commission) และค่าความคลาดเคลื่อนของการซื้อขายหลักทรัพย์ (Slippage) สำหรับค่านายหน้านี้คือ ต้นทุนที่นายหน้าซื้อขายหลักทรัพย์หรือโบรคเกอร์ (Broker) เรียกเก็บจากนักลงทุนสำหรับการใช้บริการการซื้อขายหลักทรัพย์ [20] และสำหรับค่าความคลาดเคลื่อนของการซื้อขายหลักทรัพย์นี้คือ ส่วนต่างของราคาที่คาดหวังว่าจะเข้าซื้อได้ซึ่งในกรณีนี้คือราคาเปิดเปรียบเทียบกับราคาที่สามารถซื้อขายได้จริง [21] โดยในงานวิจัยชิ้นนี้ได้กำหนดค่านายหน้าที่ 0.15% ต่อการซื้อหรือขายหนึ่งครั้ง (รวม 0.3% สำหรับการซื้อและขาย) และกำหนดค่าความเคลื่อนของการซื้อขายที่ 20% ตามความผันผวนของราคาหลักทรัพย์ โดยราคาซื้อและราคาขายจะถูกคำนวณดังสมการที่ 3 และ 4 ตามลำดับ ภาพที่ 13 แสดงตัวอย่างการคำนวณความคลาดเคลื่อนการซื้อขายของราคาตามความผันผวนซึ่งจะเป็นพลวัตร (Dynamic) ตามช่วงการเคลื่อนไหวของหลักทรัพย์รายวันกล่าวคือ ในกรณีที่หลักทรัพย์มีความผันผวนสูงจะมีช่วงการเคลื่อนไหวของราคาภายในวันที่กว้าง ส่งผลให้ค่าความคลาดเคลื่อนของการซื้อขายเพิ่มสูงขึ้น ในขณะที่หลักทรัพย์ที่มีความผันผวนต่ำจะมีช่วงการเคลื่อไหนของราคาภายในวันที่แคบ ส่งผลให้ค่าความคลาดเคลื่อนของการซื้อขายลดต่ำลง

$$Buy\ Price = Open\ + (\ slippage\ threshold * (\ High - Open\ )) \quad 3)$$

$$Sell\ Price = Open\ - \big(\ slippage\ threshold * (\ Open - Low\ )\big) \quad (4)$$



ภาพที่ 13 ตัวอย่างการคำนวณค่าความคลาดเคลื่อนของการซื้อขายแบบความผันผวน

โดยตัวอย่างการคำนวณค่าความคลาดเคลื่อนของการเข้าซื้อตามที่แสดงในตารางที่ 3 พบว่าวันที่ 2 มกราคม ค.ศ. 2020 ราคาของบริษัท เอสซี แอสเสท คอร์ปอเรชั่น จำกัด (มหาชน) หรือ SC มีราคาเปิดที่ 1.83 บาท โดยระหว่างวันราคามีการปรับตัวขึ้นไปทำจุดสูงสุดที่ 1.91 บาทหรือมีช่วงการเคลื่อนไหวจากราคาเปิดจนถึงราคาสูงสุดของวันเท่ากับ 8 สตางค์ ดังนั้นเมื่อคำนวณค่าความคลาดเคลื่อนที่ 20% จากช่วงการแกว่งของราคาจะได้เท่ากับ 2 สตางค์และได้ราคาต้นทุนของการเข้าซื้อที่ 1.85 บาทหรือมีค่าความคลาดเคลื่อนระหว่างราคาซื้อกับราคาเปิดเท่ากับ 0.88% ในทางกลับกันราคาของบริษัท เอสซีจี เซรามิกส์ จำกัด (มหาชน) หรือ COTTO ซึ่งมีราคาเปิดเท่ากับ 1.47 บาทและมีราคาสูงสุดของวันเท่ากับ 1.47 บาทเช่นกัน หรือช่วงความผันผวนระหว่างราคาเปิดจนถึงราคาสูงสุดมีค่าเป็นศูนย์ ทำให้เมื่อคำนวราคาต้นทุนการเข้าซื้อของหลักทรัพย์ดังกล่าวจะได้ค่าเท่ากับ 1.47 บาท โดยมีค่าความคลาดเคลื่อนระหว่างราคาซื้อกับราคาเปิดเป็นศูนย์ เป็นต้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3 ตัวอย่างการคำนวณค่าความคลาดเคลื่อนของราคาสำหรับการเข้าซื้อหลักทรัพย์

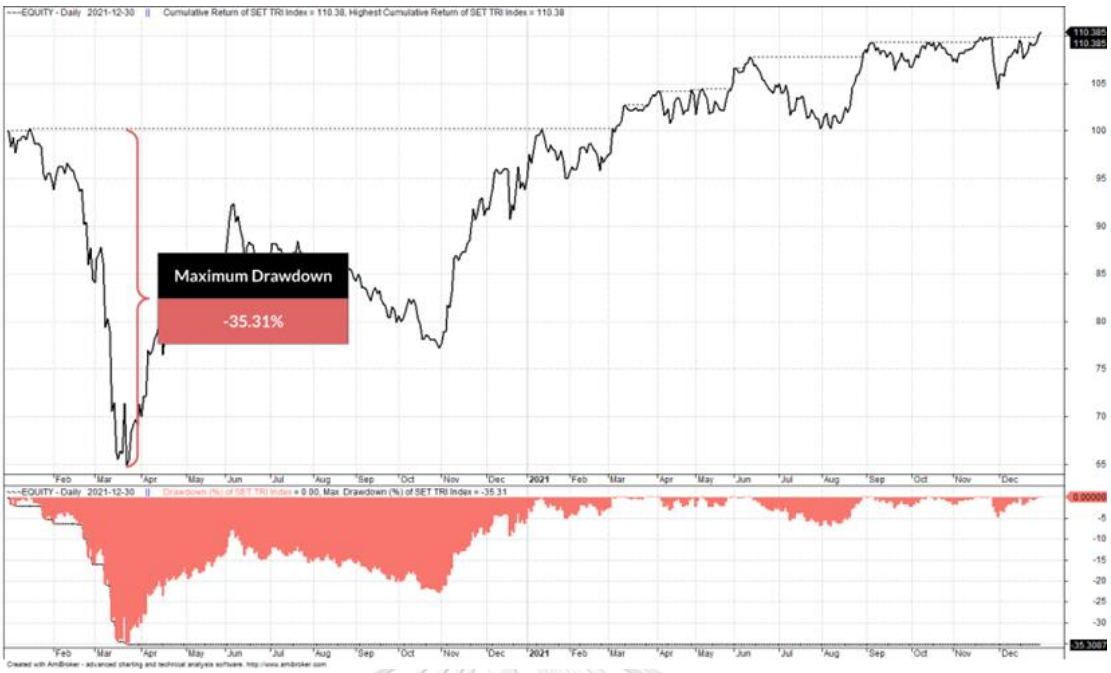
Symbol	SC	COTTO
Date	2020-01-02	2020-01-02
Open Price	1.83	1.47
High Price	1.91	1.47
Range High to Open	0.08	0
20% of Volatility	0.02	0
Buy Price	1.85	1.47
% Slippage	0.88	0

3.9 กำหนดค่าสถิติหรือดัชนีชี้วัดจากมิติต่างๆ สำหรับการประเมินพอร์ตโฟลิโอ (Specify Portfolio Evaluation Metrics)

กำหนดค่าสถิติสำหรับการประเมินผลลัพธ์การลงทุนประกอบด้วยดัชนีชี้วัดการเทรด (Trade Metrics) และดัชนีชี้วัดพอร์ตโฟลิโอ (Portfolio Metrics) โดยตัวชี้วัดการเทรดประกอบด้วยจำนวนครั้งของการซื้อขาย (# Trades), เปอร์เซ็นต์ของเทรดที่ทำกำไรได้ (% Win), เปอร์เซ็นต์ของเทรดที่ขาดทุน (% Loss), ค่าเฉลี่ยผลตอบแทนคาดหวัง (Trade Expectancy) ซึ่งคำนวณจากค่าเฉลี่ยผลตอบแทนจากทุกการซื้อขาย และค่าเฉลี่ยช่วงระยะเวลาการถือครองหลักทรัพย์ (Avg. Bar Held) ซึ่งคำนวณโดยระยะเวลาเฉลี่ยตั้งแต่การเข้าซื้อหลักทรัพย์จนกระทั่งขายหลักทรัพย์ดังกล่าวออกจากพอร์ตโฟลิโอ และสำหรับดัชนีชี้วัดพอร์ตโฟลิโอประกอบด้วยค่าจากมิติต่างๆ ดังนี้

1. ผลตอบแทน (Return) ประเมินด้วยอัตราผลตอบแทนโดยเฉลี่ยทบต้นต่อปี (Compounding Annual Growth Return หรือ CAGR)
2. ความเสี่ยง (Risk) ประเมินจากค่าความผันผวนของพอร์ตโฟลิโอ (Annualized Standard Deviation หรือ Ann. STD) และอัตราการถดถอยสูงสุดของพอร์ตโฟลิโอ (Maximum Drawdown หรือ Max. DD) เนื่องจากความเสี่ยงของการเงินดั้งเดิมนั้นจะอนุมานถึงความผันผวน (Volatility) อย่างไรก็ตาม Pardo มองว่าการลงทุนก็เหมือนกับการทำธุรกิจ ดังนั้น ความเสี่ยงในการลงทุนจึงหมายถึงความเสี่ยงในด้านของตัวเงินที่มีโอกาสจะขาดทุนสูงสุด [22] ดังตัวอย่างภาพที่ 14 แสดงถึงผลตอบแทนสะสมของดัชนี SET TRI ในช่วงปี ค.ศ. 2020 ถึงปี ค.ศ. 2021 พบว่า ในช่วงตอนต้นปี ค.ศ. 2020 ซึ่งเป็นช่วงการแพร่ระบาดของไวรัส COVID-19 ส่งผลให้ดัชนี SET TRI มีการปรับตัวลดลงมากที่สุดเท่ากับ -35.31% ดังนั้น หากนักลงทุนลงทุนในดัชนี SET TRI ก็จะพบช่วงระยะเวลาที่เงินลงทุนที่มีการปรับตัวลดลงมากถึง -35.31%
3. อัตราผลตอบแทนต่อความเสี่ยง (Risk-Adjusted Return/Risk Reward Ratio) ประเมินจากอัตราส่วน Sharpe Ratio ซึ่งถูกนำเสนอในปี ค.ศ. 1994 โดย William F. Sharpe [23] โดยคำนวณจากผลตอบแทนส่วนเกินจากอัตราการลงทุนที่ปราศจากความเสี่ยง (Risk free rate) หารด้วยความผันผวนของพอร์ตโฟลิโอ และอัตราส่วน MAR Ratio ซึ่งคำนวณจากอัตราผลตอบแทนโดยเฉลี่ยทบต้นต่อปีหารด้วยค่า Max. DD

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

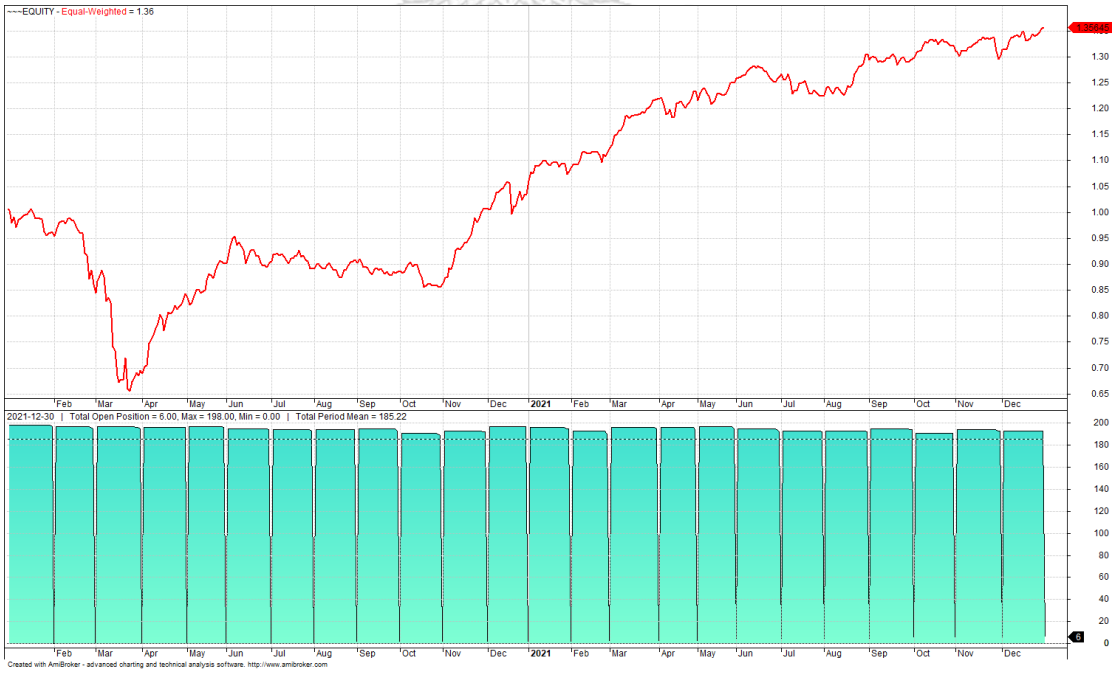


ภาพที่ 14 ผลตอบแทนสะสมของดัชนี SET TRI และค่าการถดถอยของพอร์ตโฟลิโอ

3.10 การประเมินผลพอร์ตโฟลิโอ (Portfolio Evaluation)

ผู้วิจัยได้ประเมินผลการสร้างพอร์ตโฟลิโอที่ลงทุนแบบพหุปัจจัยเปรียบเทียบกับสมรรถนะเกณฑ์มาตรฐาน (Benchmark) ประกอบด้วยดัชนีผลตอบแทนรวม SET TRI (SET Total Return Index), ดัชนีที่ถ่วงน้ำหนักแบบเท่าเทียมกัน (Equal-Weighted Index) และ ดัชนีผลตอบแทนตามปัจจัยต่างๆ (Factor Index หรือ Style Index) โดยสำหรับดัชนีผลตอบแทนรวม SET TRI (SET Total Return Index) นี้ถือเป็นดัชนีอ้างอิงของมาตรฐานอุตสาหกรรมการเงินซึ่งถูกคำนวณจากตลาดหลักทรัพย์แห่งประเทศไทยเพื่อใช้วัดผลตอบแทนรวมจากการลงทุนในหลักทรัพย์ ได้แก่ ผลตอบแทนที่เกิดจากการเปลี่ยนแปลงของราคา (Capital Gain/Loss) ซึ่งอาจเป็นได้ทั้งกำไรหรือขาดทุน และเงินปันผล (Dividends) โดยมีสมมติฐานเพิ่มเติมว่าเงินปันผลที่ได้รับนี้จะถูกนำไปลงทุนในหลักทรัพย์ด้วย (Reinvest) [24] อย่างไรก็ตาม เนื่องจากดัชนี SET TRI Index อ้างอิงงานวิจัย [25] ถูกคำนวณด้วยวิธีการถ่วงน้ำหนักตามมูลค่า ณ ราคาตลาด (Market Capitalization-weighted) ทำให้การเคลื่อนไหวของหลักทรัพย์ที่มีขนาด Market Cap. ใหญ่มีน้ำหนักมากกว่าหลักทรัพย์ที่มีขนาด Market Cap. เล็ก ดังนั้น เพื่อป้องกันอคติในการเปรียบเทียบกับดัชนี SET TRI Index ผู้วิจัยจึงเพิ่มดัชนีที่ถ่วงน้ำหนักแบบเท่าเทียมกัน (Equal-Weighted Index) เพื่อเป็นตัวแทนผลลัพธ์การลงทุนของหลักทรัพย์ทั้งหมด

สำหรับดัชนีที่ถ่วงน้ำหนักแบบเท่าเทียมกัน (Equal-Weighted Index) ผู้วิจัยสร้างจากพอร์ตโฟลิโอหมุนเวียนหลักทรัพย์รายเดือน โดยลงทุนบนหลักทรัพย์ทุกตัวเพื่อให้สะท้อนถึงผลลัพธ์การลงทุนของหลักทรัพย์ทั้งหมดในกลุ่มตัวอย่างงานวิจัยชิ้นนี้ และมีการกำหนดขนาดการลงทุนของหลักทรัพย์ทุกตัวอย่างเท่าเทียมกัน เช่น หากมีหลักทรัพย์ที่มีการซื้อขาย ณ ขณะนั้นทั้งหมด 200 ตัว ในการจำลองการซื้อหลักทรัพย์เข้าโฟลิโอจะซื้อหลักทรัพย์ทั้ง 200 ตัวนี้ ด้วยน้ำหนักการลงทุนตัวละ 0.5% ของพอร์ตโฟลิโอเพื่อให้สามารถซื้อหลักทรัพย์ได้ทั้งหมด เป็นต้น และมีการปรับน้ำหนักของการลงทุนในทุกๆเดือน (Monthly Rebalance) เพื่อป้องกันกรณีที่หลักทรัพย์บางตัวซึ่งมีการปรับตัวของราคาเพิ่มสูงขึ้นเป็นอย่างมากจนทำให้หลักทรัพย์ดังกล่าวมีขนาดน้ำหนักการลงทุนที่สูงเมื่อเปรียบเทียบกับหลักทรัพย์ตัวอื่นๆในพอร์ตโฟลิโอ หรือการเคลื่อนไหวของพอร์ตโฟลิโอเกิดจากการหลักทรัพย์เพียงบางกลุ่มเท่านั้น จากภาพที่ 15 พบว่า พอร์ตโฟลิโอมีจำนวนหลักทรัพย์ที่ถือครองหลักทรัพย์ (Total Open Position) โดยเฉลี่ยเท่ากับ 185.22 ตัวตลอดช่วงระยะเวลาปี ค.ศ. 2020 ถึง 2021 และในทุกๆปลายเดือนจำนวนหลักทรัพย์ที่ถือครองจะมีค่าลดลงเข้าใกล้ศูนย์ เนื่องจากพอร์ตโฟลิโอทำการขายหลักทรัพย์ที่มีทั้งหมดออกไปแล้วเข้าซื้อใหม่ตอนต้นเดือนเพื่อเป็นการปรับน้ำหนักการลงทุนใหม่ในทุกๆเดือน ส่งผลให้มีจำนวนครั้งของการซื้อขายหลักทรัพย์ทั้งหมดเท่ากับ 4,613 ครั้ง และมีค่าเฉลี่ยการถือครองหลักทรัพย์เท่ากับ 20.44 วันทำการหรือราวๆหนึ่งเดือน ดังแสดงในตารางที่ 4



ภาพที่ 15 ผลตอบแทนสะสมของ Equal-Weighted Portfolio และจำนวนหลักทรัพย์ที่ถือครอง

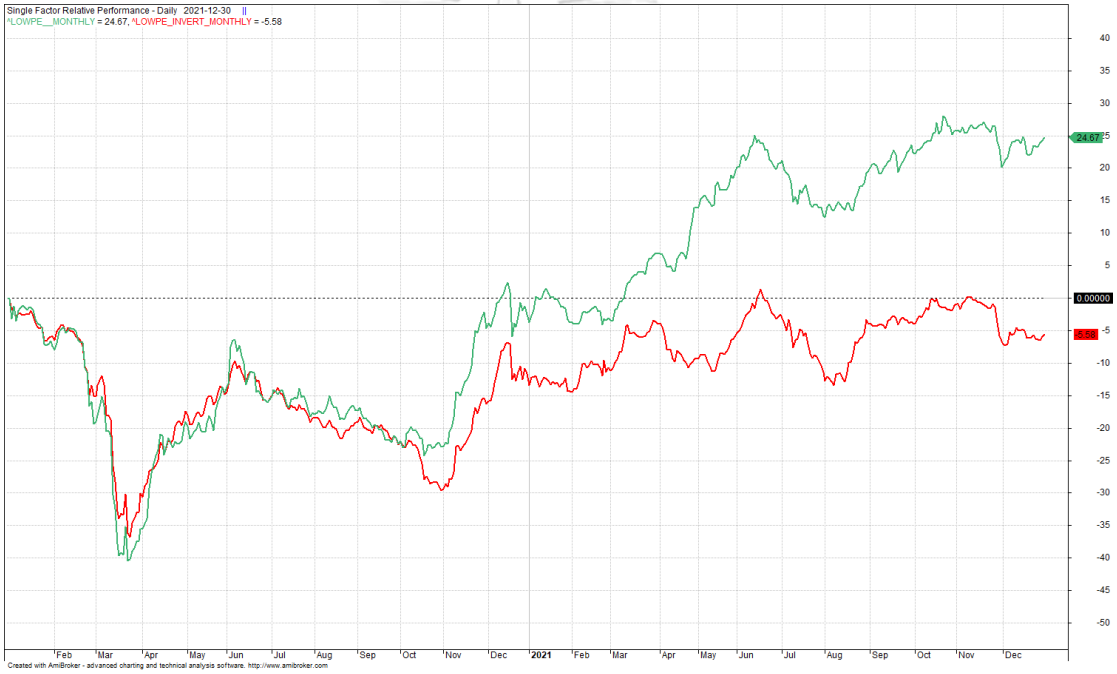
ตารางที่ 4 ค่าสถิติของพอร์ตโฟลิโอ Equal-Weighted

Portfolio statistics	**Equal-Weighted**
CAGR (%)	16.82
Max. Drawdown (%)	-34.9
Ann. STD (%)	19.95
Sharpe Ratio	0.64
MAR Ratio	0.48
Trade statistics	**Equal-Weighted**
#Trades	4,613
% Win	49.32
% Loss	50.68
Trade Expectancy %	1.59
Avg. Bar Held	20.44

นอกจากนี้ เพื่อเป็นการยืนยันว่าการคัดเลือกหลักทรัพย์แบบพหุปัจจัยด้วย XGBoost นั้นมีประสิทธิผลที่สูงกว่าการคัดเลือกหลักทรัพย์จากปัจจัยเพียงปัจจัยเดียว ดังนั้น ผู้วิจัยจึงเพิ่มดัชนีผลตอบแทนตามปัจจัยต่างๆ (Factor Index หรือ Style Index) ซึ่งเป็นวิธีการลงทุนอย่างเป็นระบบ (Systematic Approach) โดยจะทำการคัดเลือกหลักทรัพย์ที่มีคุณลักษณะหรือปัจจัยเหมือนๆกันมาสร้างพอร์ตโฟลิโอตามนิยามของ AQR [26] นั้น อย่างไรก็ตาม สำหรับการเปรียบเทียบเชิงการเงินนั้นมักจะนิยมสร้างเป็นพอร์ตโฟลิโอในลักษณะ (High & Low Portfolio) [5] เนื่องจากกลุ่มของหลักทรัพย์ที่มีค่าของปัจจัยที่สูงและต่ำนั้น มักจะพฤติกรรมของราคาที่แตกต่างกันจนส่งผลให้คุณลักษณะของพอร์ตโฟลิโอดังกล่าวมีความแตกต่างกันด้วย

สำหรับตัวอย่างของการสร้างพอร์ตโฟลิโอการลงทุนแบบปัจจัยเดียว เช่น การสร้างพอร์ตโฟลิโอจากปัจจัยเชิงคุณค่า (Value Factor) นิยามโดยอัตราส่วน Price-to-Earnings Ratio โดยผู้วิจัยจะทำการสร้างทั้งพอร์ตโฟลิโอที่ประกอบด้วยกลุ่มของหลักทรัพย์ที่มีอัตราส่วน PE สูง และกลุ่มของหลักทรัพย์ที่มีอัตราส่วน PE ต่ำ ดังภาพที่ 16 และตารางที่ 5 พบว่า ในช่วงระยะเวลาตั้งแต่ปี ค.ศ. 2020-2021 พอร์ตโฟลิโอทั้งสองมีการซื้อขายหลักทรัพย์ไปทั้งหมดราวๆ 450 ครั้ง และมีช่วงระยะเวลาการถือหลักทรัพย์โดยเฉลี่ยประมาณ 20 วันทำการหรือหนึ่งเดือน ซึ่งค่าเฉลี่ยผลตอบแทนคาดหวังต่อการซื้อขายหนึ่งของครั้งพอร์ตโฟลิโอ Low PE มีค่าที่สูงกว่าพอร์ตโฟลิโอ High PE ซึ่งมีค่า

เท่ากับ 1.43% และ 0.17% ตามลำดับ ส่งผลให้ผลตอบแทนโดยเฉลี่ยทบต้นของพอร์ตโฟลิโอ Low PE มีค่าที่สูงถึง 12.98% ในขณะที่พอร์ตโฟลิโอ High PE นั้นกลับให้ผลตอบแทนทบต้นต่อปีติดลบเท่ากับ -2.26% อย่างไรก็ตาม ในด้านความเสี่ยงของพอร์ตโฟลิโอกลับพบว่าพอร์ตโฟลิโอ High PE มีค่าความเสี่ยงที่ต่ำกว่ากล่าวคือ พอร์ตโฟลิโอ High PE มีค่าความผันผวนและอัตราการถดถอยสูงสุดของพอร์ตโฟลิโอเท่ากับ 20.59% และ -36.8% ตามลำดับ ในขณะที่พอร์ตโฟลิโอ High PE กลับมีค่าความผันผวนและอัตราการถดถอยสูงสุดของพอร์ตโฟลิโอที่มากกว่าพอร์ตโฟลิโอ High PE เป็น โดยผู้วิจัยทำการสร้างพอร์ตโฟลิโอการลงทุนแบบปัจจัยเดียวในลักษณะ High และ Low นี้จากปัจจัยตั้งต้นทั้งหมด 27 ปัจจัย ทำให้ได้พอร์ตโฟลิโอการลงทุนปัจจัยเดียวทั้งหมด 54 พอร์ตโฟลิโอสำหรับเป็นดัชนีอ้างอิงในการประเมินผลลัพธ์ของพอร์ตโฟลิโอ



ภาพที่ 16 ผลตอบแทนสะสมของพอร์ตโฟลิโอ Low PE (เส้นเขียว) และ High PE (เส้นแดง)

ตารางที่ 5 ค่าสถิติของพอร์ตโฟลิโอ Low PE และ High PE

Portfolio statistics	**High PE**	**Low PE**
CAGR (%)	-2.26	12.98
Max. Drawdown (%)	-36.8	-40.4
Ann. STD (%)	20.59	23.86
Sharpe Ratio	-0.3	0.38
MAR Ratio	0.32	0.32
Trade statistics	**High PE**	**Low PE**
#Trades	455	453
% Win	47.69	46.14
% Loss	52.31	53.86
Trade Expectancy %	0.17	1.43
Avg. Bar Held	20.21	20.34

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 4
ผลลัพธ์และการอภิปรายผล

4.1 ประเมินสมรรถนะแบบจำลอง

ในส่วนนี้เป็นการประเมินสมรรถนะแบบจำลองการคัดเลือกหลักทรัพย์แบบรายเดือนและรายไตรมาสจากพหุปัจจัยด้วย XGBoost บนชุดข้อมูลตรวจสอบ (Validation set) หรือช่วงตั้งแต่ปี ค.ศ. 2012 ถึงปี ค.ศ. 2019 และบนชุดข้อมูลทดสอบ (Test set) หรือช่วงตั้งแต่ปี ค.ศ. 2020 ถึงปี ค.ศ. 2021 โดยมีค่า ROC Curve (Receiver Operating Characteristics Curve) และ AUC (Area under the ROC Curve) เป็นดัชนีชี้วัด โดยมีรายละเอียดดังต่อไปนี้

4.1.1 ประเมินสมรรถนะแบบจำลองบนชุดข้อมูลตรวจสอบ

ในระหว่างการสอนให้โมเดลเรียนรู้ความสัมพันธ์ของปัจจัยต่างๆที่มีต่อผลตอบแทนในอนาคตของหลักทรัพย์ ผู้วิจัยได้นำกระบวนการค้นหาแบบกริดโดยมีการแบ่งข้อมูลออกเป็น 3 พับ (Grid Search with 3-fold Cross Validation) มาประยุกต์ใช้เพื่อหาค่าไฮเปอร์พารามิเตอร์ที่เหมาะสมของแบบจำลอง พบว่า ค่าความลึกที่สุดของต้นไม้ (max_depth) เป็นพารามิเตอร์ที่มีอิทธิพลต่อการจำแนกมากที่สุด และมีค่าความลึกที่เหมาะสมของแบบจำลองการทำนายผลตอบแทนของหลักทรัพย์ทั้งแบบรายเดือนและรายไตรมาสเท่ากันคือ มีค่าเท่ากับ 8 โดยมีค่าไฮเปอร์พารามิเตอร์อื่นๆของแบบจำลองการคัดเลือกหลักทรัพย์ทั้งสอง ดังตารางที่ 6

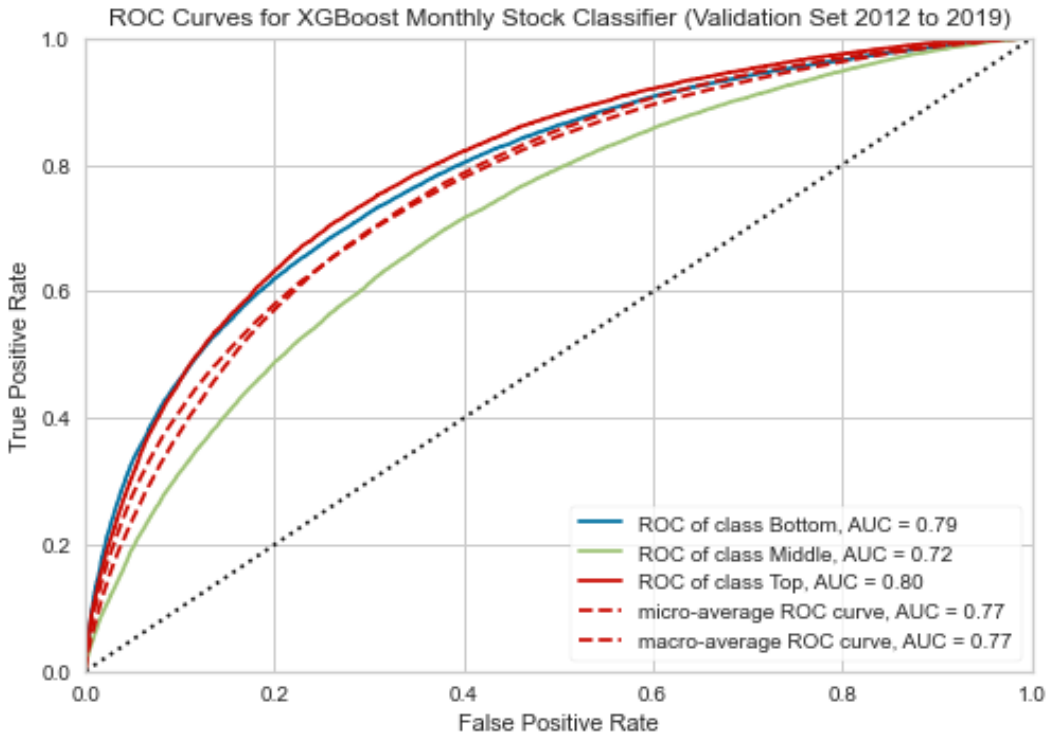
ตารางที่ 6 ค่าไฮเปอร์พารามิเตอร์ของแบบจำลองการคัดเลือกหลักทรัพย์รายเดือนและรายไตรมาส

CHULALONGKORN UNIVERSITY

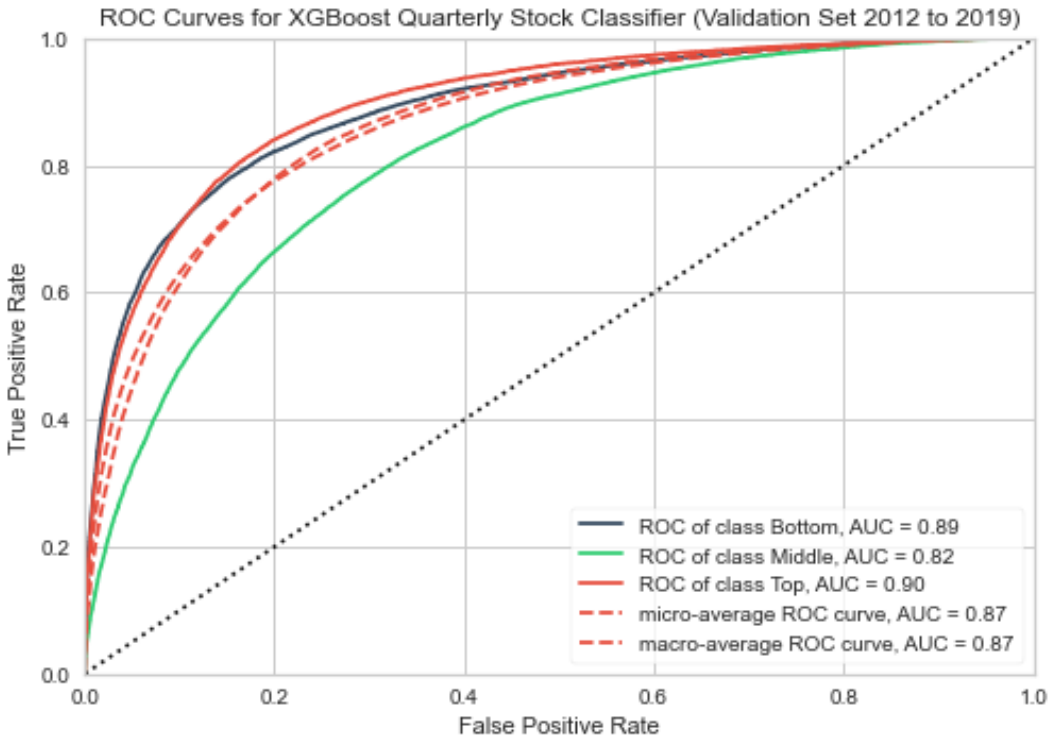
	Monthly Stock-Selection	Quarterly Stock-Selection
eta	0.01	0.01
n_estimators	700	625
max_depth	8	8
min_child_weight	1	1
gamma	0.2	0.2
colsample_bytree	0.6	0.6
colsample_bylevel	0.6	0.6
colsample_bynode	0.6	0.6

สำหรับการประเมินสมรรถนะแบบจำลองการทำนายผลตอบแทนในอนาคตหนึ่งเดือน พบว่า การทำนายกลุ่มหลักทรัพย์ที่จะให้ผลตอบแทนมากที่สุดในอนาคต (Top-Class), กลุ่มที่ให้ผลตอบแทนตรงกลาง (Middle-Class), และกลุ่มที่ให้ผลตอบแทนน้อยที่สุด (Bottom-Class) มีค่าพื้นที่ใต้กราฟของ ROC (AUC ROC) เท่ากับ 0.80, 0.72 และ 0.79 ตามลำดับ ส่งผลให้ค่าเฉลี่ยการทำนายในภาพรวมซึ่งวัดโดยค่า macro-average ROC Curve หรือค่าเฉลี่ยแบบไม่มีการถ่วงน้ำหนักตามสัดส่วนของกลุ่มตัวอย่างมีค่าเท่ากับ 0.77 และค่า micro-average ROC Curve หรือค่าเฉลี่ยแบบมีการถ่วงน้ำหนักตามสัดส่วนของกลุ่มตัวอย่างมีค่าเท่ากับ 0.77 ดังภาพที่ 17

อย่างไรก็ตาม สำหรับสมรรถนะแบบจำลองการทำนายผลตอบแทนในอนาคตหนึ่งไตรมาส พบว่า เส้นโค้งของการทำนายผลตอบแทนของหลักทรัพย์ทั้งสามกลุ่มมีการปรับตัวเพิ่มขึ้นอย่างรวดเร็วและเส้นโค้งจะอยู่ใกล้กับขอบบนของกราฟมากกว่าเมื่อเปรียบเทียบกับการทำนายผลตอบแทนแบบรายเดือน ส่งผลให้พื้นที่ใต้กราฟของแบบจำลองการทำนายผลตอบแทนของหลักทรัพย์แบบรายไตรมาสมีค่าที่สูงกว่าการทำนายผลตอบแทนของหลักทรัพย์แบบรายเดือน โดยมีค่าพื้นที่ใต้กราฟของ ROC (AUC ROC) ของการทำนายผลตอบแทนรายไตรมาสทั้งสามกลุ่มประกอบด้วย Top-Class, Middle-Class และ Bottom-Class มีค่าเท่ากับ 0.90, 0.82 และ 0.89 ตามลำดับ และมีค่าเฉลี่ยการทำนายในภาพรวมซึ่งวัดโดยค่า macro-average และ micro-average ROC Curve เท่ากับ 0.87 และ 0.87 ดังภาพที่ 18



ภาพที่ 17 AUC ROC การทำนายผลตอบแทนหลักทรัพย์ในอนาคตหนึ่งเดือนบน Validation set

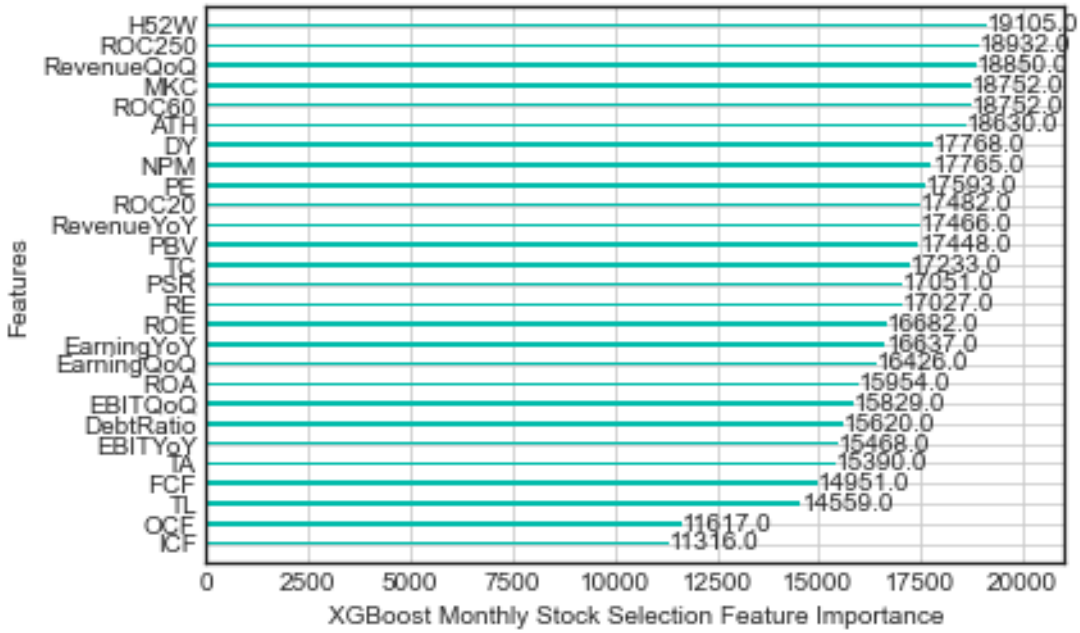


ภาพที่ 18 AUC ROC การทำนายผลตอบแทนหลักทรัพย์ในอนาคตหนึ่งไตรมาสบน Validation set

4.1.2 ปัจจัยที่สำคัญสำหรับการทำนายผลตอบแทนของหลักทรัพย์

ในการพิจารณาปัจจัยที่สำคัญสำหรับการทำนายผลตอบแทนของหลักทรัพย์ ผู้วิจัยใช้คำสั่ง plot_importance จากไลบรารี XGBoost บนภาษา Python โดยผลลัพธ์ที่ได้จะเป็นกราฟแสดงความสำคัญของปัจจัยต่างๆ ซึ่งปัจจัยที่มีค่าความสำคัญสูงหมายความว่าปัจจัยดังกล่าวมีอิทธิพลในการทำนายที่สูงตามไปด้วย โดยจากภาพที่ 19 พบว่า ปัจจัยที่มีความสำคัญสูงที่สุดสองอันดับแรกในการทำนายผลตอบแทนแบบรายเดือนคือ อัตราส่วน 52 Weeks High และผลตอบแทนในอดีตของหลักทรัพย์คำนวณย้อนหลัง 250 วันทำการหรือหนึ่งปี (ROC250) ตามด้วยปัจจัยด้านการเติบโตของกิจการ วัดจากการเติบโตของรายได้รวมย้อนหลังสี่ไตรมาสเปรียบเทียบกับไตรมาสก่อนหน้า (Rolling twelve-month revenue growth Quarter-on-Quarter), ปัจจัยด้านขนาดของหลักทรัพย์ วัดจากมูลค่า ณ ราคาตลาดของหลักทรัพย์ (Stock Market Capitalization), และปัจจัยด้านแนวโน้มของราคา วัดจากผลตอบแทนในอดีตของหลักทรัพย์คำนวณย้อนหลัง 60 วันทำการหรือหนึ่งไตรมาส (ROC60) และอัตราส่วน All Time High Ratio (ATH) โดยจะสังเกตได้ว่าปัจจัยที่มีความสำคัญสูงสุด 6 อันดับแรกในการทำนายผลตอบแทนแบบรายเดือนประกอบไปด้วยปัจจัยด้านแนวโน้มของราคาซึ่งจัดเป็นปัจจัยเชิงเทคนิคทั้งหมด 4 ปัจจัย หรือสามารถสรุปได้ว่าปัจจัยเชิงเทคนิคมีผลต่อการทำนายแนวโน้มผลตอบแทนในอนาคตได้เป็นอย่างดี

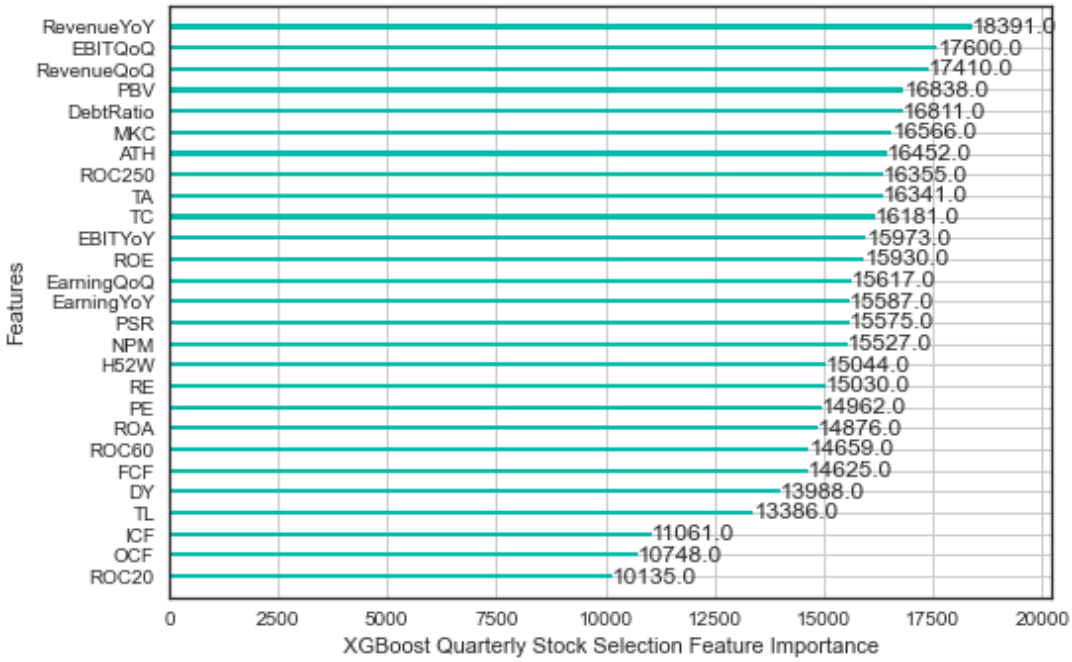
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 19 ปัจจัยที่มีอิทธิพลต่อการทำนายผลตอบแทนหลักทรัพย์หนึ่งเดือนข้างหน้า

ทั้งนี้ สำหรับการทำนายผลตอบแทนหนึ่งไตรมาสข้างหน้ากลับพบว่า ปัจจัยงบการเงินมีอิทธิพลในการทำนายมากกว่าปัจจัยเชิงเทคนิค โดยจากภาพที่ 20 พบว่า ปัจจัยที่มีความสำคัญมากที่สุดในการทำนายผลตอบแทนรายไตรมาสสามอันดับแรกล้วนแล้วแต่เป็นปัจจัยด้านการเติบโตของกิจการ วัดจากการเติบโตของรายได้รวมย้อนหลัง 4 ไตรมาสเปรียบเทียบกับปีก่อนหน้า (Rolling twelve-month revenue growth Year-on-Year), การเติบโตของรายได้ก่อนหักดอกเบี้ยและภาษีรวมย้อนหลัง 4 ไตรมาสเปรียบเทียบกับไตรมาสก่อนหน้า (Rolling twelve-month EBIT growth Quarter-on-Quarter), และการเติบโตของรายได้รวมย้อนหลังสี่ไตรมาสเปรียบเทียบกับไตรมาสก่อนหน้า (Rolling twelve-month revenue growth Quarter-on-Quarter) ตามมาด้วยปัจจัยเชิงคุณค่า วัดจากอัตราส่วนราคาของหลักทรัพย์เมื่อเปรียบเทียบกับมูลค่าทางบัญชี (Price-to-Book Value หรือ PBV) และอัตราส่วนหนี้สิน (Debt Ratio) ซึ่งเป็นสัดส่วนหนี้สินเมื่อเปรียบเทียบกับสินทรัพย์ของบริษัท และปัจจัยด้านขนาดของหลักทรัพย์ วัดจากมูลค่า ณ ราคาตลาดของหลักทรัพย์ (Stock Market Capitalization) โดยจะสังเกตได้ว่าปัจจัยที่มีความสำคัญสำหรับการทำนายผลตอบแทนในอนาคตหนึ่งไตรมาสนี้ล้วนแล้วแต่เป็นปัจจัยงบการเงินด้วยกันทั้งสิ้น โดยจากหลักฐานดังกล่าวทำให้สามารถสรุปได้ว่าแนวโน้มของราคาสินทรัพย์ในช่วงรายไตรมาสมักจะขึ้นกับปัจจัยพื้นฐาน (Fundamental factor) มากกว่าปัจจัยเชิงเทคนิค (Technical factor)

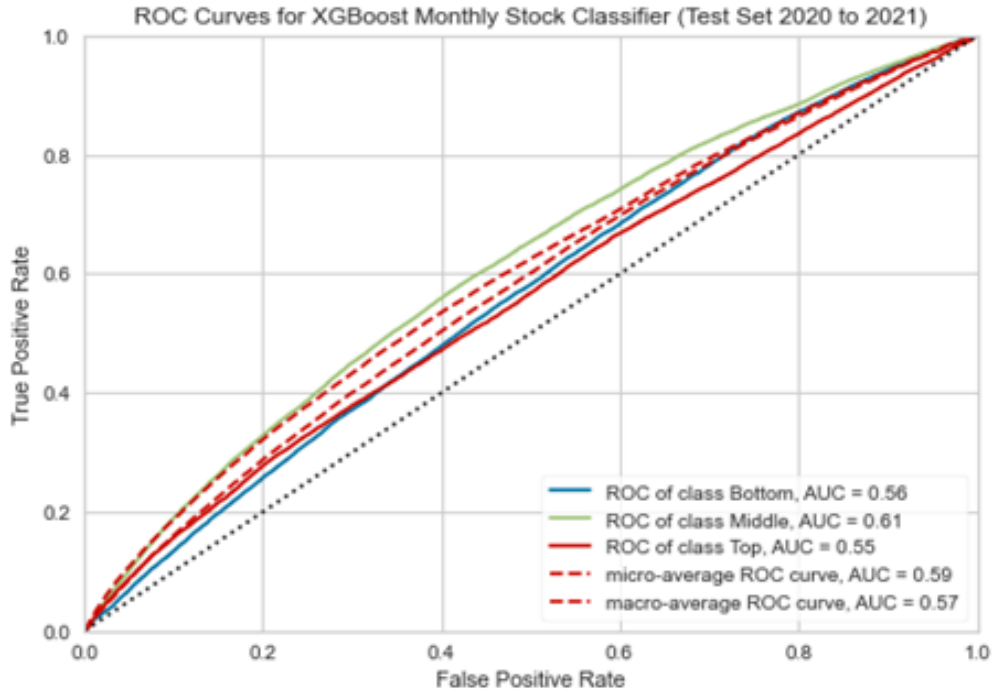
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



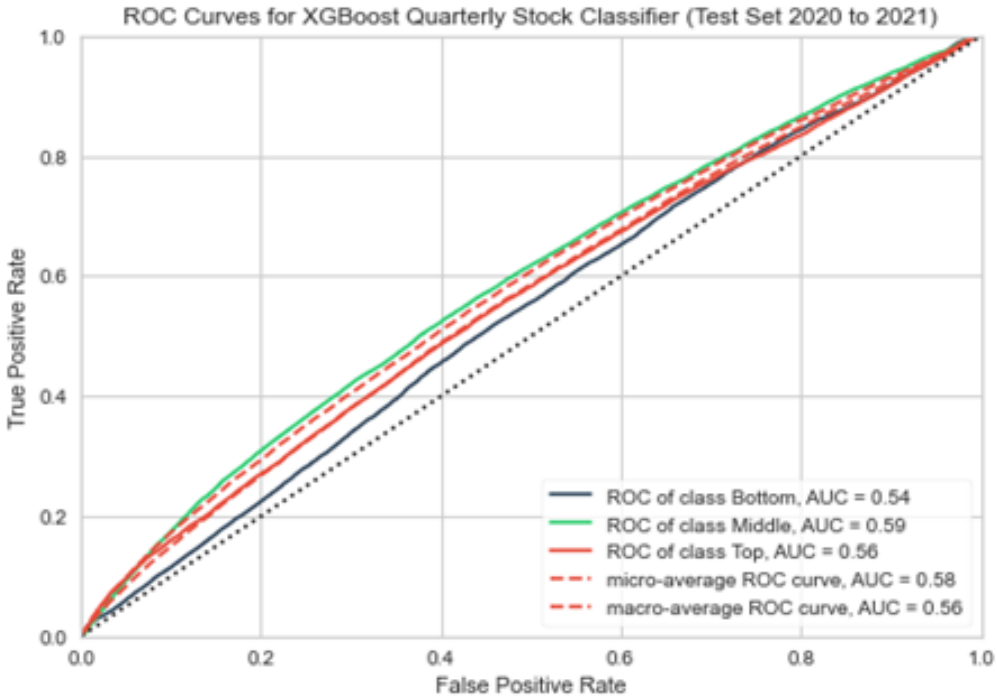
ภาพที่ 20 ปัจจัยที่มีอิทธิพลต่อการทำนายผลตอบแทนหลักทรัพย์หนึ่งไตรมาสข้างหน้า

4.1.3 สมรรถนะการทำนายหลักทรัพย์บนชุดข้อมูลทดสอบ

สำหรับการประเมินสมรรถนะการทำนายหลักทรัพย์บนชุดข้อมูลทดสอบ (ปี ค.ศ. 2020 ถึงปี ค.ศ. 2021) พบว่า สมรรถนะการจำแนกผลตอบแทนในอนาคตหนึ่งเดือนและหนึ่งไตรมาสนี้มีค่าลดลงเมื่อเปรียบเทียบกับการประเมินบนชุดมูลตรวจสอบ (Validation set) เนื่องจากภาวะความผันผวนของตลาดมีที่เพิ่มสูงขึ้นจากการแพร่ระบาดของไวรัส COVID-19 ทำให้ดัชนี SET TRI มีการปรับตัวลดลงอย่างรุนแรงและรวดเร็วในช่วงไตรมาสที่หนึ่งของปี ค.ศ. 2020 ก่อนที่ดัชนีจะมีการปรับตัวเพิ่มสูงขึ้นจากภาวะเศรษฐกิจที่เริ่มมีการฟื้นตัวในช่วงปลายปี ค.ศ. 2020 จนถึงปี ค.ศ. 2021 ซึ่งภาวะความผันผวนนี้ทำให้สมรรถนะการทำนายผลตอบแทนในอนาคตหนึ่งเดือนข้างหน้าวัดโดยค่าเฉลี่ย micro-average และ macro-average มีค่าเท่ากับ 0.59 และ 0.57 ตามลำดับดังภาพที่ 21 และค่าเฉลี่ย micro-average และ macro-average ของการทำนายผลตอบแทนในอนาคตหนึ่งไตรมาสมีค่าเท่ากับ 0.58 และ 0.56 ตามลำดับดังภาพที่ 22 อย่างไรก็ตาม แบบจำลองยังคงให้ผลลัพธ์จากทำนายผลตอบแทนที่ดีในช่วงระยะเวลาของการแพร่ระบาดไวรัส COVID-19 ดังกล่าว



ภาพที่ 21 AUC ROC การทำนายผลตอบแทนหลักทรัพย์ในอนาคตหนึ่งเดือนบน Test set



ภาพที่ 22 AUC ROC การทำนายผลตอบแทนหลักทรัพย์ในอนาคตหนึ่งไตรมาสบน Test set

4.2 ประเมินผลลัพธ์ของพอร์ตโฟลิโอการลงทุน

สำหรับการประเมินผลลัพธ์พอร์ตโฟลิโอการลงทุนแบบหมุนเวียนจากการคัดเลือกหลักทรัพย์ด้วย XGBoost ประกอบด้วยพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบรายเดือนและรายไตรมาส ซึ่งกฎในการคัดเลือกหลักทรัพย์เข้าพอร์ตโฟลิโอคือ คัดเลือกหลักทรัพย์ที่มีโอกาสจะให้ผลตอบแทนสูงสุดมาเป็นจำนวนทั้งหมด 10% ของหลักทรัพย์ที่มีอยู่ในขณะนั้น และมีการกำหนดน้ำหนักการลงทุนของหลักทรัพย์แต่ละตัวอย่างเท่าเทียมกัน โดยมีผลลัพธ์การประเมิน ดังต่อไปนี้

4.2.1 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์รายเดือนแบบพหุปัจจัยด้วย XGBoost กับพอร์ตโฟลิโอรายไตรมาส

ในส่วนแรกนี้เป็นการเปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์รายเดือนและรายไตรมาสแบบพหุปัจจัยด้วย XGBoost พบว่า ผลตอบแทนของพอร์ตโฟลิโอทั้งสองมีการเติบโตขึ้นอย่างต่อเนื่องตลอดช่วงระยะเวลาของการทดสอบ ดังแสดงในภาพที่ 23 โดยพอร์ตโฟลิโอแบบรายเดือนมีผลตอบแทนโดยเฉลี่ยทบต้นต่อปี (CAGR) ที่สูงกว่าพอร์ตโฟลิโอแบบรายไตรมาสโดยมีค่าเท่ากับ 48.19% และ 30.96% ตามลำดับ ดังที่แสดงในตารางที่ 7 ซึ่งจะมีเพียงช่วงต้นปี ค.ศ. 2020 ที่เป็นช่วงจากการแพร่ระบาดระบาดของไวรัส COVID-19 ทำให้ผลตอบแทนของพอร์ตโฟลิโอทั้งสองมีการปรับตัวลดลงจนมีค่า Max. Drawdown เท่ากับ -37.16% และ -42.18% โดยการที่พอร์ตโฟลิโอแบบรายเดือนมีระดับผลตอบแทนและค่าความเสี่ยงที่ดีกว่า เนื่องจากพอร์ตโฟลิโอแบบรายเดือนมีโอกาสในการคัดเลือกหลักทรัพย์ใหม่ๆเข้ามาในพอร์ตโฟลิโอได้มากกว่าพอร์ตโฟลิโอแบบรายไตรมาสซึ่งจำเป็นจะต้องรอให้ถือครองหลักทรัพย์ครบหนึ่งไตรมาสจึงจะมีการหมุนเวียนหลักทรัพย์ที่ลงทุนได้ และจากเหตุผลที่กล่าวมาข้างต้นนี้ ทำให้อัตราผลตอบแทนต่อความเสี่ยงวัดโดยอัตราส่วน Sharpe และอัตราส่วน MAR ของพอร์ตโฟลิโอแบบรายเดือนมีค่าที่ดีกว่าพอร์ตโฟลิโอแบบรายไตรมาส

สำหรับค่าสถิติของการซื้อขาย (Trade Statistics) พบว่า พอร์ตโฟลิการลงทุนแบบรายเดือนมีจำนวนการซื้อขายที่มากกว่าพอร์ตโฟลิโอรายไตรมาสโดยมีจำนวนการซื้อขายเท่ากับ 439 และ 149 ครั้งตามลำดับ โดยถึงแม้ว่าสัดส่วนของเทรดที่กำไรและขาดทุนของทั้งพอร์ตโฟลิโอจะอยู่ในระดับที่ใกล้เคียงกัน แต่เมื่อพิจารณาถึงค่าเฉลี่ยผลกำไร/ขาดทุนของทุกการซื้อขาย (Trade Expectancy) กลับพบว่าพอร์ตโฟลิโอแบบรายไตรมาสมีค่าที่สูงกว่าพอร์ตโฟลิโอแบบรายเดือนอย่างมีนัยยะ เนื่องจากระยะเวลาการถือครองหลักทรัพย์ที่นานขึ้น (Avg. Bar Held) จะสามารถทำให้ได้รับผลตอบแทนที่สูงกว่าการถือครองหลักทรัพย์ในช่วงระยะเวลาสั้นๆ นอกจากนี้ ค่าเฉลี่ยระยะเวลาการถือครองหลักทรัพย์ ที่ 20.57 (รายเดือน) และ 61.85 (รายไตรมาส) เป็นการยืนยันระยะเวลาเฉลี่ยตั้งแต่ซื้อจนถึงขายของพอร์ตโฟลิโอทั้งหมด นั่นคือ 20 วันทำการและ 60 วันทำการสำหรับการซื้อขายแบบหมุนเวียนรายเดือนและรายไตรมาส

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาพที่ 23 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือน (เส้นเขียว) และแบบรายไตรมาส (เส้นน้ำเงิน)

ตารางที่ 7 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือนและแบบรายไตรมาส

Portfolio statistics	**Monthly Top-Class**	**Quarterly Top-Class**
CAGR (%)	**48.19**	30.96
Max. Drawdown (%)	**-37.16**	-42.17
Ann. STD (%)	**25.16**	25.68
Sharpe Ratio	**1.76**	1.05
MAR Ratio	**1.3**	0.73
Trade statistics	**Monthly Top-Class**	**Quarterly Top-Class**
#Trades	439	149
% Win	52.16	**54.36**
% Loss	47.84	**45.64**
Trade Expectancy %	3.78	**9.49**
Avg. Bar Held	20.57	61.85

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

โดยตัวอย่างของหลักทรัพย์ที่สร้างผลตอบแทนที่สูงให้กับพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบรายเดือน เช่น บริษัท ศรีตรังแอโกรอินดัสทรี จำกัด (มหาชน) หรือ “STA” ซึ่งประกอบธุรกิจผลิตและจัดจำหน่ายยางธรรมชาติแบบครบวงจร โดยถือเป็นหนึ่งในกลุ่มธุรกิจที่ได้ประโยชน์จากการแพร่ระบาดของไวรัส COVID-19 เนื่องจากเป็นสินค้าที่มีความต้องการสูงในช่วงระยะเวลาดังกล่าว โดยภายหลังจากที่แบบจำลองได้มีการแนะนำให้ซื้อหลักทรัพย์ดังกล่าวในช่วงต้นเดือนกุมภาพันธ์ ค.ศ. 2020 พบว่า ราคาของหลักทรัพย์ได้มีการปรับตัวเพิ่มขึ้นโดยมีผลตอบแทนภายในระยะเวลาหนึ่งเดือนเท่ากับ 81.97% ดังแสดงในภาพที่ 24 รวมถึงบริษัท เอเชีย เอวิเอชั่น จำกัด (มหาชน) หรือ “AAV” ซึ่งเป็นบริษัทที่อยู่ในกลุ่มอุตสาหกรรมท่องเที่ยว โดยถือเป็นกลุ่มอุตสาหกรรมที่ได้รับผลกระทบโดยตรงในช่วงการแพร่ระบาดของไวรัส COVID-19 จนทำให้ราคาของหลักทรัพย์มีการปรับตัวลดลงอย่างต่อเนื่องในช่วงต้นปี ค.ศ. 2020 จนราคามาทำจุดต่ำสุดหรือมีสัญญาณการหยุดตกลงของราคาในช่วงเดือนมีนาคม ค.ศ. 2020 ทำให้แบบจำลองแนะนำให้ซื้อหลักทรัพย์ดังกล่าวในช่วงต้นเดือนเมษายน ค.ศ. 2020 โดยหลังจากนั้นราคาของหลักทรัพย์ดังกล่าวมีการปรับตัวเพิ่มขึ้นเท่ากับ 75% ภายในระยะเวลาหนึ่งเดือน ดังแสดงในภาพที่ 25 เป็นต้น

ภาพที่ 24 การเคลื่อนไหวของราคาบริษัท ศรีตรังแอโกรอินดัสทรี จำกัด (มหาชน)

ภาพที่ *25* การเคลื่อนไหวของราคาบริษัท เอเซีย เอวิเอชั่น จำกัด (มหาชน)

ตัวอย่างของหลักทรัพย์ที่สร้างผลตอบแทนที่สูงให้กับพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบรายไตรมาส เช่น บริษัทเดลต้า อีเลคโทรนิคส์ (ประเทศไทย) จำกัด (มหาชน) หรือ “DELTA” และบริษัท ทีคิวเอ็ม อัลฟา จำกัด (มหาชน) หรือ “TQM” ซึ่งเป็นบริษัทในกลุ่มธุรกิจประกันภัยและอิเล็กทรอนิกส์ตามลำดับ โดยกลุ่มธุรกิจทั้งสองล้วนแล้วแต่เป็นกลุ่มธุรกิจที่มีการเติบโตภายหลังการแพร่ระบาดของไวรัส COVID-19 ส่งผลให้ราคาของหลักทรัพย์มีการปรับตัวเพิ่มขึ้น 92.86% และ 206.31% ภายในระยะเวลาหนึ่งไตรมาสดังแสดงในภาพที่ 26 และ 27 เป็นต้น

ภาพที่ 26 การเคลื่อนไหวของราคาบริษัทเดลต้า อีเลคโทรนิคส์ (ประเทศไทย) จำกัด (มหาชน)

ภาพที่ 27 การเคลื่อนไหวของราคาบริษัท ทีคิวเอ็ม อัลฟา จำกัด (มหาชน)

จากตัวอย่างการคัดเลือกหลักทรัพย์ที่กล่าวมาข้างต้น ประกอบกับการกระจายความเสี่ยงในการลงทุนที่เหมาะสมคือ ลงทุนในหลักทรัพย์จำนวนทั้งหมด 10% ของหลักทรัพย์ที่มีอยู่ในขณะนั้น จึงทำให้พอร์ตโฟลิโอคัดเลือกหลักทรัพย์ทั้งแบบรายเดือนและรายไตรมาสสามารถสร้างผลตอบแทนที่เติบโตอย่างต่อเนื่องได้ตลอดทั้งช่วงระยะเวลาการทดสอบ

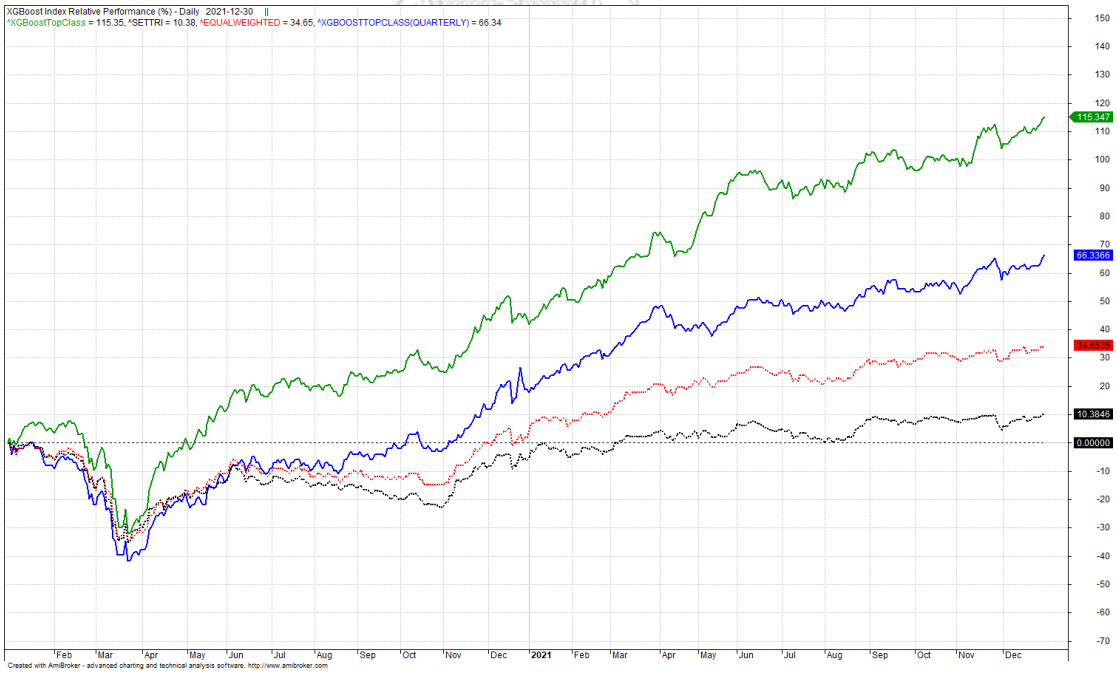
4.2.2 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์รายเดือนและรายไตรมาสแบบพหุปัจจัยด้วย XGBoost กับดัชนีอ้างอิง

สำหรับการเปรียบเทียบผลตอบแทนของพอร์ตโฟลิโอคัดเลือกหลักทรัพย์รายเดือนและรายไตรมาสแบบพหุปัจจัยด้วย XGBoost กับดัชนีอ้างอิงประกอบไปด้วยดัชนี SET TRI และดัชนีที่ให้น้ำหนักการลงทุนอย่างเท่าเทียม (Equal-Weighted Index) จากภาพที่ 28 และตารางที่ 8 พบว่าผลตอบแทนของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบพหุปัจจัยนี้ยังคงให้ระดับผลตอบแทนที่สูงกว่าดัชนีทั้งสองซึ่งมีระดับผลตอบแทนโดยเฉลี่ยทบต้นต่อปีเท่ากับ 5.28% และ 16.82% ตามลำดับ ทั้งนี้ การที่ผลตอบแทนของดัชนีที่ให้น้ำหนักการลงทุนอย่างเท่าเทียมกันมีระดับผลตอบแทนที่สูงกว่าดัชนี SET TRI เป็นเพราะดัชนี SET TRI ถูกคำนวณแบบถ่วงน้ำหนัก (Market Capitalization Weighted) ทำให้หลักทรัพย์ที่มีขนาดราคา ณ มูลค่าตลาดสูงมีอิทธิพลต่อการเคลื่อนไหวของดัชนีมากกว่าหรือดัชนีสะท้อนผลตอบแทนของกลุ่มหลักทรัพย์ขนาดใหญ่ อย่างไรก็ตาม เนื่องจากกลุ่มตัวอย่างของงานวิจัยนี้ประกอบด้วยหลักทรัพย์ขนาดใหญ่และขนาดกลางของประเทศไทย ดังนั้น เมื่อทำการสร้างพอร์ตโฟลิโอที่กำหนดน้ำหนักการลงทุนของหลักทรัพย์ทุก

ตัวอย่างเท่าเทียมกันจึงทำให้สะท้อนภาพรวมผลตอบแทนของหลักทรัพย์ได้ดีกว่าดัชนี SET TRI ซึ่งมักสะท้อนผลตอบแทนเฉพาะกลุ่มหลักทรัพย์ขนาดใหญ่

ด้านความเสี่ยงวัดโดยอัตราการถดถอยสูงสุดของพอร์ตโฟลิโอ (Max. Drawdown) พบว่า ในช่วงต้นปี ค.ศ. 2020 ซึ่งเป็นช่วงที่มีการแพร่ระบาดของไวรัส COVID-19 ส่งผลให้ดัชนี SET TRI Index และดัชนีให้น้ำหนักการลงทุนอย่างเท่าเทียมมีการปรับตัวลดลงเท่ากับ -35.31% และ -34.9% โดยพอร์ตโฟลิโอการลงทุนแบบรายเดือนนี้มีการปรับตัวลดลงในระดับที่ใกล้เคียงกับดัชนีอ้างอิงทั้งสอง และอยู่ในระดับที่ต่ำกว่าพอร์ตโฟลิโอการลงทุนแบบรายไตรมาสเพราะพอร์ตโฟลิโอการลงทุนแบบรายเดือนสามารถหมุนเวียนหลักทรัพย์ได้ในทุกๆเดือนตลอดช่วงระยะเวลาของการแพร่ระบาด อย่างไรก็ตาม พบว่า พอร์ตโฟลิโอการลงทุนทั้งสองมีระดับความผันผวนที่สูงกว่าดัชนีอ้างอิง เนื่องจากค่าความผันผวนของพอร์ตโฟลิโอถูกคำนวณจากการปรับตัวเพิ่มขึ้นและลดลงของพอร์ตโฟลิโอ ดังนั้น การที่พอร์ตโฟลิโอมีการเติบโตในระดับที่สูงอย่างต่อเนื่องจึงทำให้ค่าความผันผวนสูงขึ้นด้วย

เนื่องจากภาพรวมผลตอบแทนของพอร์ตโฟลิโอทั้งสองอยู่ในระดับที่สูงกว่าดัชนีอ้างอิงอย่างมีนัยยะ ซึ่งถึงแม้ว่าค่าความเสี่ยงจะอยู่ในระดับที่ใกล้เคียงกันกับดัชนีอ้างอิง แต่เมื่อพิจารณาถึงอัตราผลตอบแทนต่อความเสี่ยงทั้งอัตราส่วน Sharpe และอัตราส่วน MAR จึงส่งผลให้ค่าของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost ทั้งสองนี้มีค่าที่สูงกว่าดัชนีอย่างมีนัยยะ



ภาพที่ 28 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือน (เส้นเขียว), รายไตรมาส (เส้นน้ำเงิน), ดัชนี SET TRI (เส้นประสีดำ) และดัชนี Equal-Weighted (เส้นประสีแดง)

ตารางที่ 8 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือน, รายไตรมาส ระหว่าง XGBoost, ดัชนี SET TRI และดัชนี Equal-Weighted

Portfolio statistics	Monthly Top-Class	Quarterly Top-Class	Equal-Weighted	SET TRI
CAGR (%)	**48.19**	30.96	16.82	5.28
Max. Drawdown (%)	-37.16	-42.17	**-34.9**	-35.31
Ann. STD (%)	25.16	25.68	**19.95**	22.81
Sharpe Ratio	**1.76**	1.05	0.64	0.06
MAR Ratio	**1.3**	0.73	0.48	0.15
Trade statistics	**Monthly Top-Class**	**Quarterly Top-Class**	**Equal-Weighted**	**SET TRI**
#Trades	439	149	4,613	N/A
% Win	52.16	**54.36**	49.32	N/A
% Loss	47.84	**45.64**	50.68	N/A
Trade Expectancy %	3.78	**9.49**	1.59	N/A
Avg. Bar Held	20.57	61.85	20.44	N/A

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2.3 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบพหุปัจจัยด้วย XGBoost กับพอร์ตโฟลิโอการลงทุนแบบปัจจัยเดียว

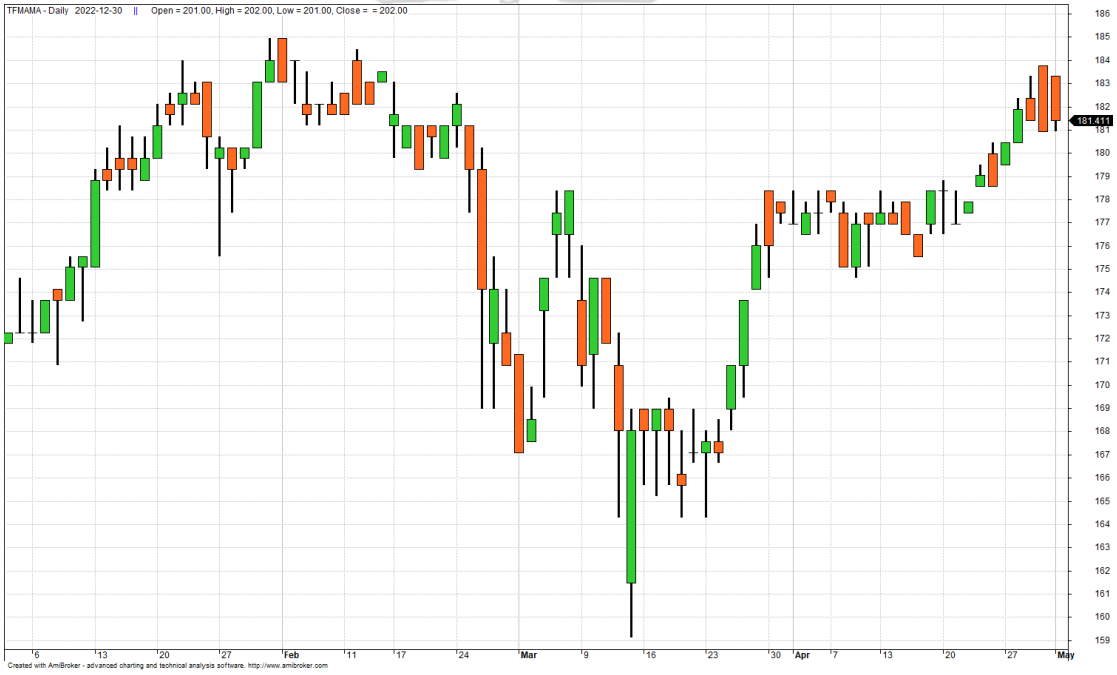
เพื่อเป็นการยืนยันว่าการคัดเลือกหลักทรัพย์แบบพหุปัจจัยนั้นให้ผลลัพธ์ที่ดีกว่าการคัดเลือกหลักทรัพย์แบบปัจจัยเดียว ดังนั้น ในส่วนนี้จึงเป็นการเปรียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบพหุปัจจัยเปรียบเทียบกับปัจจัยเดียว โดยจากภาพที่ 32 และตารางที่ 9 พบว่า ในด้านผลตอบแทนของพอร์ตโฟลิโอการลงทุนด้วยปัจจัยเดียวทั้งหมด 54 ปัจจัยนี้มีค่าเฉลี่ยผลตอบแทนโดยเฉลี่ยทบต้นต่อปีเท่ากับ 9.58% โดยพอร์ตโฟลิโอการลงทุนแบบปัจจัยเดี่ยวที่มีอัตราผลตอบแทนสูงสุดสองอันดับแรก คือ ปัจจัยด้านแนวโน้มของราคาวัดจากกลุ่มของหลักทรัพย์ที่ให้ผลตอบแทนย้อนหลังในอดีต 60 วันต่ำสุด และกลุ่มของหลักทรัพย์ที่ให้ผลตอบแทนย้อนหลังในอดีต 250 วันสูงที่สุด โดยมีผลตอบแทนเฉลี่ยทบต้นต่อปีเท่ากับ 25.78% และ 25.4% ตามลำดับ แสดงให้เห็นว่านอกจากกลุ่มหลักทรัพย์ที่ให้ผลตอบแทนในระดับที่สูงในอดีตจะยังคงให้ผลตอบแทนที่สูงต่อไปในอนาคตแล้ว ยังพบพฤติกรรมของราคาในลักษณะการพลิกกับสู่ค่าเฉลี่ย (Mean Reversion) กล่าวคือ

กลุ่มหลักทรัพย์ขนาดใหญ่และขนาดกลางในประเทศไทยที่ในอดีตถูกแรงเทขายจนทำให้ราคาของหลักทรัพย์มีการปรับตัวลดลงอย่างรุนแรงแล้ว มักจะมีแรงซื้อกลับจากนักลงทุนจนทำให้ราคาของหลักทรัพย์มีการปรับตัวเพิ่มสูงขึ้นในอนาคตด้วย ดังแสดงในภาพที่ 29 ซึ่งเป็นภาพราคาของบริษัท แพลน บี มีเดีย จำกัด (มหาชน) โดยมีสัญลักษณ์ย่อคือ “PLANB” โดยตั้งแต่ช่วงต้นปี ค.ศ. 2020 พบว่า ราคาของหลักทรัพย์มีการปรับตัวลดลงจาก 7.10 บาท ลงมาเหลือเพียง 2.83 บาท ณ สิ้นเดือนมีนาคม ค.ศ. 2020 หรือมีการปรับตัวลดลงเท่ากับ -60.14% ภายในระยะเวลาเพียงหนึ่งไตรมาสเท่านั้น อย่างไรก็ตาม ภายหลังที่ราคาของหลักทรัพย์มีการปรับตัวลดลงอย่างรุนแรงนี้ กลับมาแรงซื้อกลับของนักลงทุนจนทำให้ราคาของหลักทรัพย์มีการปรับตัวเพิ่มสูงขึ้นจนมีราคาเท่ากับ 5.74 บาท ณ สิ้นเดือนมิถุนายน ค.ศ. 2020 หรือมีการปรับตัวเพิ่มขึ้น 102.83% ภายในระยะเวลาหนึ่งไตรมาส

ภาพที่ 29 การเคลื่อนไหวของราคาหลักทรัพย์ PLANB

สำหรับในด้านความเสี่ยงของพอร์ตโฟลิโอวัดโดยอัตราการถดถอยสูงสุดและค่าความผันผวน พบว่า พอร์ตโฟลิโอการลงทุนแบบปัจจัยเดียวที่ให้ค่าความเสี่ยงต่ำที่สุดคือ พอร์ตโฟลิโอที่ลงทุนด้วยปัจจัยสภาพคล่อง โดยคัดเลือกกลุ่มของหลักทรัพย์ที่มีอัตราส่วนหนี้สินต่ำ (Low Debt Ratio) ซึ่งพอร์ตโฟลิโอดังกล่าวมีค่าความเสี่ยงวัดโดยความผันผวนของพอร์ตโฟลิโอและอัตราการถดถอยสูงสุดของพอร์ตโฟลิโอเท่ากับ 11.63% และ -23.62% ทั้งนี้ สำหรับการคำนวณอัตราส่วนหนี้สินนี้คำนวณจากหนี้สินทั้งหมดของบริษัท (Total Liabilities) เปรียบเทียบกับสินทรัพย์ทั้งหมดของบริษัท (Total Assets) โดยบริษัทที่มีอัตราส่วนหนี้สินที่สูงมักจะเป็นบริษัทที่มีความเสี่ยงสูง เนื่องจากมีส่วนของหนี้สินมากกว่าสินทรัพย์ของบริษัท ในทางกลับกัน สำหรับบริษัทที่มีอัตราส่วนหนี้สินต่ำก็หมายความ

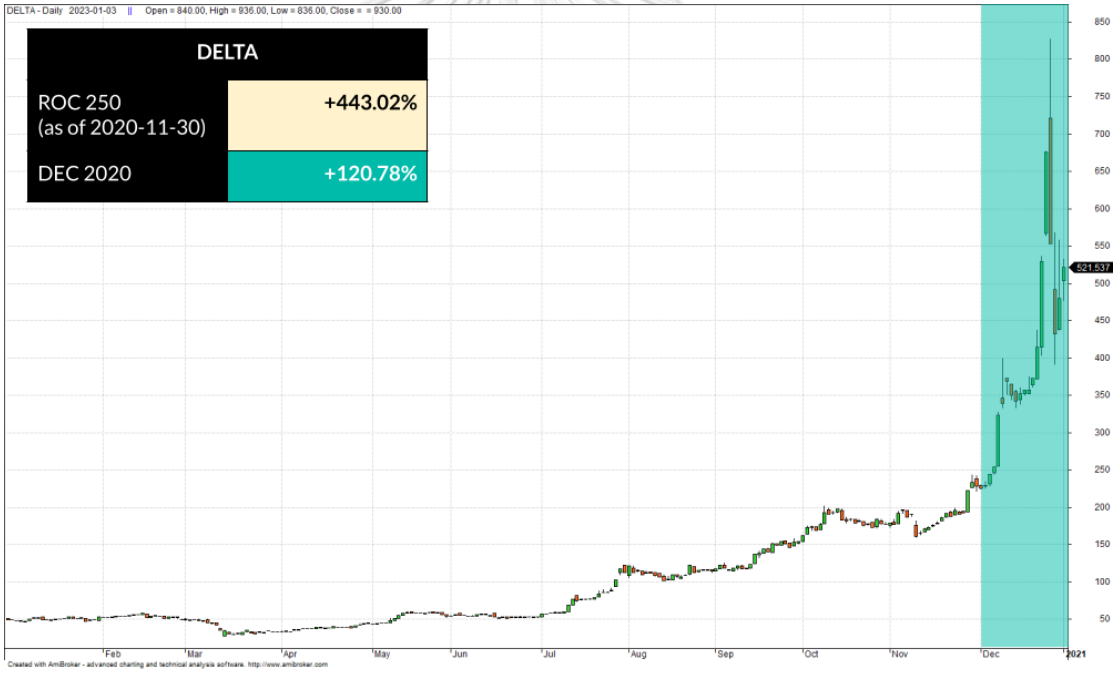
ว่าเป็นบริษัทที่มีความเสี่ยงในเชิงสภาพคล่องที่ต่ำ เช่น บริษัท ไทยเพรซิเดนท์ฟูดส์ จำกัด (มหาชน) โดยมีสัญลักษณ์ย่อคือ “TFMAMA” ซึ่งมีการประกาศงบการเงินประจำไตรมาส 4 ของปี ค.ศ. 2019 พบว่า บริษัทมีหนี้สินทั้งหมด 4,208.69 ล้านบาท ในขณะที่มีสินทรัพย์ทั้งหมดเท่ากับ 34,253.47 ล้านบาท หรือมีอัตราส่วนหนี้สินเท่ากับ 0.12 ซึ่งถือเป็นบริษัทที่มีความเสี่ยงเชิงสภาพคล่องที่ต่ำ โดยถึงแม้ว่าในช่วงการแพร่ระบาดของไวรัส COVID-19 ซึ่งภาพรวมผลตอบแทนของหลักทรัพย์ในตลาดมีการปรับตัวลดลงราวๆ -35% อ้างอิงจากค่า Max. Drawdown ของดัชนีที่ให้น้ำหนักแบบเท่าเทียม อย่างไรก็ตาม ราคาของหลักทรัพย์ TFMAMA กลับมาการปรับตัวลดลงจากจุดสูงสุด ณ วันที่ 4 กุมภาพันธ์ ค.ศ. 2022 ที่ราคา 184 บาท มาทำจุดต่ำสุดที่ราคา 165.69 บาท ณ วันที่ 19 มีนาคม ค.ศ. 2020 หรือมีการปรับตัวลดลงเพียง -9.95% ดังแสดงในภาพที่ 30 ซึ่งถือว่ามีการปรับตัวลดลงจากจุดสูงสุดน้อยกว่าตลาดเป็นอย่างมาก



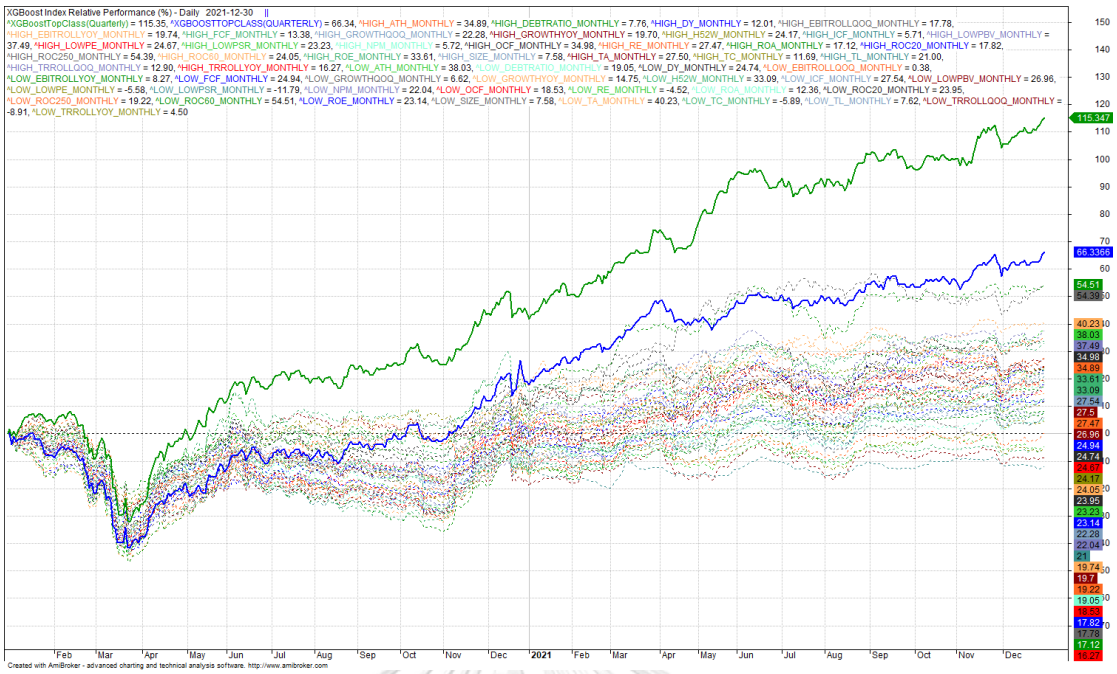
ภาพที่ 30 การเคลื่อนไหวของหลักทรัพย์ TFMAMA

สุดท้าย ในด้านอัตราผลตอบแทนต่อความเสี่ยง พบว่า พอร์ตโฟลิโอการลงทุนจากปัจจัยแนวโน้มของราคานิยามโดยกลุ่มของหลักทรัพย์ที่มีผลตอบแทนในอดีต 250 วันทำการสูงที่สุด (High ROC 250) มีอัตราผลตอบแทนต่อความเสี่ยงวัดโดยอัตราส่วน Sharpe และ MAR ที่ดีที่สุดเมื่อเปรียบเทียบกับพอร์ตโฟลิโอการลงทุนแบบปัจจัยเดียวทั้งหมด โดยมีค่าเท่ากับ 0.85 และ 0.75 ตามลำดับ โดยมีตัวอย่างหลักทรัพย์เช่น บริษัทเดลต้า อีเลคโทรนิคส์ (ประเทศไทย) จำกัด (มหาชน) โดยมีสัญลักษณ์ย่อคือ “DELTA” พบว่า ราคาของหลักทรัพย์ได้มีการปรับตัวเพิ่มขึ้นจาก 42.02 บาท ณ วันที่ 21 พฤศจิกายน ค.ศ. 2019 เป็น 228.17 บาท ณ สิ้นเดือนพฤศจิกายน ค.ศ. 2020 หรือมี

การปรับตัวเพิ่มสูงขึ้น 443.02% ในช่วง 250 วันที่ผ่านมา อย่างไรก็ตาม ภายหลังการปรับตัวเพิ่มสูงขึ้นดังกล่าว พบว่า ในเดือนธันวาคม ค.ศ. 2020 ราคาของหลักทรัพย์ดังกล่าวยังคงปรับตัวเพิ่มขึ้นจาก 228.17 บาท ณ สิ้นเดือนพฤศจิกายน ค.ศ. 2020 เป็น 480.05 บาท ณ สิ้นเดือนธันวาคม ค.ศ. 2020 หรือมีการปรับตัวเพิ่มสูงขึ้นถึง 120.78% ภายในระยะเวลาเพียงหนึ่งเดือน ดังภาพที่ 31 ซึ่งแสดงให้เห็นถึงปัจจัยด้านแนวโน้มของราคาที่หลักทรัพย์ที่มีการปรับตัวในระดับที่สูงกว่าหลักทรัพย์ตัวอื่นๆในอดีต ก็มีแนวโน้มที่จะให้ผลตอบแทนที่สูงในอนาคตด้วยเช่นกัน อย่างไรก็ตาม เมื่อเปรียบเทียบกับพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบพหุปัจจัยด้วย XGBoost ทั้งรายเดือนและรายไตรมาส พบว่า พอร์ตโฟลิโอทั้งสองยังคงมีค่าอัตราผลตอบแทนต่อความเสี่ยงวัดโดยอัตราส่วน Sharpe และ MAR ที่สูงกว่าการลงทุนแบบปัจจัยเดียวทั้งหมด ซึ่งเป็นหลักฐานสนับสนุนสมมติฐานการคัดเลือกหลักทรัพย์แบบพหุปัจจัยนั้นจะให้ผลลัพธ์ในการลงทุนที่ดีกว่าการคัดเลือกหลักทรัพย์จากปัจจัยเพียงปัจจัยเดียว



ภาพที่ 31 การเคลื่อนไหวของหลักทรัพย์ DELTA



ภาพที่ 32 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือน (เส้นเขียว), รายไตรมาส (เส้นน้ำเงิน) และดัชนีผลตอบแทนตามปัจจัยต่างๆ (เส้นประ)

ตารางที่ 9 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบรายเดือนและแบบรายไตรมาสกับค่าสถิติเชิงบรรยายของดัชนีผลตอบแทนตามปัจจัยต่างๆ

Portfolio statistics	CAGR (%)	Max. DD (%)	Ann. STD (%)	Sharpe Ratio	MAR Ratio
Monthly Top-Class	48.19	-37.16	25.16	1.76	1.3
Quarterly Top-Class	30.96	-42.17	25.68	1.05	0.73
Single Factor Portfolio Performance					
#Factors	54	54	54	54	54
Average	9.58	-37.63	23.82	0.22	0.26
Median	10.14	-38.26	24.33	0.25	0.27
Minimum	-6.00	-46.38	11.63	-0.58	-0.18
Maximum	25.78	-23.62	29.12	0.85	0.75
St. Dev.	6.74	4.82	3.54	0.29	0.19

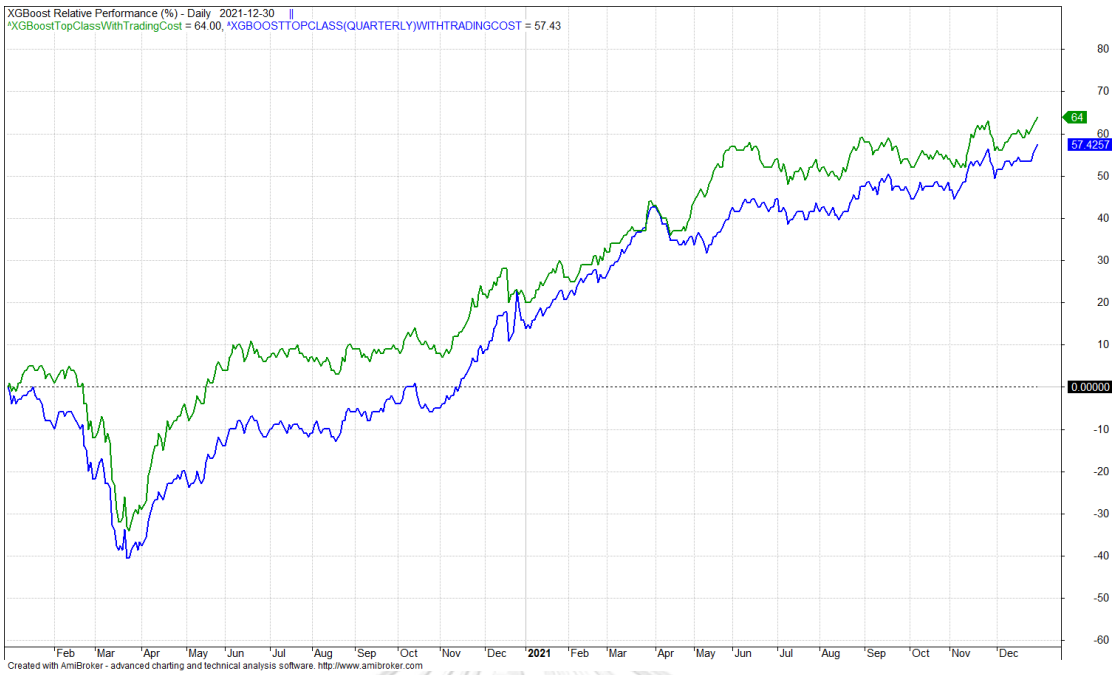
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

4.2.4 ศึกษาผลกระทบต้นทุนการซื้อขายที่มีต่อพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบพหุปัจจัยด้วย XGBoost ในเชิงปฏิบัติ

ในส่วนสุดท้ายนี้เป็นการศึกษาผลกระทบต้นทุนการซื้อขายที่มีต่อพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบพหุปัจจัยด้วย XGBoost โดยเมื่อมีการกำหนดเงื่อนไขการซื้อขายให้สอดคล้องกับการลงทุนจริง คือ เพิ่มค่านายหน้า (Commission) ที่ 0.15% และเพิ่มค่าความคลาดเคลื่อนของการซื้อขายแบบความผันผวน (Volatility Slippage) ที่ 20% ของทุกๆการซื้อและขาย จากภาพที่ 33 และตารางที่ 10 พบว่า ผลตอบแทนโดยเฉลี่ยทบต้นต่อปีของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์แบบรายเดือนมีการปรับตัวลดลงอย่างเห็นได้ชัดจาก 48.19% (ไม่มีต้นทุนการซื้อขาย) เหลือเพียง 29.26% (มีต้นทุนการซื้อขาย) ในขณะที่พอร์ตโฟลิโอที่มีการคัดเลือกหลักทรัพย์แบบรายไตรมาสได้รับผลกระทบเพียงเล็กน้อย คือ มีระดับผลตอบแทนโดยเฉลี่ยทบต้นต่อปีลดลงจาก 30.96% (ไม่มีต้นทุนการซื้อขาย) เหลือ 27.26% (มีต้นทุนการซื้อขาย) เนื่องจากต้นทุนการซื้อขายจะกระทบต่อค่าเฉลี่ยผลตอบแทนคาดหวังต่อการซื้อขายโดยตรง (Trade Expectancy) ดังนั้น พอร์ตโฟลิโอที่มีความถี่ของการซื้อขายที่มากทำให้ต้นทุนการซื้อขายไปลดระดับผลตอบแทนคาดหวังให้น้อยลงจนทำให้ผลตอบแทนของพอร์ตโฟลิโอลดลงตามลำดับ อย่างไรก็ตาม พอร์ตโฟลิโอแบบรายเดือนยังคงให้ผลตอบแทนโดยเฉลี่ยทบต้นที่สูงกว่าพอร์ตโฟลิโอแบบรายไตรมาสเล็กน้อย

ทั้งนี้ สำหรับค่าสถิติของพอร์ตโฟลิโอในด้านความเสี่ยง นิยามโดยความผันผวนของพอร์ตโฟลิโอและอัตราการถดถอยสูงสุดของพอร์ตโฟลิโอ พบว่า ค่าดังกล่าวไม่ได้มีความแตกต่างกับการจำลองพอร์ตโฟลิโอที่ไม่มีต้นทุนการซื้อขายมากนัก เนื่องจากต้นทุนการซื้อขายจะกระทบต่อผลตอบแทนของพอร์ตโฟลิโอเป็นหลัก ซึ่งไม่ได้ทำให้ค่าความเสี่ยงของพอร์ตโฟลิโอเปลี่ยนไปเป็นอย่างมาก อย่างไรก็ตาม ในด้านอัตราผลตอบแทนต่อความเสี่ยงของพอร์ตโฟลิโอทั้งสองกลับมีค่าที่ลดลง เนื่องจากผลตอบแทนของพอร์ตโฟลิโอทั้งสองลดต่ำลง ทำให้มีอัตราส่วน Sharpe ของพอร์ตโฟลิโอแบบรายเดือนและรายไตรมาสเท่ากับ 1.05 และ 0.79 ตามลำดับ และมีอัตราส่วน MAR ของพอร์ตโฟลิโอทั้งสองเท่ากับ 0.94 และ 0.67 ซึ่งแสดงให้เห็นว่า ถึงแม้พอร์ตโฟลิโอแบบรายเดือนจะได้รับผลกระทบจากต้นทุนการซื้อขายที่มากกว่าพอร์ตโฟลิโอแบบรายไตรมาส แต่ภาพรวมค่าสถิติของพอร์ตโฟลิโอแบบรายเดือนนั้นยังให้ค่าที่ดีกว่าพอร์ตโฟลิโอแบบรายไตรมาสในทุกมิติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 33 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost หลังต้นทุนการซื้อขายแบบรายเดือน (เส้นเขียว) และแบบรายไตรมาส (เส้นน้ำเงิน)

ตารางที่ 10 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost หลังต้นทุนการซื้อขายแบบรายเดือน และแบบรายไตรมาส

Portfolio statistics	**Monthly Top-Class**	**Quarterly Top-Class**
CAGR (%)	**29.26**	27.26
Max. Drawdown (%)	**-37.15**	-40.46
Ann. STD (%)	**24.02**	24.78
Sharpe Ratio	**1.05**	0.94
MAR Ratio	**0.79**	0.67
Trade statistics	**Monthly Top-Class**	**Quarterly Top-Class**
#Trades	424	144
% Win	49.53	**54.17**
% Loss	50.27	**45.83**
Trade Expectancy %	2.64	**8.82**
Avg. Bar Held	20.59	61.91

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สุดท้ายนี้ เมื่อทำการเปรียบเทียบพอร์ตโฟลิโอทั้งสองภายหลังต้นทุนการซื้อขายกับดัชนีอ้างอิงประกอบด้วย ดัชนี SET TRI และดัชนีที่ให้น้ำหนักการลงทุนอย่างเท่าเทียมกัน (Equal-Weighted)) ดังแสดงในภาพที่ 34 และตารางที่ 11 พบว่า พอร์ตโฟลิโอทั้งสองก็ยังคงมีระดับผลตอบแทนโดยเฉลี่ยทบต้นต่อปีที่สูงกว่าดัชนีอ้างอิงทั้งสอง ซึ่งเป็นหลักฐานที่สนับสนุนว่าการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบพหุปัจจัยทั้งสองนี้สามารถสร้างผลตอบแทนที่สูงกว่าผลตอบแทนของตลาดในเชิงปฏิบัติ โดยถึงแม้ในเชิงปฏิบัติจะมีค่าความเสี่ยงซึ่งวัดโดยความผันผวนของพอร์ตโฟลิโอและอะตราการถดถอยสูงสุดที่มากกว่าตลาดเล็กน้อย อย่างไรก็ตาม เมื่อพิจารณาในด้านอัตราผลตอบแทนต่อความเสี่ยง ซึ่งวัดจากอัตราส่วน Sharpe และ MAR พบว่า ค่าของพอร์ตโฟลิโอทั้งสองยังคงมีระดับอัตราผลตอบแทนต่อความเสี่ยงที่ดีกว่าดัชนีอ้างอิงทั้งสอง

ภาพที่ 34 ผลตอบแทนสะสมของพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost หลังต้นทุนการซื้อขายแบบรายเดือน (เส้นเขียว), แบบรายไตรมาส (เส้นน้ำเงิน), ดัชนี SET TRI (เส้นประสีดำ) และดัชนี Equal-Weighted (เส้นประสีแดง)

ตารางที่ 11 เปรียบเทียบพอร์ตโฟลิโอการคัดเลือกหลักทรัพย์ด้วย XGBoost หลังต้นทุนการซื้อขาย แบบรายเดือน รายไตรมาส, ดัชนี SET TRI และดัชนี Equal-Weighted

Portfolio statistics	Monthly Top-Class	Quarterly Top-Class	Equal-Weighted	SET TRI
CAGR (%)	**29.26**	27.26	16.82	5.28
Max. Drawdown (%)	**-37.15**	-40.46	-34.9	-35.31
Ann. STD (%)	24.02	24.78	**19.95**	22.81
Sharpe Ratio	**1.05**	0.94	0.64	0.06
MAR Ratio	**0.79**	0.67	0.48	0.15
Trade statistics	**Monthly Top-Class**	**Quarterly Top-Class**	**Equal-Weighted**	**SET TRI**
#Trades	424	144	4,613	N/A
% Win	49.53	**54.17**	49.32	N/A
% Loss	50.47	**45.83**	50.68	N/A
Trade Expectancy %	2.64	**8.82**	1.59	N/A
Avg. Bar Held	20.59	61.91	20.44	N/A

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 5
บทสรุป

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอวิธีการคัดเลือกหลักทรัพย์ขนาดกลางและขนาดใหญ่แบบพหุปัจจัยโดยการใช้แบบจำลอง XGBoost ที่มีประสิทธิภาพในการทำนายสูง ทั้งนี้ การศึกษามุ่งเน้นไปที่การสร้างพอร์ตโฟลิโอหมุนเวียนหลักทรัพย์ (Rotational Trading Portfolio) แบบรายเดือนและรายไตรมาส เพื่อประเมินสมรรถนะการทำนายในการใช้งานจริง (Usable) ซึ่งแบบจำลอง XGBoost ทั้งสองได้ถูกพัฒนาเพื่อให้เรียนรู้ความสัมพันธ์ที่ซับซ้อนของปัจจัยทั้ง 27 ปัจจัยบนฐานข้อมูลหลักทรัพย์ขนาดใหญ่และขนาดกลางในประเทศไทย และทำการทำนายผลตอบแทนในอนาคตของหลักทรัพย์ โดยแบ่งออกเป็นสามกลุ่ม ประกอบด้วย 1) กลุ่มที่ให้ผลตอบแทนสูงสุด 30% (Top30%) 2) กลุ่มที่ให้ผลตอบแทนตรงกลาง 40% (Middle 40%) และ 3) กลุ่มที่ให้ผลตอบแทนต่ำสุด 30% (Bottom 30%) การประเมินสมรรถนะแบบจำลองบนชุดข้อมูลทดสอบ พบว่า สามารถทำนายผลตอบแทนในอนาคตหนึ่งเดือนและหนึ่งไตรมาส โดยให้ค่า AUC-ROC ที่สูงถึง 0.59% และ 0.58% ตามลำดับ นอกจากนี้ ยังพบว่า ปัจจัยเชิงเทคนิคเป็นปัจจัยที่มีความสำคัญสำหรับการทำนายการเคลื่อนไหวของราคาสินทรัพย์แบบรายเดือน ซึ่งจะแตกต่างกับการเคลื่อนไหวของราคาสินทรัพย์แบบรายไตรมาสที่ปัจจัยพื้นฐานจะมีความสำคัญมากกว่าปัจจัยเชิงเทคนิค

นอกจากนี้ เมื่อจำลองการสร้างพอร์ตโฟลิโอแบบการหมุนเวียนหลักทรัพย์แบบรายเดือนและรายไตรมาสด้วยกระบวนการ Backtesting ตั้งแต่ปี ค.ศ. 2020 ถึง ปี ค.ศ. 2021 พบว่า พอร์ตโฟลิโอแบบรายเดือนให้ค่าสถิติในเชิงพอร์ตโฟลิโอที่ดีกว่า ซึ่งเป็นเพราะพอร์ตโฟลิโอดังกล่าวมีโอกาสในการคัดเลือกหลักทรัพย์ใหม่ๆเข้ามาในพอร์ตโฟลิโอได้มากกว่า อย่างไรก็ตาม ในด้านค่าสถิติของการซื้อขายกลับกลับพบว่าพอร์ตโฟลิโอแบบรายไตรมาสกลับให้ค่าที่ดีกว่า เนื่องจากระยะเวลาการถือครองหลักทรัพย์ที่นานขึ้นทำให้มีค่าผลตอบแทนคาดหวังต่อการซื้อขายที่สูงขึ้น ทั้งนี้ พอร์ตโฟลิโอทั้งสองยังคงให้ระดับผลตอบแทนต่อความเสี่ยงที่สูงกว่าดัชนีอ้างอิงทั้งหมดประกอบด้วย 1) ดัชนี SET TRI 2) ดัชนีที่ให้น้ำหนักการลงทุนแบบเท่าเทียม (Equal-Weighted Index) และ 3) ดัชนีผลตอบแทนแบบปัจจัยเดียว (Factor Index หรือ Style Index) นอกจากนี้ ผู้วิจัยได้ทำการศึกษาผลลัพธ์ในเชิงการปฏิบัติใช้งานสำหรับการลงทุนจริงที่ต้องพิจารณาตัวแปรด้านต้นทุนการซื้อขายเพิ่มด้วย พบว่าพอร์ตโฟลิโอการคัดเลือกหุ้นแบบรายไตรมาสได้รับผลกระทบจากต้นทุนการซื้อขายที่น้อยกว่าพอร์ตโฟลิโอแบบรายเดือน อันเนื่องมาจากความถี่ของการซื้อขายน้อยกว่าพอร์ตโฟลิโอแบบรายเดือน

5.2 ข้อจำกัดในงานวิจัย

งานวิจัยนี้ได้จัดทำขึ้นเพื่อศึกษาการคัดเลือกหลักทรัพย์ด้วย XGBoost แบบพหุปัจจัย โดยมีขอบเขตการศึกษาคือ กลุ่มหลักทรัพย์ขนาดใหญ่และขนาดกลางในประเทศไทย ซึ่งไม่ได้ครอบคลุมถึงหลักทรัพย์ขนาดเล็ก ดังนั้น การประยุกต์ใช้สำหรับการคัดเลือกหลักทรัพย์กลุ่มดังกล่าวอาจจะมีรายละเอียดที่แตกต่างไปจากงานวิจัยนี้ นอกจากนี้ แบบจำลองทั้งสองถูกพัฒนาบนชุดข้อมูลสอน ช่วงระยะเวลา ค.ศ. 2012-2019 ความเป็นทั่วไป (Generalization) ของแบบจำลองจึงอาจมีสมรรถนะที่จำกัด รวมทั้งเป็นข้อจำกัดของการสร้างพอร์ตโฟลิโอหมุนเวียนหลักทรัพย์ที่เหมาะสมสำหรับรายเดือนและรายไตรมาสเท่านั้น

5.3 งานวิจัยในอนาคต

งานวิจัยในอนาคตที่ควรนำมาต่อยอด ได้แก่ การขยอยขอบเขตกลุ่มตัวอย่าง (Universe) รวมถึงการเพิ่มปัจจัยที่เป็นตัวแปรต้น เช่น ปัจจัยเศรษฐกิจมหภาค (Macroeconomics), ปัจจัยด้านภาวะของตลาด (Sentiment factor), และอื่นๆ เพื่อให้การเรียนรู้ของเครื่อง (Machine learning) สามารถค้นหาแบบหรือแพตเทิร์นเพื่อช่วยยกระดับสมรรถนะของแบบจำลองให้ดียิ่งขึ้น

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

สูตรคำนวณค่าสถิติและดัชนีชี้วัดพอร์ตโฟลิโอ

ค่าสถิติดัชนีชี้วัดพอร์ตโฟลิโอ (Portfolio Metrics) และดัชนีชี้วัดการซื้อขาย (Trade Metrics) มีรายละเอียดการคำนวณดังนี้

1. สูตรผลตอบแทนโดยเฉลี่ยทบต้นต่อปี (Compounding Annual Growth Rate หรือ CAGR)

$$CAGR\ (\%) = \left(\left(\left(\frac{Ending\ Value}{Begining\ Value} \right)^{\frac{1}{nYears}} \right) - 1 \right) * 100$$

โดย

- $Ending\ Value$ คือ มูลค่าพอร์ตโฟลิโอ ณ วันสิ้นงวดหรือวันสุดท้าย
- $Begining\ Value$ คือ มูลค่าพอร์ตโฟลิโอ ณ วันต้นงวดหรือวันแรก
- $nYears$ คือ จำนวนปีในการลงทุน

2. สูตรค่าความผันผวนของพอร์ตโฟลิโอ (Annualized Standard Deviation หรือ Ann. STD)

$$Ann.STD = \sqrt{\frac{\sum_{i=1}^{n}(r_i - \mu)}{n-1}} \times \sqrt{nDIY}$$

โดย

- r_i คือ ผลตอบแทนรายวัน (Daily Return) ณ วันที่ i
- μ คือ ค่าเฉลี่ยผลตอบแทนรายวัน
- n คือ จำนวนวันทั้งหมด
- $nDIY$ คือ จำนวนวันทั้งหมดที่มีในหนึ่งปี โดยในที่นี้กำหนดไว้ที่ 252 วันซึ่งเป็นค่ามาตรฐานจำนวนวันทำการในแต่ละปี

3. สูตรอัตราการถดถอยสูงสุดของพอร์ตโฟลิโอ (Maximum Drawdown หรือ Max. DD)

$$Drawdown\ (\%) = \left(\left(\frac{Current\ Value}{Highest\ Value} \right) - 1 \right) * 100$$

$$Maximum\ Drawdown = Lowest(Drawdown)$$

โดย

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

- Current Value คือ มูลค่าพอร์ตโฟลิโอ ณ ปัจจุบัน
- Highest Value คือ มูลค่าที่สูงที่สุดของพอร์ตโฟลิโอ

4. สูตรอัตราส่วน Sharpe Ratio

$$Sharpe\ Ratio = \frac{r_p - r_f}{\sigma_p}$$

โดย

- r_p คือ ผลตอบแทนของพอร์ตโฟลิโอการลงทุน โดยในที่นี้ใช้ผลตอบแทนโดยเฉลี่ยทบต้นต่อปี
- r_f คือ ผลตอบแทนของการลงทุนที่ปราศจากความเสี่ยง โดยในที่นี้กำหนดไว้ที่ *4%* ต่อปี
- σ_p คือ ค่าความผันผวนของพอร์ตโฟลิโอ

5. สูตรอัตราส่วน MAR Ratio

$$MAR\ Ratio = \frac{CAGR}{Max.Drawdown}$$

โดย

- $CAGR$ คือ ผลตอบแทนโดยเฉลี่ยทบต้นต่อปี
- $Max.Drawdown$ คือ อัตราการถดถอยสูงสุดของพอร์ตโฟลิโอ

6. จำนวนครั้งการซื้อขายทั้งหมด (*#Trades*) คำนวณจากผลรวมจำนวนหลักทรัพย์ที่มีการซื้อขายเข้ามาในพอร์ตโฟลิโอทั้งหมดตลอดช่วงระยะเวลาการจำลองการสร้างพอร์ตโฟลิโอ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$\#Trades = Total\ Number\ of\ Trades$$

7. เปอร์เซ็นต์ของเทรดที่ได้กำไร (% Win) คำนวณจาก

$$\%\ Win = \left(\frac{Number\ of\ Profitable\ Trades}{Total\ Number\ of\ Trades}\right) * 100$$

โดย

- $Number\ of\ Profitable\ Trades$ คือ จำนวนหลักทรัพย์ที่ซื้อแล้วได้กำไร
- $Total\ Number\ of\ Trades$ คือ จำนวนหลักทรัพย์ที่มีการซื้อเข้ามาในพอร์ตโฟลิโอ

8. เปอร์เซ็นต์ของเทรดที่ขาดทุน (% Loss) คำนวณจาก

$$\%\ Loss = \left(\frac{Number\ of\ Loss\ Trades}{Total\ Number\ of\ Trades}\right) * 100$$

โดย

- $Number\ of\ Loss\ Trades$ คือ จำนวนหลักทรัพย์ที่ซื้อแล้วขาดทุน
- $Total\ Number\ of\ Trades$ คือ จำนวนหลักทรัพย์ที่มีการซื้อเข้ามาในพอร์ตโฟลิโอ

9. ค่าเฉลี่ยผลตอบแทนคาดหวัง (Trade Expectancy) คำนวณจากค่าเฉลี่ยผลตอบแทนจากทุกการซื้อขาย

$$Trade\ Expectancy = (\%\ Win * Average\ Win) - (\%\ Loss * Average\ Loss)$$

โดย

- $\%\ Win$ คือ เปอร์เซ็นต์ของเทรดที่ได้กำไร
- $Average\ Win$ คือ ค่าเฉลี่ยผลตอบแทน (%) ของทุกเทรดที่ได้กำไร
- $\%\ Loss$ คือ เปอร์เซ็นต์ของเทรดที่ขาดทุน
- $Average\ Loss$ คือ ค่าเฉลี่ยผลตอบแทน (%) ของทุกเทรดที่ขาดทุน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

10. สูตรค่าเฉลี่ยช่วงระยะเวลาการถือครองหลักทรัพย์ (Avg. Bar Held)

$$Avg.Bar\ Held = \left(\frac{Sum\ Bar\ Held\ of\ All\ Trades}{Total\ Number\ of\ Trades}\right)$$

โดย

- $Sum\ Bar\ Held\ of\ All\ Trades$ คือ ผลรวมช่วงระยะเวลาการถือครองหลักทรัพย์ของทุกการซื้อขายในพอร์ตโฟลิโอ
- $Total\ Number\ of\ Trade$ คือ จำนวนหลักทรัพย์ที่มีการซื้อเข้ามาในพอร์ตโฟลิโอ

บรรณานุกรม

1. Bernard, V.L. and J.K. Thomas, *Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?* Journal of Accounting Research, 1989. **27**: p. 1-36.
2. Jegadeesh, N. and S. Titman, *Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency.* The Journal of Finance, 1993. **48**(1 (Mar., 1993)): p. 65-91.
3. Fama, E.F., *Efficient Capital Markets: A Review of Theory and Empirical Work.* The Journal of Finance, 1970. **25**(2): p. 383-417.
4. Investopedia. *Multi-Factor Model: Definition and Formula for Comparing Factors*. [cited 2023 11 February]; Available from: https://www.investopedia.com/terms/m/multifactor-model.asp.
5. Fama, E.F. and K.R. French, *The cross-section of expected stock returns.* The Journal of Finance, 1992. **47**(2): p. 427-465.
6. Carhart, M.M., *On Persistence in Mutual Fund Performance.* The Journal of Finance, 1997. **52**(1 (Mar., 1997)): p. 57-82.
7. Fama, E.F. and K.R. French, *A five-factor asset pricing model.* Journal of Financial Economics, 2015. **116**: p. 1-22.
8. Zhang, R., et al., *Multi Factor Stock Selection Model Based on LSTM.* International Journal of Economics and Finance, 2018. **10**(8).
9. Zhong, Y., et al., *Multi-factor Stock Selection Model Based on Machine Learning.* Engineering Letters, 2021. **29**(1).
10. Chen, T. and C. Guestrin, *XGBoost: A scalable tree boosting system.* In Proceedings of the 22nd acm sigkkd international conference on knowledge discovery and data mining, 2016: p. 785-794.
11. Hsieh, C., et al., *Feature Selection Framework for XGBoost Based on Electrodermal Activity in Stress Detection.* IEEE International Workshop on Signal Processing Systems (SiPS), 2019.
12. Li, H., et al., *XGBoost Model and Its Application to Personal Credit Evaluation.* IEEE Intelligent Systems, 2020. **35**(3): p. 52-61.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

13. Zhange, X. and W. Chen, *Stock Selection based on Extreme Gradient Boosting.* 2019 the 38th Chinese Control Conference (CCC2019), 2019.
14. Li, J. and R. Zhang, *Dynamic Weighting Multi Factor Stock Selection Strategy Based on XGboost Machine Learning.* 2018 IEEE International Conference of Safety Produce Informatization (IICSPI), 2018.
15. Dai, J. and J. Zhou, *A Novel Quantitative Stock Selection Model Based on Support Vector Regression.* 2019 International Conference on Economics Management and Model Engineering (ICEMME), 2019.
16. Yang, M., *Research on quantitative multi-factor stock selection model based on machine learning.* 2021 the 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST), 2021.
17. Baquero, G., J.T. Horst, and M. Verbeek, *Survival, Look-Ahead Bias, and Persistence in Hedge Fund Performance.* Journal of Financial and Quantitative Analysis, 2005. **40**(3).
18. Mandelbrot, B.B., *The variation of certain speculative prices.* Journal of Business, 1963. **36**(4): p. 392-417.
19. Chong, T.T.L., Y. Wu, and J. Su, *The Unusual Trading Volume and Earnings Surprises in China's Market.* Journal of Risk and Financial Management, 2020.
20. Investopedia. *How Do I Keep Commissions and Fess Eating Trading Profits?* [cited 2023 09 April]; Available from: https://www.investopedia.com/ask/answers/08/trading-frequency-commissions.asp.
21. Investopedia. *Slippage: What It Means in Finance*. [cited 2023 09 April]; Available from: https://www.investopedia.com/terms/s/slippage.asp.
22. Pardo, R., *The Evaluation and Optimization Trading Strategies*. 2008: Wiley; 2nd edition.
23. Sharpe, W.F., *The Sharpe Ratio.* The Journal of Portfolio Management Fall 1994, 1994. **21**(1): p. 49-58.
24. Investory. *SET TRI Index*. [cited 2022 31 October]; Available from: https://investory.set.or.th/news/details/020cc009-305d-466a-b836-55206bf4a5f5.
25. SET. *SET Index Series*. [cited 2022 31 October]; Available from:

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

https://classic.set.or.th/en/products/index/setindex_p1.html.

26. Funds, A. *Understanding Style Investing*. [cited 2022 31 October]; Available from: https://funds.aqr.com/Insights/Fund-Promo/Style-Investing.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล	ธนดล ประพฤทธิกุล
วัน เดือน ปี เกิด	5 ธันวาคม 2537
สถานที่เกิด	กรุงเทพมหานคร
ที่อยู่ปัจจุบัน	234/185 หมู่บ้านมัณฑนา ถนนวงแหวนเทพารักษ์ ตำบลบางพลีใหญ่ อำเภอบางพลี จังหวัดสมุทรปราการ 10540
ผลงานตีพิมพ์	T. Praphutikul, and Y. Limpiyakorn, “XGBoost for Smart Portfolio Management Based on Multi Factor Stock Selection”, 9th International Conference on Computer Technology Applications (ICCTA 2023), May 10-12, 2023, Vienna, Austria. T. Praphutikul, and Y. Limpiyakorn, “XGBoost-Based Multi-Factor Stock Selection Model for Rotational Trading”, 5th International Conference on Information Technology and Computer Communications (ITCC 2023), June 15-17, 2023, Tianjin, China.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University