

การทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและ
การวิเคราะห์เชิงเทคนิค

STOCK PRICE DIRECTION PREDICTION OF SET50 USING NEWS AND
TECHNICAL ANALYSIS



การค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูลและการวิเคราะห์
ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2566

KMITL-2023-SC-M-017-038

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

STOCK PRICE DIRECTION PREDICTION OF SET50 USING NEWS AND
TECHNICAL ANALYSIS



AN INDEPENDENT STUDY SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE
IN DATA SCIENCE AND ANALYTICS

KMITL DIGITAL ANALYTICS AND INTELLIGENCE CENTER SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2023

KMITL-2023-SC-M-017-038

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2023

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อการค้นคว้าอิสระ	การทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและการวิเคราะห์เชิงเทคนิค
ชื่อนักศึกษา	ศุภณัฐ ก้อนศิลา
รหัสประจำตัว	64605110
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการข้อมูลและการวิเคราะห์) ศูนย์วิเคราะห์ข้อมูลดิจิทัลอัจฉริยะพระจอมเกล้าลาดกระบัง
พ.ศ.	2566
อาจารย์ที่ปรึกษาการค้นคว้าอิสระ	ผู้ช่วยศาสตราจารย์ ดร.บุษยมาส พิมพ์พรรณชาติ

บทคัดย่อ

เนื่องจากปัจจุบันนี้ตลาดหุ้นมีการเปลี่ยนแปลงตลอดเวลาตามข่าวสารที่เกิดขึ้นทั่วโลก จึงเป็นเรื่องยากที่เราจะพยากรณ์ทิศทางราคาหุ้นได้อย่างแม่นยำ ผู้วิจัยจึงมีความสนใจในการพยากรณ์ทิศทางของราคาหุ้นในกลุ่ม SET50 โดยใช้ข้อมูลข่าวสารและการวิเคราะห์เชิงเทคนิคเป็นตัวแปรอิสระตั้งแต่เดือนพฤษภาคม พ.ศ.2561 ถึง มกราคม พ.ศ.2566 รวมทั้งสิ้น 631 วัน ในการสร้างแบบจำลองเพื่อมาช่วยประกอบการตัดสินใจหรือการวางแผนกลยุทธ์การซื้อขาย ในงานวิจัยมีวัตถุประสงค์เพื่อวิเคราะห์และพิจารณาความสามารถของแบบจำลองร่วมกับปัจจัยภายนอกต่างๆ ในการพยากรณ์ทิศทางของราคาหุ้นในกลุ่ม SET50 ผ่านการนำเข้าข้อมูลข่าวสารที่เกี่ยวข้องกับตลาดหุ้น รวมถึงข้อมูลทางเทคนิคที่ได้รับค่านิยมในการวิเคราะห์หุ้น ตัวอย่างเช่น เทรนด์ของราคาหุ้น ปริมาณการซื้อขาย และค่าสถิติทางเทคนิคอื่นๆ โดยผู้วิจัยได้นำการประมวลผลภาษาธรรมชาติเข้ามาวิเคราะห์บทความข่าวสารรายวัน เพื่อนำตัวแปรเหล่านี้มาสร้างแบบจำลองการเรียนรู้ของเครื่องจักร อาทิ ซัพพอร์ตตเวกเตอร์แมชชีน ป่าไม้สุ่ม อดาบูสต์ เกรเดียนท์บูสต์ดิง และแคสบูสต์ เพื่อเปรียบเทียบประสิทธิภาพความแม่นยำของแบบจำลองและเปรียบเทียบประสิทธิภาพการใช้ปัจจัยต่างๆ ในการพยากรณ์ทิศทางราคาหุ้น จากการทดลองพบว่าแบบจำลองอดาบูสต์ตอบสนองกับหุ้นในกลุ่ม SET50 มากที่สุด รองลงมาคือแบบจำลองซัพพอร์ตตเวกเตอร์แมชชีน ในด้านปัจจัยที่มีต่อการพยากรณ์มากที่สุดคือ ปัจจัยเชิงเทคนิค เช่น ราคาดัชนีอุตสาหกรรม เป็นต้น ดังนั้นผลการวิจัยนี้จะช่วยให้นักลงทุนและผู้สนใจตลาดหุ้นในกลุ่ม SET50 ได้มีข้อมูลและความรู้เพิ่มเติมเกี่ยวกับการพยากรณ์ทิศทางราคาหุ้นและสามารถใช้ข้อมูลเหล่านี้ในการประกอบการตัดสินใจการลงทุนให้เหมาะสมกับตัวเองมากขึ้น

คำสำคัญ : หุ้น SET50 ซัพพอร์ตตเวกเตอร์แมชชีน ป่าไม้สุ่ม อดาบูสต์ เกรเดียนท์บูสต์ดิง แคสบูสต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Independent Study Title	Stock Price Direction Prediction of SET50 Using News and Technical Analysis
Student Name	Supanut Konsila
Student ID	64605110
Degree	Master of Science (Data Science and Analytics) KMITL Digital Analytics and Intelligence Center
Year	2023
Independent Study Advisor	Asst.Prof.Dr.Busayamas Pimpunchat

Abstract

The stock market is constantly changing based on news events occurring worldwide, making it difficult to accurately predict stock price directions. This research focuses on predicting the direction of SET50 stock prices using news and technical analysis data as independent variables from May 2018 to January 2023, totaling 631 days. The study aims to analyze and evaluate the performance of a predictive model in conjunction with various external factors for SET50 stock price prediction. The research utilizes relevant news data and popular technical indicators for stock analysis, such as price trends, trading volumes, and other technical statistics. Natural language processing techniques are applied to analyze daily news articles and extract variables for machine learning models, including Support Vector Machine, Random Forest, AdaBoost, Gradient Boosting, and CatBoost. The models' accuracy and the effectiveness of different factors in stock price prediction are compared. The results indicate that the AdaBoost model performs the best in predicting SET50 stock prices, followed by the SVM model. Among the factors, technical indicators demonstrate the highest predictive power, such as industry index prices. Therefore, this research provides valuable insights and knowledge for investors and individuals interested in the SET50 stock market, enabling them to make more informed investment decisions.

Keywords: SET50, Support Vector Machine, Random Forest, AdaBoost, Gradient Boosting, CatBoost



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

การศึกษาค้นคว้าอิสระหัวข้อ การทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและการวิเคราะห์เชิงเทคนิค ผู้วิจัยขอขอบพระคุณ ผศ.ดร.บุษยามาส พิมพ์พรรณชาติ อาจารย์ที่ปรึกษาการค้นคว้าอิสระ ที่คอยเสียสละเวลาในการให้คำปรึกษา แนะนำแนวทาง และการแก้ปัญหาต่างๆ ให้การค้นคว้าอิสระผ่านสำเร็จลุล่วงไปด้วยดี

ขอขอบพระคุณผู้ทรงคุณวุฒิ รศ.ดร.กนกณัฐภูษ วัฒนแจ่มศรี ที่กรุณาเป็นประธานในการสอบ และ ผศ.ดร.กนกกรรณ ลีโรจนาที่กรุณาเป็นกรรมการในการสอบในการค้นคว้าอิสระนี้ และช่วยตรวจสอบแก้ไขให้การค้นคว้าอิสระเล่มนี้สมบูรณ์

ขอขอบพระคุณเพื่อนๆ และเจ้าหน้าที่ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่าน ที่คอยช่วยอำนวยความสะดวกตลอดในหลายๆ เรื่อง

ขอขอบพระคุณพี่ๆ นักลงทุนทุกท่านที่ถ่ายทอดประสบการณ์ความรู้ในเรื่องการลงทุนต่างๆ ให้ผู้วิจัยสามารถนำความรู้เหล่านั้นมาประยุกต์ใช้ในงานการศึกษาค้นคว้าอิสระหัวข้อนี้

และขอขอบพระคุณบิดา มารดา และครอบครัวของข้าพเจ้า ที่ให้การสนับสนุน และเป็นกำลังใจที่สำคัญในการทำการค้นคว้าอิสระเล่มนี้ให้ผ่านไปได้โดยดี

นายศุภณัฐ ก้อนศิลา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ข
กิตติกรรมประกาศ	ง
สารบัญ	จ
สารบัญตาราง	ช
สารบัญรูป	ญ
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 ข้อมูล SET50	4
2.1.1 เกณฑ์การคำนวณดัชนี SET50	4
2.1.2 เกณฑ์ในการคัดเลือกของหลักทรัพย์ใน SET50	5
2.2 ปัจจัยส่งผลกระทบต่อตลาดหุ้น (Factors affecting the stock market)	7
2.2.1 เศรษฐกิจ (Economic)	7
2.2.2 กิจกรรมทางการเมือง (Political Events)	8
2.2.3 กิจกรรมทางบริษัท (Company Events)	8
2.2.4 อารมณ์ของนักลงทุน (Investor Sentiment)	9
2.2.5 แนวโน้มของตลาด (Market Moment)	9
2.2.6 ภัยธรรมชาติ (Natural Disasters)	10
2.2.7 กิจกรรมในภาคอุตสาหกรรม (Industry Events)	10
2.3 ปัจจัยเชิงเทคนิค (Technical Analysis)	10
2.3.1 แท่งเทียน (Candle Stick)	10
2.3.2 ปริมาณการซื้อขาย (Volume)	11

สารบัญ (ต่อ)

	หน้า
2.3.3 เส้นค่าเฉลี่ยเคลื่อนที่แบบเอ็กซ์โพเนนเชียล (Exponential Moving Average: EMA)	12
2.3.4 เอ็มเอซีดี (Moving Average Convergence Divergence: MACD)	14
2.3.5 อาร์เอสไอ (Relative Strength Index: RSI)	15
2.4 เทคนิคการเลือกตัวแปร (Feature Selection)	16
2.5 การเรียนรู้ของเครื่องจักร (Machine Learning)	18
2.6 การประมวลภาษาธรรมชาติ (Natural Language Processing: NLP)	22
2.7 การประเมินประสิทธิภาพ (Model Evaluation)	24
บทที่ 3 วิธีการดำเนินงานวิจัย	27
3.1 ขั้นตอนการดำเนินงาน	27
3.2 การรวบรวมข้อมูล	28
3.2.1 ปัจจัยข่าว	28
3.2.2 ปัจจัยเชิงเทคนิค	29
3.3 การจัดเตรียมข้อมูล	30
3.3.1 ปัจจัยข่าว	30
3.3.2 ปัจจัยเชิงเทคนิค	33
3.4 ตัวแปรที่ใช้ในการวิเคราะห์	43
3.4.1 ตัวแปรตาม (Dependent Variable)	43
3.4.2 ตัวแปรอิสระ (Independent Variable)	43
3.5 การสร้างแบบจำลอง	58
3.6 เครื่องมือที่ใช้ในการวิเคราะห์	59
บทที่ 4 ผลการวิจัยและอภิปรายผล	61
4.1 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น DELTA	62
4.2 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น AOT	70
4.3 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น PTT	76
4.4 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น ADVANC	83

สารบัญ (ต่อ)

	หน้า
4.5 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น GULF	89
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	97
5.1 สรุปผลการดำเนินงาน	97
5.2 ข้อเสนอแนะ	99
เอกสารอ้างอิง	100
ภาคผนวก	104
ภาคผนวก ก	105
ประวัติผู้เขียน	135



สารบัญตาราง

ตารางที่	หน้า
ตารางที่ 2.1 รายชื่อหลักทรัพย์ดัชนี SET50 ช่วงเดือน พฤศจิกายน พ.ศ.2565	5
ตารางที่ 2.2 เมทริกซ์ความสับสน (Confusion Matrix) ขนาด 3x3	25
ตารางที่ 3.1 ตัวแปรตาม	43
ตารางที่ 3.2 ตัวแปรอิสระปัจจัยข่าว	43
ตารางที่ 3.3 ตัวแปรอิสระปัจจัยเชิงเทคนิครายวัน	49
ตารางที่ 3.4 ตัวแปรอิสระดัชนีราคาต่างประเทศและดัชนีราคาในประเทศไทยรายวัน	56
ตารางที่ 3.5 โลจิสติกส์สำหรับการวิเคราะห์ข้อมูล	59
ตารางที่ 4.1 รายชื่อหุ้นใน SET50 ทั้งหมด 5 ตัว ตามมูลค่าหลักทรัพย์ตามราคาตลาด	61
ตารางที่ 4.2 การตั้งค่าพารามิเตอร์สำหรับแบบจำลอง	62
ตารางที่ 4.3 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น DELTA	62
ตารางที่ 4.4 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น DELTA	63
ตารางที่ 4.5 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ DELTA	64
ตารางที่ 4.6 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันแรกในการพยากรณ์ทิศทางของราคาของ DELTA	64
ตารางที่ 4.7 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น AOT	70
ตารางที่ 4.8 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น AOT	71
ตารางที่ 4.9 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ AOT	71
ตารางที่ 4.10 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันแรกในการพยากรณ์ทิศทางของราคาของ AOT	72
ตารางที่ 4.11 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น PTT	76
ตารางที่ 4.12 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น PTT	78
ตารางที่ 4.13 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ PTT	78
ตารางที่ 4.14 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันแรกในการพยากรณ์ทิศทางของราคาของ PTT	79
ตารางที่ 4.15 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น ADVANC	83

สารบัญตาราง (ต่อ)

ตารางที่	หน้า
ตารางที่ 4.16 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางของการเปลี่ยนแปลงราคาหุ้น ADVANC	84
ตารางที่ 4.17 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ ADVANC	85
ตารางที่ 4.18 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันแรกในการพยากรณ์ทิศทางของราคาของ ADVANC	85
ตารางที่ 4.19 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น GULF	90
ตารางที่ 4.20 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางของการเปลี่ยนแปลงราคาหุ้น GULF	91
ตารางที่ 4.21 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ GULF	91
ตารางที่ 4.22 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันแรกในการพยากรณ์ทิศทางของราคาของ GULF	92
ตารางที่ 5.1 ประสิทธิภาพแบบจำลองกับหุ้นในกลุ่ม SET50	97
ตารางที่ 5.2 ตัวแปรอิสระที่สำคัญอันดับที่ 1 ของการพยากรณ์ทิศทางของการเปลี่ยนแปลงราคาหุ้นในกลุ่ม SET50	99

สารบัญรูป

รูปที่	หน้า
รูปที่ 2.1 ตัวอย่างการคำนวณ SET50	5
รูปที่ 2.2 ส่วนประกอบของแท่งเทียน	11
รูปที่ 2.3 ดัชนี SET50 และปริมาณการซื้อขาย	12
รูปที่ 2.4 กราฟดัชนี SET กับเส้นค่าเฉลี่ยเคลื่อนที่	14
รูปที่ 2.5 กราฟหุ้นและการตีความค่า MACD	15
รูปที่ 2.6 กราฟหุ้นและการตีความค่า RSI	16
รูปที่ 2.7 แบบจำลองซัพพอร์ตเวกเตอร์แมทซิ่ง	18
รูปที่ 2.8 แบบจำลองป่าไม้สุ่ม	19
รูปที่ 2.9 แบบจำลองอตาบัสต์	20
รูปที่ 2.10 แบบจำลองเกรเดียนท์บูสต์ดีง	21
รูปที่ 2.11 แบบจำลองแคสบูสต์	21
รูปที่ 2.12 คำที่คล้ายกันจะอยู่ในเวกเตอร์ที่ใกล้เคียงกัน	22
รูปที่ 3.1 ขั้นตอนการดำเนินงาน	27
รูปที่ 3.2 ตัวอย่างข่าวหุ้นรายวัน	28
รูปที่ 3.3 ตัวอย่างข่าวเศรษฐกิจ	29
รูปที่ 3.4 ตัวอย่างราคาดัชนี (รายวัน)	29
รูปที่ 3.5 ตัวอย่างราคาหุ้น SET50 (รายวัน)	30
รูปที่ 3.6 ตัวอย่างราคาดัชนีอุตสาหกรรมในประเทศ (รายวัน)	30
รูปที่ 3.7 ตัวอย่างข่าวเศรษฐกิจ	30
รูปที่ 3.8 ตัวอย่างข่าวหุ้น	30
รูปที่ 3.9 ตัวอย่างการวิเคราะห์ความคล้ายคลึงกันของข้อความ (Text Similarity)	31
รูปที่ 3.10 ตัวอย่างคะแนนความคล้ายคลึงกันของข้อความ (Text Similarity) ในแต่ละวัน	32
รูปที่ 3.11 ตัวอย่างการรวมคะแนนความคล้ายคลึงกันของคำ (Text Similarity) ในแต่ละวัน	32
รูปที่ 3.12 อัตราส่วนข่าวที่เกิดขึ้นแต่ละประเภทของแต่ละวัน	33
รูปที่ 3.13 การเปลี่ยนแปลงราคาหุ้นในทิศทางขึ้น	34

สารบัญรูป (ต่อ)

รูปที่	หน้า
รูปที่ 3.14 การเปลี่ยนแปลงราคาหุ้นในทิศทางลง	34
รูปที่ 3.15 การเปลี่ยนแปลงราคาหุ้นที่ไม่มีทิศทาง	34
รูปที่ 3.16 ตัวอย่างการคำนวณ Rolling Correlation ของ S&P500 และ Crude Oil Future	35
รูปที่ 3.17 แนวโน้มหลักขาขึ้น	36
รูปที่ 3.19 แนวโน้มระยะสั้นขาขึ้น	37
รูปที่ 3.20 แนวโน้มระยะสั้นขาลง	38
รูปที่ 3.21 ระยะห่างของเส้น EMA และราคา	39
รูปที่ 3.22 จุดที่มีนัยสำคัญสำหรับแนวโน้มหลักขาขึ้น	39
รูปที่ 3.23 จุดที่มีนัยสำคัญสำหรับแนวโน้มหลักขาลง	40
รูปที่ 3.24 การเปลี่ยนแปลงแนวโน้มหลัก	40
รูปที่ 3.25 การเปลี่ยนแปลงราคาหุ้นที่ไม่มีทิศทาง	41
รูปที่ 3.26 ปริมาณการซื้อขายที่ผิดปกติ	42
รูปที่ 4.1 ตารางเมทริกซ์สับสนของการพยากรณ์ราคา DELTA	68
รูปที่ 4.2 ประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ DELTA	68
รูปที่ 4.3 ตารางเมทริกซ์สับสนของการพยากรณ์ราคา AOT	75
รูปที่ 4.4 ตารางประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ AOT	75
รูปที่ 4.5 ตารางเมทริกซ์สับสนของการพยากรณ์ราคา PTT	81
รูปที่ 4.6 ประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ PTT	82
รูปที่ 4.7 ตารางเมทริกซ์สับสนของการพยากรณ์ราคา ADVNAC	88
รูปที่ 4.8 ประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ ADVANC	88
รูปที่ 4.9 ตารางเมทริกซ์สับสนของการพยากรณ์ทิศทางราคา GULF	95
รูปที่ 4.10 ประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ GULF	95

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ตลาดหลักทรัพย์ หรือตลาดหุ้นนั้นเป็นตลาดศูนย์กลางสำหรับการซื้อขายหลักทรัพย์ อาทิเช่น หุ้นสามัญ (Common Stock) หุ้นบุริมสิทธิ (Preferred Stock) ใบสำคัญแสดงสิทธิ (Warrant) เป็นต้น ซึ่งตลาดหลักทรัพย์จะประกอบด้วยบริษัทต่างๆ ที่ต้องการออกหลักทรัพย์เพื่อระดมเงินทุนจากนักลงทุน ตลาดหลักทรัพย์แห่งประเทศไทยถูกควบคุมดูแลโดยสำนักงานคณะกรรมการกำกับหลักทรัพย์และตลาดหลักทรัพย์ นักลงทุนสามารถซื้อขายหลักทรัพย์ผ่านระบบอิเล็กทรอนิกส์ (Streaming) ได้โดยตรงในช่วงเวลาทำการ บริษัทที่จดทะเบียนในตลาดหลักทรัพย์มีการประกอบธุรกิจที่แตกต่างกันออกไป ผลดำเนินงานของแต่ละบริษัทจึงมีความแตกต่างกันด้วย ผลการดำเนินงานเหล่านั้นล้วนมีความสอดคล้องกับสรุปเศรษฐกิจในการดำเนินงานธุรกิจไม่มากก็น้อยและยังมีผลต่อการตัดสินใจของนักลงทุนที่จะซื้อขายหลักทรัพย์ต่างๆ เหล่านั้น ทำให้ราคาของหลักทรัพย์มีการเปลี่ยนแปลงไปในทิศทางต่างๆ ที่ยากต่อการคาดการณ์ การที่เราสามารถคาดการณ์ทิศทางราคาหุ้นได้อย่างแม่นยำนั้นล้วนเป็นประโยชน์อย่างมากสำหรับการวางแผนกลยุทธ์การซื้อขาย ทั้งผู้ที่ต้องบริหารหรือควบคุมความเสี่ยง ผู้ที่ต้องเก็งกำไร และผู้ที่ต้องการลงทุน แต่เนื่องจากตลาดหุ้นนั้นมีปัจจัยต่างๆ ที่ส่งผลกระทบต่อ การเปลี่ยนแปลงราคาหุ้น เช่น เศรษฐกิจ การเมือง ภัยธรรมชาติ เป็นต้น สิ่งต่างๆ เหล่านี้ทำให้ทิศทางการเปลี่ยนแปลงราคาหุ้นมีการเคลื่อนที่แบบผันผวน ไม่คงที่และไม่แน่นอนตลอดเวลา จึงยากต่อการคาดการณ์ทิศทางราคาหุ้นเป็นอย่างมาก

ดังนั้นการวางแผนกลยุทธ์การซื้อขายอย่างมีประสิทธิภาพได้นั้น จึงจำเป็นต้องเก็บรวบรวมข้อมูล ที่ช่วยให้นักลงทุนสามารถตัดสินใจและวางแผนกลยุทธ์การซื้อขายได้อย่างมีประสิทธิภาพ ทั้งข้อมูลข่าวสารและเหตุการณ์ต่างๆ ที่เกิดขึ้นในช่วงเวลานั้นรวมถึงการใช้ข้อมูลด้านเทคนิค คอลอนาไลซิส (Technical Analysis) เข้ามาประกอบกับการตัดสินใจ ตามหลักเศรษฐศาสตร์การที่ราคา จะเปลี่ยนแปลงได้นั้นเกิดจากความต้องการซื้อ (Demand) ความต้องการขาย (Supply) เมื่อไรที่ความต้องการซื้อ มากกว่าความต้องการขายราคาจะมีการเปลี่ยนแปลงในทิศทางขึ้น เมื่อไรที่ความต้องการซื้อ น้อยกว่าความต้องการขายราคาจะมีการเปลี่ยนแปลงในทิศทางลง การที่นักลงทุนจะตัดสินใจซื้อหรือขาย

ได้นั้น ล้วนเกิดรับรู้ถึงข่าวสารต่างๆ หรือสิ่งที่เคยพบเจอในอดีต สำหรับการประเมินความเสี่ยงหรือความเสียหายที่จะเกิดขึ้น ณ ปัจจุบัน เพื่อให้เกิดความคุ้มค่าในการลงทุน

ผู้วิจัยจึงมีความสนใจที่จะศึกษาการพยากรณ์ทิศทาง การเปลี่ยนแปลงราคาหุ้นโดยการนำปัจจัยข่าวสารร่วมกับปัจจัยด้านเทคนิคคอลอนาไลซิส (Technical Analysis) ในการสร้างแบบจำลองการเรียนรู้ของเครื่องจักร (Machine Learning) เพื่อช่วยสำหรับการตัดสินใจวางแผนกลยุทธ์การซื้อขายให้ได้ผลตอบแทนที่คุ้มค่า

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 ศึกษาปัจจัยที่มีผลต่อการเปลี่ยนแปลงทิศทางราคาหุ้น เช่น ปัจจัยข่าว และปัจจัยด้านเทคนิคในตลาดหุ้น

1.2.2 สร้างแบบจำลองสำหรับการพยากรณ์ทิศทาง การเปลี่ยนแปลงราคาหุ้นในประเทศไทย โดยใช้ปัจจัยเชิงคุณภาพและปัจจัยเชิงปริมาณ

1.2.3 เปรียบเทียบประสิทธิภาพการใช้ปัจจัยต่างๆ ในการสร้างแบบจำลองที่ช่วยการพยากรณ์ทิศทาง การเปลี่ยนแปลงราคาหุ้นได้อย่างแม่นยำ

1.3 ขอบเขตของงานวิจัย

1.3.1. ขอบเขตระยะเวลาดำเนินการ

ระยะเวลาตั้งแต่เดือนพฤศจิกายน พ.ศ. 2565 ถึง เดือนพฤษภาคม พ.ศ. 2566 ทั้งหมด
ระยะเวลา 6 เดือน

1.3.2 ขอบเขตด้านข้อมูล

1) รวบรวมข้อมูลตัวแปรเชิงคุณภาพ ได้แก่ ข่าวด้านเศรษฐกิจ และข่าวหุ้นรายตัวของต่างประเทศรายวัน ตั้งแต่วันที่ 25 พฤษภาคม พ.ศ.2561 ถึง วันที่ 15 มกราคม พ.ศ.2566

2) รวบรวมข้อมูลตัวแปรเชิงปริมาณ ได้แก่ ราคาดัชนีต่างประเทศรายวัน (DJIA Index) ราคาดัชนีในประเทศรายวัน (SET Index) ราคากลุ่มอุตสาหกรรมในประเทศ (Sector Index) ราคาหุ้นรายวันในประเทศไทยในกลุ่ม SET50 และ ปริมาณการซื้อขายรายวัน (Volume) เป็นต้น ตั้งแต่วันที่ 25 พฤษภาคม พ.ศ.2561 ถึง วันที่ 15 มกราคม พ.ศ.2566 โดยใช้ข้อมูลสำหรับวิเคราะห์ในช่วงเดือน มกราคม พ.ศ.2566

1.3.3 ขอบเขตเครื่องมือการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1) เครื่องมือสำหรับการจัดการข้อมูลและสร้างแบบจำลองโดยใช้โปรแกรมภาษาไพธอน (Python)
- 2) เครื่องมือสำหรับการเก็บข้อมูล เช่น โปรแกรมไมโครซอฟต์เอ็กเซล (Excel) และโปรแกรมโน้ตแพด (Notepad)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 เข้าใจหลักการประยุกต์ใช้ปัจจัยเชิงคุณภาพและปัจจัยเชิงปริมาณในการสร้างแบบจำลอง
- 1.4.2 เข้าใจปัจจัยที่มีผลต่อการเปลี่ยนแปลงของราคาหุ้น
- 1.4.3 สามารถนำแบบจำลองเพื่อกำหนดกลยุทธ์ซื้อขายหุ้นในกลุ่ม SET50



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาหัวข้องานวิจัยเรื่อง “การทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและการวิเคราะห์เชิงเทคนิค” ผู้วิจัยได้ทำศึกษาค้นคว้าเกี่ยวกับแนวคิดทฤษฎี เอกสารและงานวิจัยต่างๆ ที่เกี่ยวข้อง ดังต่อไปนี้

1. ข้อมูล SET50
2. ปัจจัยข่าวทางเศรษฐกิจ (Economic Factor)
3. ปัจจัยเชิงเทคนิค (Technical Analysis)
4. เทคนิคการเลือกตัวแปร (Feature Selection)
5. การเรียนรู้ของเครื่องจักร (Machine Learning)
6. การประมวลภาษาธรรมชาติ (Natural Language Processing)
7. การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation)

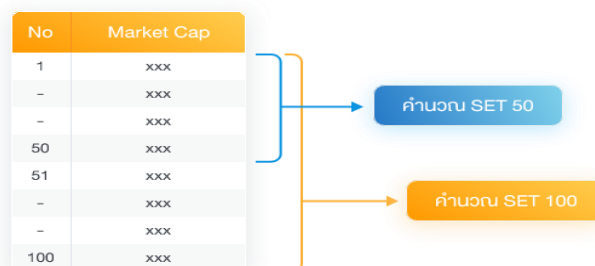
2.1 ข้อมูล SET50

ดัชนีราคา SET50 Index เพื่อส่งเสริมการออกตราสารอนุพันธ์และเป็นเครื่องมือวัดสถานะตลาดสำหรับกองทุนรวมต่างๆ ตลาดหลักทรัพย์ฯ จึงจัดทำ ดัชนีราคา SET50 Index เพื่อเป็นดัชนีราคาหุ้นที่ใช้แสดงระดับและความเคลื่อนไหวของราคาหุ้นสามัญ 50 ตัวที่มีมูลค่าตามราคาตลาด (Market Capitalization) สูง การซื้อขายมีสภาพคล่องสูงอย่างสม่ำเสมอและมีสัดส่วนผู้ถือหุ้นรายย่อยผ่านเกณฑ์ที่กำหนด [1]

2.1.1 เกณฑ์การคำนวณดัชนี SET50

- 1) คำนวณแบบถ่วงน้ำหนักด้วยมูลค่าหลักทรัพย์ตามราคาตลาด (Market Capitalization Weight)
- 2) หลักทรัพย์ที่ถูกขึ้นเครื่องหมาย SP เกิน 20 วัน ไม่นำมารวมในการคำนวณ
- 3) SET50 Index คำนวณโดยใช้หุ้นสามัญจดทะเบียนที่ผ่านการคัดเลือก 50 อันดับแรก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.1 ตัวอย่างการคำนวณ SET50

ที่มา : <https://www.set.or.th/th/market/index/set50/profile>

2.1.2 เกณฑ์ในการคัดเลือกของหลักทรัพย์ใน SET50

- 1) คำนวณแบบถ่วงน้ำหนักด้วยมูลค่าหลักทรัพย์ตามราคาตลาด
- 2) หลักทรัพย์ที่ถูกขึ้นเครื่องหมาย SP เกิน 20 วัน ไม่นำมารวมในการคำนวณ
- 3) สัดส่วนผู้ถือหลักทรัพย์รายย่อย (Free-float) ไม่น้อยกว่าร้อยละ 20 ของทุนชำระแล้ว
- 4) SET50 คำนวณโดยใช้หุ้นสามัญจดทะเบียนที่ผ่านการคัดเลือก 50 อันดับแรก

ตารางที่ 2.1 รายชื่อหลักทรัพย์ดัชนี SET50 ช่วงเดือน พฤศจิกายน พ.ศ.2565 [2]

หลักทรัพย์	เปิด	สูงสุด	ต่ำสุด	ล่าสุด	ปริมาณ (หุ้น)	มูลค่า ('000 บาท)
ADVANC	188	190	187.5	190	7,922,992	1,499,378.55
AOT	74.5	74.5	73.5	74	25,501,398	1,883,592.44
AWC	6	6	5.8	5.85	37,686,329	221,712.74
BANPU	12.5	12.5	12.3	12.4	43,296,642	537,998.24
BBL	143.5	144	142.5	142.5	4,421,217	632,038.11
BDMS	29	29.5	28.75	29.25	44,755,913	1,304,572.66
BEM	9.45	9.5	9.35	9.45	32,264,614	304,405.27
BGRIM	36.75	37	36.25	36.75	6,543,673	239,515.58
BH	213	216	209	212	7,573,465	1,608,493.61
BLA	30	30.75	30	30.25	5,561,271	169,639.32
BTS	8.15	8.2	8.05	8.1	28,530,414	231,682.29
CBG	96.5	97	95.75	96.5	4,980,472	480,101.55
CPALL	62.5	63.5	62.5	62.75	33,161,911	2,086,690.96
CPF	24.3	24.4	24.2	24.4	14,900,044	362,145.22

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

CPN	68	69	67.75	68.75	5,218,024	357,950.45
CRC	42	42.25	41.75	42	7,404,082	310,940.24
DTAC	44.5	44.75	44.25	44.25	3,354,027	149,082.06
EA	95.25	95.5	93	94	8,968,982	843,552.28
EGCO	168.5	169	167.5	168	1,041,193	175,079.17
GLOBAL	19.9	20	19.4	19.9	11,400,314	224,099.51
GPSC	66.75	67.25	66.25	66.5	3,078,286	205,134.70
GULF	50.75	51.25	50.25	50.5	9,024,233	456,822.17
HMPRO	14.8	14.9	14.6	14.8	29,389,518	433,862.05
INTUCH	73.75	74.75	73.75	74.25	3,328,992	247,570.70
IRPC	3.04	3.04	2.98	3	104,867,464	314,810.00
IVL	42.25	42.5	42.25	42.5	4,716,312	200,046.63
JMART	45	45.5	43.75	44	7,105,068	314,706.41
JMT	66.5	66.75	64.5	65	5,119,540	335,371.55
KBANK	143.5	143.5	142	143	9,260,498	1,321,367.56
KCE	48.25	52	48.25	51.25	43,894,922	2,198,099.11
KTB	17.6	17.6	17.4	17.4	28,039,914	489,995.60
KTC	59	59	58	58.5	2,686,615	157,227.34
LH	9.3	9.35	9.25	9.3	17,439,907	162,257.03
MINT	31	31	30	30.25	17,222,872	522,246.15
MTC	35.5	35.5	34.75	35.25	13,123,583	461,847.91
OR	24.2	24.4	24.2	24.2	11,037,953	267,594.08
OSP	28	28.25	27.75	28.25	2,828,270	79,441.64
PTT	33	33.5	32.75	33.5	71,045,189	2,355,661.12
PTTEP	186	187.5	185	187	8,725,193	1,626,874.97
PTTGC	46.5	47.25	46	47.25	12,587,556	588,201.25
SAWAD	41.75	42.5	41.5	42	8,607,831	360,465.19
SCB	106	106	104.5	105.5	6,300,636	662,982.37
SCC	345	346	343	344	1,789,399	616,254.00
SCGP	56.25	56.75	56	56.25	4,065,579	229,684.27
TIDLOR	26	26.25	25.75	26	5,033,120	130,774.79
TISCO	96.75	96.75	96	96.5	3,273,794	315,517.28
TOP	53.75	55	53.5	54.5	13,874,967	752,433.34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

TRUE	4.94	4.96	4.9	4.94	31,298,200	154,478.17
TTB	1.37	1.38	1.35	1.36	92,258,723	125,952.80
TU	17	17.4	16.9	17.2	27,325,604	469,978.42

ที่มา : <https://www.set.or.th/th/market/index/set50/overview>

2.2 ปัจจัยส่งผลกระทบต่อตลาดหุ้น (Factors affecting the stock market)

2.2.1 เศรษฐกิจ (Economic)

ตลาดหลักทรัพย์ได้รับผลกระทบจากปัจจัยทางเศรษฐกิจหลายประการ เช่น เหตุการณ์ทางการเมือง นโยบายของบริษัท สภาวะเศรษฐกิจทั่วไป คาดการณ์ของนักลงทุน การเลือกลงทุนของนักลงทุนสถาบัน การเคลื่อนไหวของตลาดหลักทรัพย์อื่น ๆ และจิตวิทยาของนักลงทุน ฯ (Tan, Quek, & Ng, 2007) [3]

งานศึกษา “ความผันแปรของเวลาในความสัมพันธ์ระหว่างราคาหุ้นกับกิจกรรมทางเศรษฐกิจ” ได้ทดสอบว่าราคาหุ้นตอบสนองอย่างไรต่อการเปลี่ยนแปลงในการผลิต และพบว่ามีผลบวกที่มีนัยสำคัญต่อราคาหุ้น อันเนื่องจากการเพิ่มกิจกรรมทางเศรษฐกิจหรือการลงทุนทางธุรกิจส่งผลให้มีกระแสเงินเกิดขึ้นในอนาคต ดังนั้นนักลงทุนที่คาดหวังว่าจะได้รับเงินปันผลที่สูง จะมีความต้องการซื้อหุ้นที่เพิ่มขึ้น (McMillan, 2005) [4]

ปานตา ฉัตรมาศ (2565) ดัชนี Dow Jones หรือ Dow Jones Industrial Average (DJIA) ที่ประกอบด้วยหุ้นขนาดใหญ่ 30 บริษัทในตลาดหุ้น New York (NYSE) และตลาดหุ้น Nasdaq โดยได้รับการคัดเลือกจากคณะกรรมการของบริษัท S&P Dow Jones Indices ซึ่งเป็นเจ้าของดัชนี Dow Jones ในปัจจุบัน หุ้นที่อยู่ในดัชนี Dow Jones มาจากหลากหลายกลุ่มอุตสาหกรรม เช่น หุ้น 3M (กลุ่มอุตสาหกรรม), Apple (กลุ่ม เทคโนโลยี), Coca-Cola (กลุ่มสินค้าอุปโภคบริโภค) และ Goldman Sachs (กลุ่มธุรกิจการเงิน) การคำนวณ ดัชนี Dow Jones แตกต่างจากการคำนวณดัชนีอื่นๆ โดยเป็นการนำราคาของหุ้นทั้ง 30 บริษัท มาบวกรวมกันแล้วหารด้วย 30 ดังนั้น หุ้นของบริษัทที่มีราคาสูงจะมีผลกับดัชนีมากกว่าหุ้นของบริษัทที่มีราคาต่ำส่งผลให้การเคลื่อนไหวของดัชนี ไม่สามารถสะท้อนรูปรวมของการลงทุนได้ดีมากนัก [5]

จากงานวิจัย “ปัจจัยที่ส่งผลต่อดัชนีราคาตลาดหลักทรัพย์ SET 50” รัชณี รุ่งศรีรัตนวงศ์ (2553) ได้อธิบายว่า ดัชนีอุตสาหกรรมดาวโจนส์ สามารถอธิบายการเปลี่ยนแปลงดัชนีราคาตลาดหลักทรัพย์ SET 50 ในทิศทางเดียวกัน ได้อย่างมีนัยสำคัญทางสถิติ เป็นไปตามสมมติฐานและเป็นไปตามแนวคิด ทฤษฎีประสิทธิภาพตลาด ดังนี้ จากทฤษฎีประสิทธิภาพตลาด แสดงให้เห็นว่า มูลค่าที่ควรเป็นของหุ้นจะเท่ากับ

ราคาตลาด เสมอ แสดงให้เห็นว่าการเปลี่ยนแปลงของราคาหุ้น ณ เวลาใดเวลาหนึ่งจะเป็นข้อมูลสะท้อนข่าวสารอย่างสมบูรณ์หากการตัดสินใจซื้อขายหุ้นของนักลงทุนในตลาดตั้งอยู่บนพื้นฐานของการคาดคะเนด้วยเหตุผล (Rational expectations) ราคาหุ้นจะปรับตัวสูงขึ้นหรือลดลงอย่างรวดเร็วเมื่อมีข้อมูลข่าวสารใหม่ๆ เข้ามา ผลกำไรที่นักลงทุนได้รับจะเป็นกำไรในระดับปกติ (Normal Profit) ดังนั้นการที่ดัชนีตลาดอุตสาหกรรมดาวโจนส์เพิ่มสูงขึ้น หมายถึงสัญญาณที่ดี ทั้งในทิศทางเศรษฐกิจและความเชื่อมั่นของนักลงทุนเพิ่มสูงขึ้น ข่าวสารและข้อมูลดังกล่าวจึงสะท้อนมายังดัชนีราคาตลาดหลักทรัพย์ SET 50 เพิ่มสูงขึ้นเช่นเดียวกัน [6]

2.2.2 กิจกรรมทางการเมือง (Political Events)

งานศึกษา “ผลกระทบของเหตุการณ์ทางการเมืองต่อปริมาณการซื้อขายและหุ้นของดัชนี KSE” ได้ตรวจสอบผลกระทบของความไม่แน่นอนทางการเมืองต่อปริมาณการซื้อขายในตลาดหุ้นและผลตอบแทนของหุ้นรายวัน การศึกษาในช่วงเวลาตั้งแต่ปี 2551 ถึง 2552 ผลการวิจัยชี้ให้เห็นถึงความสัมพันธ์เชิงลบระหว่างปริมาณการซื้อขายและผลตอบแทนของหุ้น หลายเหตุการณ์ถูกระบุว่าเป็นกิจกรรมทางการเมืองมีอิทธิพลต่อตลาดหุ้น เช่น การทำสงคราม การโจมตีระหว่างประเทศ การลอบสังหารรัฐมนตรี นั้นส่งผลกระทบต่อปริมาณการซื้อขายหุ้น เหตุการณ์เหล่านี้ถูกสังเกตว่ามีผลกระทบอย่างมีนัยสำคัญต่อการเคลื่อนไหวของ KSE 100-INDEX เพื่อดูว่าดัชนีได้รับอิทธิพลจากเหตุการณ์ที่เป็นบวกหรือลบ (Malik, Hussain, & Ahmed, 2009) [7]

2.2.3 กิจกรรมทางบริษัท (Company Events)

การศึกษาของ Halsey (2000) พบว่าราคาหุ้นมีแนวโน้มสูงขึ้นและลดลงต่ำกว่าแนวโน้มในระยะยาวเมื่อมีการรายงานเกี่ยวกับรายได้ ความผันผวนตามวัฏจักรเหล่านี้ถูกสังเกตว่ามีการเกิดขึ้นหลายช่วงเวลา ซึ่งไม่ได้รวมถึงการรายงานการคงค้างในระยะสั้น แต่จะเป็นตัวแทนของความเบี่ยงเบนในแนวโน้มระยะยาวของราคาและรายได้ การศึกษายังเผยให้เห็นว่ารายได้ของบริษัทมีความผันผวนทั้งด้านบวกและด้านลบของแนวโน้มในระยะยาว และพบว่ารายงานรายได้เหล่านี้มีความเกี่ยวข้องกับการเปลี่ยนแปลงกับราคาหุ้น นอกจากนี้รายได้ของบริษัทยังเป็นสัญญาณบ่งชี้ว่ารายได้ของบริษัทมีความเชื่อมโยงกับกิจกรรมทางเศรษฐกิจเพียงบางส่วน นั้นหมายความว่ายังมีปัจจัยอื่น ๆ ที่มีผลต่อความผันผวนของเศรษฐกิจอีกด้วย

ในด้านการศึกษา Fama, Fisher, Jensen และ Roll (1969) ซึ่งเป็นการศึกษาที่เก่าแก่ที่สุดชิ้นหนึ่ง แม้ว่าการศึกษาที่ตีพิมพ์ครั้งแรกโดย Ball Brown (1968) การศึกษาเหตุการณ์เหล่านี้ได้ทำการศึกษาเกี่ยวกับผลกระทบของราคาหุ้นต่อการแตกพาร์หุ้นและการประกาศผลประกอบการ โดยใช้แบบจำลอง

ตลาด (Market Model) หรือแบบจำลองการกำหนดราคาสินทรัพย์ทุนเป็นเกณฑ์มาตรฐาน (Capital Asset Pricing Model) การค้นพบว่าผลกระทบทั้งสองกรณี ตลาดมักมีการคาดการณ์ข้อมูลไว้ล่วงหน้า และส่วนใหญ่ราคามักมีการปรับตัวขึ้นก่อนที่เหตุการณ์นั้นๆ เมื่อมีข่าวออกจริงๆ ราคาก็จะมีการปรับตัวขึ้นอย่างรวดเร็ว การศึกษาของ Fama, Fisher, Jensen และ Roll แสดงให้เห็นอย่างชัดเจนว่าราคาหุ้นไม่เพียงสะท้อนถึงการประมาณการของประสิทธิภาพของบริษัทในอนาคต แต่ยังรวมถึงมูลค่าราคาที่เหมาะสมของบริษัทอีกด้วย [8-9]

2.2.4 อารมณ์ของนักลงทุน (Investor Sentiment)

การวิเคราะห์ตลาดหุ้นต้องพิจารณาถึงความเป็นไปได้ของการเปลี่ยนแปลงราคาหุ้นตามความรู้สึกของนักลงทุน แต่ละบริษัทอาจมีการตอบสนองแตกต่างกัน การเก็งกำไรในตลาดหุ้นยากและคาดการณ์ความรู้สึกของนักลงทุนยังเป็นเรื่องที่ทำนาย ตัวอย่างเช่น กลุ่มนักลงทุนบางกลุ่มอาจมีความเชื่อมั่นต่อกำไรบริษัทสูงมาก ซึ่งจะส่งผลให้ราคาหุ้นมีแนวโน้มการเปลี่ยนแปลงมากขึ้น คล้ายกับการเล่นลอตเตอรี่เพื่อเสี่ยงดวงถูกรางวัล ดังนั้น ความเชื่อมั่นเป็นปัจจัยที่ส่งผลต่อความต้องการและเมื่อความเชื่อมั่นเพิ่มส่งผลให้ราคามีการเปลี่ยนแปลงอย่างมีนัยและผลตอบแทนที่สูงขึ้น (Baker & Wurgler, 2007) [10]

2.2.5 แนวโน้มของตลาด (Market Moment)

เครื่องมือทางเทคนิคที่ใช้ในการวิเคราะห์ตลาดหุ้นมีจุดมุ่งหมายที่แตกต่างกัน โดยเครื่องมือที่ใช้บอกถึงการเคลื่อนไหวของแนวโน้มราคา เช่น Simple Moving Average (SMA) Weighted Moving Average (WMA) และ Exponential Moving Average (EMA) จึงมีความเหมาะสมกับสถานการณ์ของตลาดที่แตกต่างกันในแต่ละช่วงเวลา ดังนั้นจึงมีการใช้เครื่องมือทางเทคนิคอื่นเพื่อให้ความเชื่อถือได้มากขึ้นอย่าง Moving Average Convergence Divergence (MACD) และ Relative Strength Index (RSI) ซึ่งเครื่องมือเหล่านี้ให้ข้อมูลเกี่ยวกับแนวโน้มและเสถียรรูปของราคาหุ้น รวมถึงความเป็นไปได้ในการเปลี่ยนแปลงแนวโน้มของราคา ในขณะเดียวกัน เครื่องมือ Stochastic Oscillators K และ D เป็นการเปรียบเทียบราคาปิดกับช่วงราคาในระยะเวลาที่กำหนด เพื่อให้ข้อมูลเกี่ยวกับเคลื่อนที่ของราคา โดยสรุปว่าเครื่องมือเหล่านี้ให้ข้อมูลสำหรับการวิเคราะห์และเข้าใจแนวโน้มและการเคลื่อนไหวของตลาดหุ้น (Bhargavi et al. 2017; Bustos and Pomares-Quimbaya 2020; Perry 2011; Praekhaow 2010; Vaidya 2018) [11-15]

2.2.6 ภัยธรรมชาติ (Natural Disasters)

จากการศึกษาตลาดหุ้นในออสเตรเลีย โดย Worthington และ Valadkhani (2004) โดยใช้ข้อมูลตั้งแต่วันที่ 31 ธันวาคม 2525 ถึง 01 มกราคม 2545 พบว่าไฟป่า ไซโคลน และแผ่นดินไหวมีผลกระทบอย่างมากต่อผลตอบแทนของตลาดหุ้น ในขณะที่พายุรุนแรงและน้ำท่วมไม่ได้ให้ผลในแบบเดียวกัน สุดท้ายแล้วผลกระทบที่เกิดขึ้นเป็นไปได้ทั้งเชิงบวกและหรือเชิงลบ ผลกระทบส่วนใหญ่เหล่านี้ที่ขึ้นจะมีผลกับตลาดหุ้นในทันที แต่จะมีบางเหตุการณ์ที่จะสะท้อนในวันต่อๆ ไป (Worthington & Valadkhani, 2004) [16]

2.2.7 กิจกรรมในภาคอุตสาหกรรม (Industry Events)

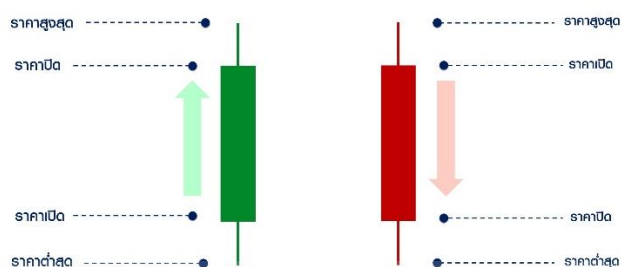
การศึกษาความสัมพันธ์ระหว่างตลาดหุ้น ราคาน้ำมัน และปรากฏการณ์การแพร่กระจาย (Spillover Effect) มีความสำคัญอย่างมาก เนื่องจากช่วยให้นักลงทุนสามารถป้องกันความเสี่ยงของพอร์ตการลงทุนจากการเปลี่ยนแปลงของตลาดอย่างกะทันหันได้ การเปลี่ยนแปลงราคาน้ำมันไม่เพียงแต่ส่งผลกระทบต่อต้นทุนเท่านั้น แต่ยังมีผลทางอ้อมต่อวัฏจักรธุรกิจและความผันผวนของตลาดหุ้นด้วย (Bastianin, Conti, & Manera, 2016) [17]

2.3 ปัจจัยเชิงเทคนิค (Technical Analysis)

2.3.1 แท่งเทียน (Candle Stick) [18]

เครื่องมือการวิเคราะห์ราคาหุ้นที่ใช้ในการแสดงข้อมูลเกี่ยวกับการเคลื่อนไหวของราคาหุ้นในช่วงเวลาหนึ่ง ซึ่งเป็นข้อมูลที่สำคัญในการวิเคราะห์แนวโน้มและการเคลื่อนไหวของราคาหุ้นในช่วงเวลานั้นๆ แท่งเทียนสามารถแบ่งออกเป็นแท่งเทียนขาขึ้น (Bullish candlestick) และแท่งเทียนขาลง (Bearish candlestick) โดยจะใช้สีแตกต่างกันเพื่อแสดงถึงแนวโน้มของราคาหุ้น ราคาปิดสูงกว่าราคาเปิดแท่งเทียนสีเขียว ราคาปิด ต่ำกว่าราคาเปิดแท่งเทียนสีแดง ดังรูปที่ 2.2 แท่งเทียนมีโครงสร้างที่ประกอบด้วยส่วนต่างๆ เช่น ราคาเปิด (Open) ราคาปิด (Close) ราคาสูงสุด (High) และราคาต่ำสุด (Low)

- ราคาเปิด เป็นราคาซื้อขายแรกที่เกิดขึ้นตั้งแต่เปิดตลาด
- ราคาสูงสุด การเคลื่อนไหวของราคาหุ้น ณ ระดับราคาสูงสุดในวัน
- ราคาต่ำสุด การเคลื่อนไหวของราคาหุ้น ณ ระดับราคาต่ำสุดในวัน
- ราคาปิด เป็นราคาสุดท้ายที่เกิดขึ้นจากการซื้อขายสิ้นสุดของวัน



รูปที่ 2.2 ส่วนประกอบของแท่งเทียน

ที่มา : <https://knowledge.bualuang.co.th/knowledge-base/candlesticks/>

2.3.2 ปริมาณการซื้อขาย (Volume) [19]

ปริมาณการซื้อขาย บ่งบอกถึงระดับความสนใจ หรือ ความกระตือรือร้นของนักลงทุน ซึ่งใช้ยืนยันการเกิดหรือการเปลี่ยนแปลงแนวโน้มหลัก โดยให้สังเกตวันหรือช่วงเวลาที่มียปริมาณการซื้อขายสูงสุด (volume peak) อาจเกิดแนวโน้มใหม่หรือจบแนวโน้มเดิมดังรูปที่ 2.3

กรณีราคาปรับตัวขึ้น

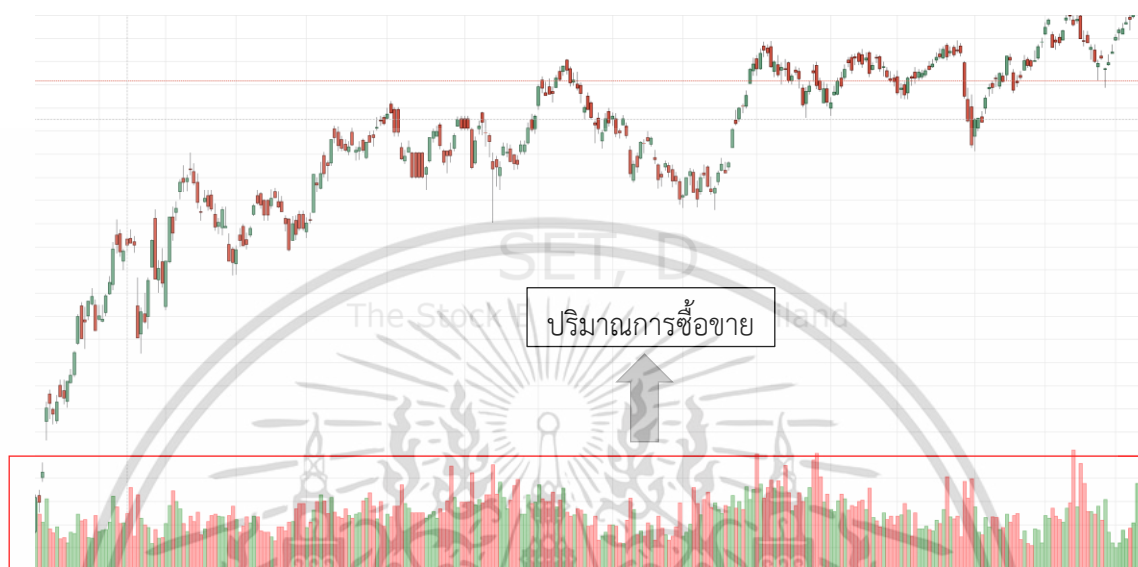
- ราคาปรับตัวขึ้น และ ปริมาณการซื้อขายขึ้น แสดงถึง ราคาขึ้นอย่างมีพลัง และมีโอกาสขึ้นต่อ
- ราคาปรับตัวขึ้น แต่ปริมาณการซื้อขายลง แสดงถึง ราคาขึ้นอย่างอ่อนพลัง
- ราคาปรับตัวขึ้นอย่างต่อเนื่อง และปริมาณการซื้อขายขึ้นอย่างมาก แสดงถึง ราคาอาจใกล้สิ้นสุดการขึ้น มีแนวโน้มการกลับตัวลงสูง
- ราคาปรับตัวขึ้น ถึงยอดเดิม แต่ปริมาณการซื้อขายปัจจุบันน้อยกว่าปริมาณการซื้อขาย ยอดเก่าในอดีต แสดงถึง ราคาขึ้นต่อได้ยาก

กรณีราคาปรับตัวลง

- ราคาปรับตัวลง และปริมาณการซื้อขายมาก แสดงถึง ราคาลงอย่างมีพลัง และราคายังมีโอกาสลงต่อ
- ราคาปรับตัวลง แต่ปริมาณการซื้อขายน้อย แสดงถึง ราคาลงตามปกติ
- ราคาปรับตัวลงมานาน ต่อมาปริมาณการซื้อขายก็ขึ้นอย่างมาก แสดงถึง ราคาอาจใกล้สิ้นสุดการลง เป็นไปได้ว่า คนขายเกือบหมด โอกาสราคาเริ่มปรับตัวขึ้นเพิ่มมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ราคาปรับตัวลงถึงฐานเดิมแต่ปริมาณการซื้อขายปัจจุบันน้อยกว่าปริมาณการซื้อขายฐานเก่าในอดีต แสดงถึง ปฏิเสธการลงต่อของราคา



รูปที่ 2.3 ดัชนี SET50 และปริมาณการซื้อขาย

ที่มา : <http://siamchart.com>

2.3.3 เส้นค่าเฉลี่ยเคลื่อนที่แบบเอ็กซ์โพเนนเชียล (Exponential Moving Average: EMA)

[20]

เส้นค่าเฉลี่ยเคลื่อนที่แบบเอ็กซ์โพเนนเชียล เป็นการคำนวณประเภทหนึ่งของเส้นค่าเฉลี่ยแบบเคลื่อนที่ (Moving Average: MA) ซึ่งการหาค่าแบบ Exponential จะมองความสัมพันธ์ของราคาหุ้นย้อนหลังแบบถ่วงน้ำหนักในรูปแบบของเลขชี้กำลังโดยให้ความสำคัญกับราคาสุดท้ายมากที่สุด ดังนั้น เส้นค่าเฉลี่ยแบบ EMA จึงเคลื่อนไหวได้เร็วกว่าและตอบสนองต่อการเปลี่ยนแปลงของราคาไวกว่าการคำนวณแบบธรรมดา (Simple Moving Average: SMA) อย่างมีนัยสำคัญ จัดได้ว่าเป็นเครื่องมือทางเทคนิคที่ยอดนิยมของนักเก็งกำไร โดยที่นักลงทุนนิยมตั้งค่าเส้น EMA แตกต่างกันไป หากเป็นนักลงทุนระยะยาวก็จะใช้เส้น EMA มากขึ้น โดยนิยมใช้เส้น EMA5 EMA10 EMA25 EMA50 หรือ EMA200 วัน

เส้น EMA มีต้นกำเนิดมาจากการหาเส้นค่าเฉลี่ยอย่างง่าย โดยการคำนวณหาค่าเฉลี่ยปกติตามสูตรดังนี้

$$SMA = \frac{A_1 + A_2 + \dots + A_n}{n}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ A_n คือ ราคาปิดของวันที่ n

n คือ จำนวนวัน หรือ แท่งเทียนที่ใช้คำนวณ

หลังจากนักเทคนิคได้ใช้ SMA ในการหาแนวโน้มของราคาหุ้น ซึ่งสามารถตอบสนองต่อการเปลี่ยนแปลงของราคาได้อย่างดีในระยะสั้น แต่อาจตอบสนองในระยะยาวได้ช้ากว่า จึงได้เกิดการคำนวณหา เส้นค่าเฉลี่ยแบบถ่วงน้ำหนัก ที่สามารถช่วยตอบสนองต่อการเปลี่ยนแปลงของราคาหุ้นได้ดียิ่งขึ้น ซึ่งได้แก่ Weighted Moving Average (WMA) และ Exponential Moving Average (EMA)

ความแตกต่างของการคำนวณเส้นค่าเฉลี่ยของ SMA และ EMA นั้น คือ EMA จะมีการสะท้อนการเปลี่ยนแปลงของราคาที่รวดเร็วกว่าแบบ SMA เนื่องจากการคำนวณแบบเอ็กซ์โพเนนเชียลทำให้ EMA ให้น้ำหนักการคำนวณที่ราคาล่าสุดมากกว่าราคาย้อนหลัง ในขณะที่การคำนวณแบบ SMA จะให้น้ำหนักความสำคัญเท่ากันทุกราคา โดยสามารถหาเส้นค่าเฉลี่ย EMA จากสูตรดังนี้

$$EMA_{Today} = Value_{Today} * \left(\frac{Smoothing}{1 + Days} \right) + EMA_{Yesterday} * \left[1 - \left(\frac{Smoothing}{1 + Days} \right) \right]$$

$Value_{Today}$ คือ ราคา ณ ปัจจุบัน

$Smoothing$ คือ การให้น้ำหนักการเคลื่อนไหวของราคา

$EMA_{Yesterday}$ คือ เส้นค่าเฉลี่ยเคลื่อนที่ของวันก่อนหน้า

$Days$ คือ จำนวนที่ถูกคิดคำนวณสำหรับเส้นค่าเฉลี่ยเคลื่อนที่

นักเทคนิคมักใช้สูตรการคำนวณ EMA ที่มีการคำนวณ Period ราคาย้อนหลังแตกต่างกันหลายแบบ เช่น

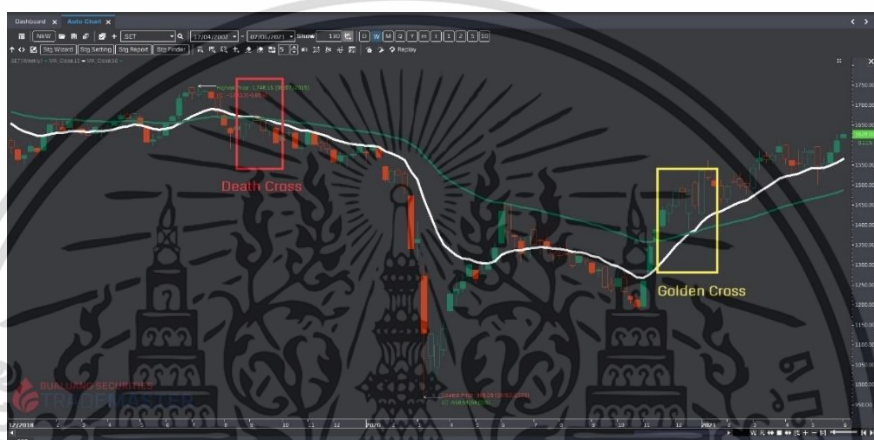
- ค่าเฉลี่ยเคลื่อนที่ 10 วัน (2 สัปดาห์) ใช้สะท้อนการเคลื่อนไหวของราคาหุ้นระยะสั้น
- ค่าเฉลี่ยเคลื่อนที่ 25 วัน (5 สัปดาห์) ใช้สะท้อนการเคลื่อนไหวของราคาหุ้นระยะสั้น ถึงระยะกลาง
- ค่าเฉลี่ยเคลื่อนที่ 50 วัน (10 สัปดาห์) ใช้สะท้อนการเคลื่อนไหวของราคาหุ้นระยะกลาง
- ค่าเฉลี่ยเคลื่อนที่ 200 วัน (40 สัปดาห์) ใช้สะท้อนการเคลื่อนไหวของราคาหุ้นระยะยาว

สัญญาณเตือนที่นักเทคนิคนิยมใช้บ่งบอกว่าหุ้นกำลังเปลี่ยนเป็น สัญญาณขาขึ้น หรือ สัญญาณขาลง โดยใช้เส้นค่าเฉลี่ย 2 เส้น เปรียบเทียบกัน แบ่งเป็น

- เส้นค่าเฉลี่ยระยะสั้น 15 วัน และเส้นค่าเฉลี่ยระยะยาว 50 วัน
 - เส้นค่าเฉลี่ยระยะสั้น 50 วัน และเส้นค่าเฉลี่ยระยะยาว 200 วัน
- หรืออาจจะใช้เป็นระยะอื่นๆ แทนได้

โดยกำหนดให้ เส้นสีขาวเป็นเส้น EMA ระยะสั้น 15 วัน และ เส้นสีเขียวเป็นเส้น EMA ระยะยาว 50 วัน ดังรูปที่ 2.4

- Golden Cross – จุดซื้อที่เกิดสัญญาณกระทิง (Bullish) หรือราคาหุ้นกำลังจะเปลี่ยนเป็นเทรนด์ขาขึ้น จะเห็นได้จากเส้น EMA ระยะสั้นตัดขึ้นเหนือเส้น EMA ระยะยาว
- Death Cross – จุดขายที่เกิดสัญญาณหมี (Bearish) หรือราคาหุ้นกำลังจะเปลี่ยนเป็นเทรนด์ขาลง จะเห็นได้จากเส้น EMA ระยะสั้นตัดลงต่ำกว่าเส้น EMA ระยะยาว



รูปที่ 2.4 กราฟดัชนี SET กับเส้นค่าเฉลี่ยเคลื่อนที่

ที่มา : <https://knowledge.bualuang.co.th/knowledge-base/whatisema>

2.3.4 เอ็มเอซีดี (Moving Average Convergence Divergence: MACD) [21]

ตัวชี้วัดที่บอกทิศทางแนวโน้มของราคาหุ้น (Trend) เป็นเครื่องมือที่มีแนวคิดจากเส้นค่าเฉลี่ยเคลื่อนที่ 2 เส้นที่มีค่าแตกต่างกัน โดยลักษณะเส้นเป็นแบบ Exponential เรียกอีกชื่อหนึ่งว่า Exponential Moving Average (EMA) หรือจะพูดอีกนัยหนึ่งว่า MACD คือ ระยะห่างระหว่างเส้นค่าเฉลี่ยเคลื่อนที่ 2 เส้น นั่นเอง และสามารถใช้ในการเคลื่อนที่ของ MACD กับ Signal line เพื่อจับจังหวะในการซื้อขายหุ้น โดยมีสูตรดังนี้

$$MACD = EMA(12) - EMA(26)$$

$EMA(12)$ คือ เส้นค่าเฉลี่ยเคลื่อนที่ 12 วัน

$EMA(26)$ คือ เส้นค่าเฉลี่ยเคลื่อนที่ 26 วัน

เส้นสัญญาณ (Signal line หรือ Trigger line) เป็นเส้นค่าเฉลี่ยเคลื่อนที่แบบ Exponential (EMA) ย้อนหลัง 9 วัน ของ MACD โดยที่ตัวมันเองไม่ใช่ Indicator อีกทั้งตัวเลขย้อนหลัง 9 วัน เป็นเพียง

การตั้งค่าเริ่มต้นที่คนส่วนใหญ่นิยมใช้กันเท่านั้น นักลงทุนสามารถปรับเปลี่ยนตัวเลขให้เหมาะกับกลยุทธ์การซื้อขายในระยะสั้น ระยะกลาง หรือระยะยาวของตนเองได้

$$\text{Signal line} = \text{EMA}(9)$$

EMA(9) คือ เส้นค่าเฉลี่ยเคลื่อนที่ 9 วัน

การตีความค่า Relative Strength Index RSI ดังรูปที่

- MACD > Signal line หรือ เส้น MACD ตัดขึ้นไปอยู่เหนือ Signal line หมายความว่า ราคาหุ้นมีแนวโน้มสูงขึ้น หรือส่งสัญญาณแนวโน้มขาขึ้น เป็นจังหวะในการเข้าซื้อ เรียกว่า “Bullish MACD”
- MACD < Signal line หรือ เส้น MACD ตัดลงมาอยู่ใต้เส้น Signal line หมายความว่า ราคาหุ้นมีแนวโน้มลดลง หรือส่งสัญญาณแนวโน้มขาลง เป็นการเตือนว่าควรขายหุ้น ออก เรียกว่า “Bearish MACD”



รูปที่ 2.5 กราฟหุ้นและการตีความค่า MACD

ที่มา : <https://knowledge.bualuang.co.th/knowledge-base/macd/>

2.3.5 อาร์เอสไอ (Relative Strength Index: RSI) [22]

เครื่องมือที่บ่งบอกสัญญาณแนวโน้มขาขึ้นและขาลงของราคา RSI ได้รับการพัฒนามาจาก Indicator ที่มีชื่อว่า Momentum ซึ่งใช้ในการวิเคราะห์ทางเทคนิคเพื่อวัดขนาดการเปลี่ยนแปลงของราคาล่าสุด และประเมินสถานะซื้อมากเกินไป (Overbought) หรือขายมากเกินไป (Oversold) ในราคาหุ้นหรือสินทรัพย์อื่นๆ โดยจะแสดงเป็นกราฟเส้นที่สามารถอ่านค่าได้ตั้งแต่ 0 – 100 โดยมีสูตรดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$RSI = 100 - \left(\frac{100}{1 + \frac{\text{Average gain}}{\text{Average loss}}} \right)$$

Average gain คือ ค่าเฉลี่ยของอัตราผลตอบแทนที่เป็นบวกจากการเปลี่ยนแปลงของราคา
ย้อนหลัง 14 แห่งเทียบ

Average loss คือ ค่าเฉลี่ยของอัตราผลตอบแทนที่เป็นลบจากการเปลี่ยนแปลงของราคา
ย้อนหลัง 14 แห่งเทียบ

หมายเหตุ: ส่วนใหญ่การคำนวณจะนิยมใช้ช่วง 14 แห่งเทียบ แต่นักลงทุนสามารถกำหนดจำนวนแห่ง
เทียบที่ใช้ในการคำนวณได้ ขึ้นอยู่กับสไตล์ของนักลงทุนแต่ละท่าน

การตีความค่า Relative Strength Index RSI

- เมื่อค่า RSI < 30 ราคาหุ้นจะถูกลง เนื่องจากมีสถานะขายมากเกินไป (Oversold) นัก
ลงทุนสามารถพิจารณาหาจังหวะในการเข้าซื้อได้
- หาก RSI > 70 ราคาหุ้นจะแพงขึ้น เนื่องจากมีสถานะซื้อมากเกินไป (Overbought) นัก
ลงทุนสามารถใช้เป็นจังหวะในการขายได้



รูปที่ 2.6 กราฟหุ้นและการตีความค่า RSI

ที่มา : <https://knowledge.bualuang.co.th/knowledge-base/rsi/>

2.4 เทคนิคการเลือกตัวแปร (Feature Selection)

เทคนิคการเลือกตัวแปร คือ การเลือกตัวแปรที่มีนัยสำคัญข้อมูลมากที่สุดจากชุดข้อมูล เมื่อ
ชุดข้อมูลขนาดใหญ่จะยากต่อการสร้างแบบจำลอง ทั้งในเรื่องระยะเวลาและประสิทธิภาพการคำนวณ
โดยเทคนิคการเลือกตัวแปรเป็นวิธีการใช้เลือกตัวแปรที่สำคัญหรือมีนัยสำคัญกับแบบจำลองเพื่อช่วยลด
ค่าใช้จ่ายในเรื่องระยะเวลาหรือไม่ให้สูญเสียความแม่นยำ [23]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประโยชน์การใช้เทคนิคการเลือกตัวแปร

1. การสอนอัลกอริทึมให้เร็วยิ่งขึ้น
2. ปรับปรุงประสิทธิภาพแบบจำลอง
3. ลดความซับซ้อนของตัวแปร
4. ลดปัญหาการเรียนรู้ที่มากเกินไป
5. ความเข้าใจตัวแปรในทางธุรกิจ
6. การตีความของแบบจำลองและข้อมูลที่ยากขึ้น

วิธีการการเลือกตัวแปรมี 2 วิธีการหลัก [24]

- การเลือกตัวแปรแบบไปข้างหน้า (Forward feature selection) โดยการพิตติงแบบจำลองกับ 1 ตัวแปร และเพิ่มตัวแปรเข้ามาในแบบจำลองจนกว่าการเพิ่มตัวแปรเหล่านั้นไม่มีผลต่อค่าวัดของแบบจำลอง อย่างเช่น การหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation analysis)
- การเลือกตัวแปรแบบไปข้างหลัง (Backward feature selection) จะแตกต่างจากวิธีการที่ 1 โดยที่เราสามารถพิตติงชุดข้อมูลทั้งหมดและทำการทำซ้ำในการลดตัวแปรตราบเท่าที่แบบจำลองยังคงเหมือนเดิมไม่เปลี่ยนแปลง

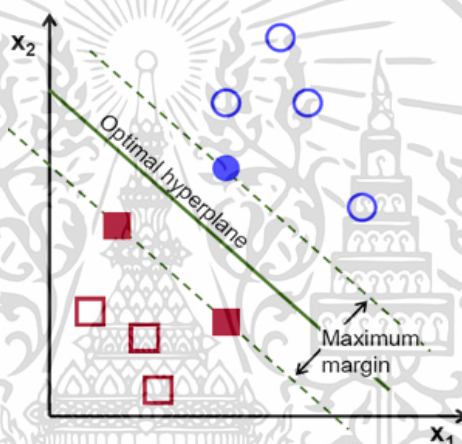
วิธีการที่มักนิยมใช้แบ่งออกเป็น 3 แบบ

1. Filtered-based เป็นวิธีการที่ตรงไปตรงมาสำหรับการเลือกตัวแปรอิสระของอัลกอริทึมของการเรียนรู้ของเครื่องจักรต่างๆ เช่น การใช้ค่าสถิติ อย่าง Pearson's Correlation, LDA เป็นต้น โดยการเลือกตัวแปรที่สำคัญที่มีผลกับเป้าหมายในการ training แบบจำลอง และใช้ระยะเวลาในการประมวลผลที่รวดเร็ว
2. Wrapper เป็นวิธีการเลือกตัวแปรจากผลลัพธ์ค่าวัดของการ training การเรียนรู้ของเครื่องจักร โดยแต่ละส่วนการเรียนรู้จะได้รับคะแนนหลังจาก training และทำการเพิ่มตัวแปรหรือลดตัวแปรจนกระทั่งการเรียนรู้หยุดหรือบรรลุวัตถุประสงค์ที่เราต้องการ วิธีการนี้มักเป็นวิธีการ forward backward หรือ recursive selection แต่ต้องใช้ระยะเวลาในการประมวลผลแบบจำลองการเรียนรู้ของเครื่องจักรจำนวนมาก
3. Embedded เป็นวิธีการที่ซับซ้อนและเป็นการผสมผสานกับวิธี Filtered-base และ wrapper เข้าด้วยกัน อย่างเช่น LASSO และ Tree

2.5 การเรียนรู้ของเครื่องจักร (Machine Learning)

การพยากรณ์ทิศทางการเปลี่ยนแปลงของราคาหุ้นนั้นมีแบบจำลองที่หลากหลายที่ถูกนำมาใช้ ซึ่งวิธีการที่นิยม คือการสร้างแบบจำลองการเรียนรู้ของเครื่องจักรแบบมีผู้สอน โดยใช้ข้อมูลด้านราคาและข่าวที่เกี่ยวข้องเป็นตัวแปรในการสอนแบบจำลอง

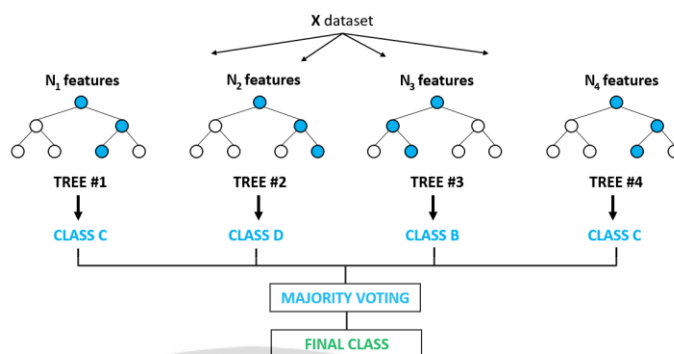
ซัพพอร์ตเวกเตอร์แมทชีน (Support Vector Machine) [25] แบบจำลองนี้ถูกใช้เป็นที่นิยมอย่างมากในการจำแนกประเภท การทำงานของแบบจำลองนี้ใช้วิธีการหาเส้นไฮเปอร์เพลน (Hyperplane) ที่ให้ระยะขอบที่เหมาะสมที่สุดในการแยกประเภททิศทางหรือลงของราคา ซึ่งแบบจำลองนี้ยังช่วยแก้ไขปัญหการเรียนรู้ที่มากเกินไป (Overfitting) ทำให้แบบจำลองถูกนำมาใช้งานได้เป็นอย่างดี



รูปที่ 2.7 แบบจำลองซัพพอร์ตเวกเตอร์แมทชีน

ที่มา : <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>

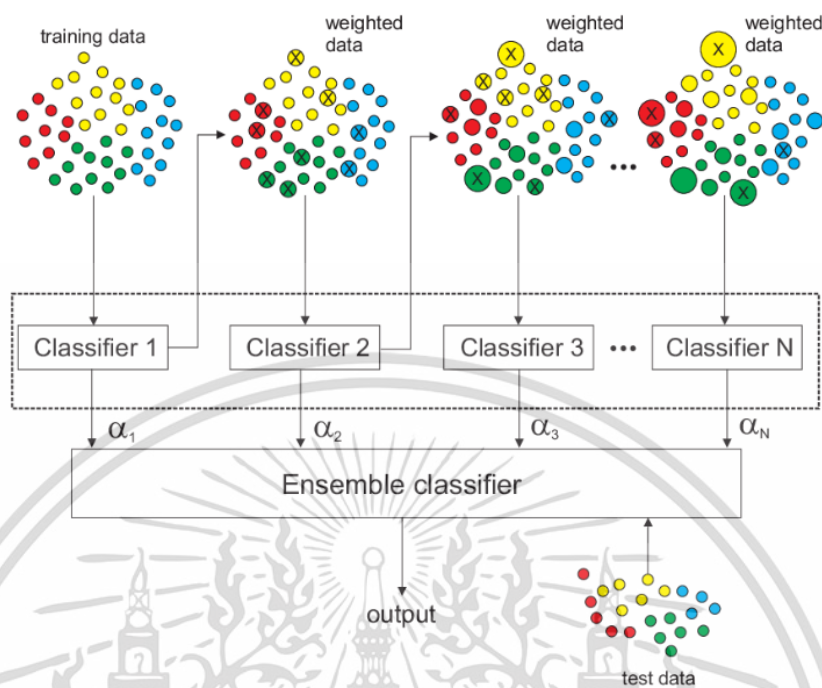
ป่าไม้สุ่ม (Random Forest) [26] ที่ใช้ในการแยกประเภทบนพื้นฐานการทำงานของต้นไม้การตัดสินใจ (Decision Trees) ซึ่งการทำงานของป่าไม้สุ่ม (Random Forest) นั้น เป็นการใช้ต้นไม้การตัดสินใจหลายๆ ต้น ในการพยากรณ์ โดยที่ชุดข้อมูลที่ถูกสอนกับแบบจำลองของต้นไม้การตัดสินใจจะถูกสุ่มเข้าไปในแต่ละต้น ทำให้ผลลัพธ์ของต้นไม้การตัดสินใจนั้นมีค่าที่แตกต่างกัน การพยากรณ์ใช้วิธีการโหวตตามเสียงข้างมากและยังช่วยป้องกันการเรียนรู้ที่มากเกินไปจากการใช้แบบจำลองต้นไม้ตัดสินใจของแต่ละต้นอีกด้วย



รูปที่ 2.8 แบบจำลองป่าไม้ผสม

ที่มา : <https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/>

อดาบู้สต์ (AdaBoost) [27] เป็นอัลกอริทึมในเครื่องมือประมวลผลทางสถิติและการเรียนรู้เชิงเส้นคือการเรียนรู้แบบสอนแบบพิเศษที่ใช้กับการจำแนกหรือคาดการณ์ โดยเฉพาะในกรณีของปัญหาการจำแนก (Classification) ซึ่งมีการเรียนรู้ร่วมกันของตัวแบบหลายตัวเพื่อปรับปรุงประสิทธิภาพของการจำแนก โดยอัลกอริทึมอดาบู้สต์ (Adaboost) ใช้แนวคิดของการสร้างตัวจำแนกแบบอ่อนแอ (Weak Classifier) ให้ตัวแบบอ่อนแอเป็นตัวจำแนกแบบตัดสินใจ (Decision Stump) โดยในการเทรนแต่ละรอบจะมีการกำหนดค่าน้ำหนักของข้อมูลแต่ละรายการของตัวแบบจำแนก (Classifier Instance) จะเรียนรู้จากค่าน้ำหนักเหล่านั้น โดยถ้าพยากรณ์ผิด ค่าน้ำหนักของรายการนั้นจะมีค่ามากขึ้น ส่งผลให้ตัวแบบจำแนก (Classifier Instance) นั้นได้รับคะแนนต่ำ แต่ในทางกลับกัน ถ้าตัวแบบจำแนกไหนพยากรณ์ถูกเป็นสัดส่วนที่มากก็จะได้คะแนนมาก การพยากรณ์ของอดาบู้สต์ (AdaBoost) คือการถ่วงคะแนนเข้ากับคำตอบของแต่ละตัวแบบจำแนก แล้วเลือกคำตอบจากผลคำตอบที่ได้รวมแล้วได้ค่าน้ำหนักมากที่สุด

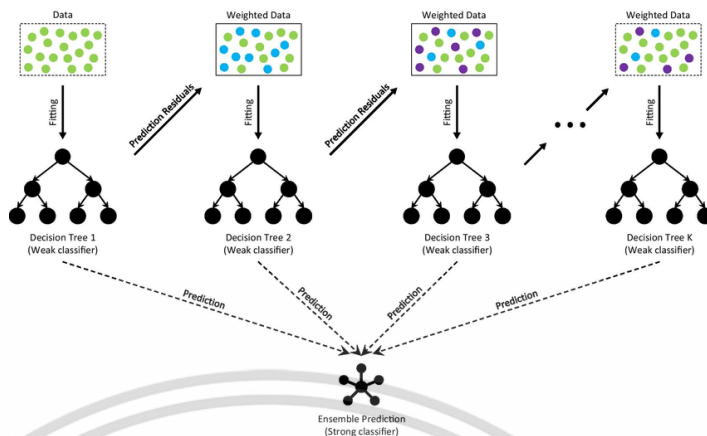


รูปที่ 2.9 แบบจำลองอดาบู้สต์

ที่มา : <https://www.researchgate.net/profile/Jorge-Perez-Aracil/publication/362065998/figure/fig2/AS:1179038517342211@1658116096852/Diagram-of-the-AdaBoost-algorithm-exemplified-for-multi-class-classification-problems.png>

Aracil/publication/362065998/figure/fig2/AS:1179038517342211@1658116096852/Diagram-of-the-AdaBoost-algorithm-exemplified-for-multi-class-classification-problems.png

เกรเดียนท์บูสต์ติง (Gradient Boosting) [27] เป็นแบบจำลองที่ถูกพัฒนาจากอัลกอริทึมบูสต์ (Boosted Machine Learning) ที่มีการประมวลผลในการคำนวณรวดเร็วยิ่งขึ้น จึงนิยมนำมาใช้เพื่อพัฒนาประสิทธิภาพของแบบจำลองเป็นอย่างมาก และยังลดโอกาสที่จะเกิดการเรียนรู้ที่มากเกินไป โดยมีการทำงานจากการบูสต์ต้นไม้ (Boosted Tree) ที่ฝึกฝนข้อมูลบนแบบจำลองตามลำดับทำให้การคำนวณต้นไม้ใหม่นั้นมีการถ่วงน้ำหนักความผิดพลาดจากต้นไม้ก่อนหน้าด้วยจึงทำให้ค่าสูญเสียมีค่าต่ำ โดยแบบจำลองจะทำการปรับตัวแบบจำแนก (Classifier Instance) ที่มาใหม่แต่ละตัว ให้มีความแม่นยำขึ้นเรื่อยๆ โดยเรียนรู้จากผลต่างค่าจริงกับค่าที่พยากรณ์ (Residual) สะสมที่เกิดจากการทำนายของตัวแบบจำแนกก่อนหน้า การพยากรณ์ของแบบจำลองคำนวณจากการบวกค่าสะสมของผลต่างค่าจริงกับค่าที่พยากรณ์ แบบจำลองนี้ยังสามารถจัดอันดับความสำคัญของตัวแปร (Feature Importance) ที่มีผลต่อการจำแนกประเภทข้อมูลได้อีกด้วย

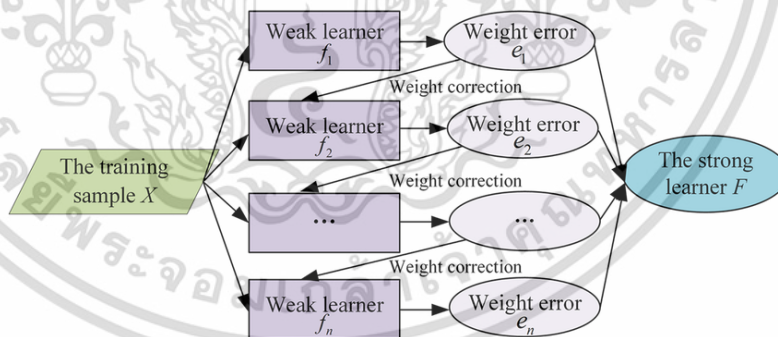


รูปที่ 2.10 แบบจำลองเกรเดียนท์บู้สต์ตั้ง

ที่มา :

<https://www.researchgate.net/publication/356698772/figure/fig2/AS:1096436418641951@163842221975/The-architecture-of-Gradient-Boosting-Decision-Tree.png>

แคตบู้สต์ (CatBoost) [28] เป็นการพัฒนาจากวิธีเกรเดียนท์บู้สต์ตั้ง (Gradient Boosting) ที่ถูกพัฒนาเพื่อจัดการกับข้อมูลจำพวกตัวแปรประเภท (Categorical data) และช่วยเพิ่มประสิทธิภาพการทำงานของแบบจำลองให้สามารถทำงานได้เร็วขึ้น



รูปที่ 2.11 แบบจำลองแคตบู้สต์

ที่มา :

<https://www.researchgate.net/publication/367098144/figure/fig1/AS:11431281153674670@1682520318800/Diagram-of-CatBoost-structure.tif>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

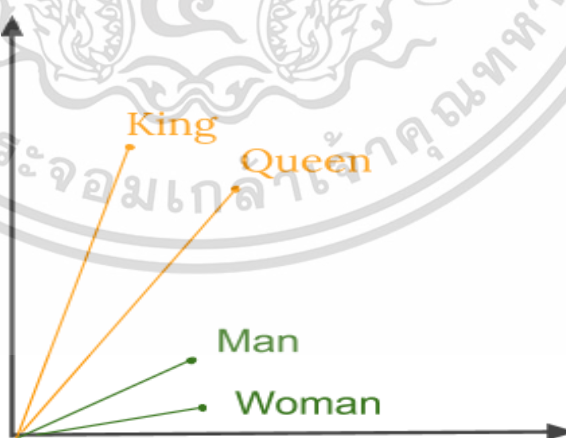
2.6 การประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP)

การประมวลผลภาษาธรรมชาติ เป็นวิทยาการแขนงหนึ่งในหมวดหมู่ของเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) ซึ่งช่วยให้คอมพิวเตอร์สามารถเข้าใจ ตลอดจนตีความและใช้งานภาษาปกติที่มนุษย์ใช้สื่อสารได้

กระบวนการทำงานของการประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาตินั้นประกอบด้วยหลากหลายวิธีการประมวลผลและแปลความหมายของภาษาปกติของมนุษย์ เช่น ระเบียบวิธีทางสถิติและการเรียนรู้ของเครื่องที่หลากหลาย ไปจนถึงกระบวนการทำงานตามขั้นตอน กฎเกณฑ์ และอัลกอริทึมที่ซับซ้อน ซึ่งเทคนิคในการรับมือข้อมูลเหล่านี้จำเป็นต่อการทำงาน เนื่องจากข้อมูลในรูปแบบข้อความ คำ และเสียงพูดนั้น มีความแตกต่างและหลากหลายค่อนข้างมาก รวมถึงสามารถใช้งานได้ในหลายรูปแบบเช่นกัน

การแปลงคำเป็นเวกเตอร์ (Word Embedding) [29] เป็นเทคนิคที่ใช้ในการแปลงคำหรือวลีในภาษาธรรมชาติเป็นเวกเตอร์หรือตัวเลข โดยที่ตัวเลขหนึ่งชุดสามารถแทนที่คำหนึ่งคำได้ วิธีการนี้ช่วยให้เครื่องคอมพิวเตอร์สามารถเข้าใจความหมายและความสัมพันธ์ระหว่างคำหรือวลีได้ ใน Word Embedding ที่ได้รับความนิยมมากคือ Word2Vec GloVe Thai2Vec และ FastText ซึ่งเป็นแบบจำลองที่สร้างเวกเตอร์คำจากข้อมูลที่ใช้ฝึก (Pre-Trained Model) วิธีการแปลงคำเป็นเวกเตอร์ (Word Embedding) เป็นหนึ่งในรูปแบบการแสดงคำที่เชื่อมต่อภาษาของมนุษย์กับเครื่องจักรได้ โดยคำที่มีความหมายเดียวกันเวกเตอร์จะมีค่าใกล้เคียงกัน ซึ่งเป็นสิ่งสำคัญอย่างมากในการแก้ไขปัญหาด้านการประมวลผลภาษาธรรมชาติ (Natural Language Processing)



รูปที่ 2.12 คำที่คล้ายกันจะอยู่ในเวกเตอร์ที่ใกล้กัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ที่มา : https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcQ03MAFBpl_RDzDHAXZtrplzD8vB80mXdHxAA&usqp=CAU

ความคล้ายคลึงของข้อความ (Text Similarity) [30] เป็นการวัดความคล้ายคลึงระหว่างข้อความสองข้อความ ซึ่งอาจใช้ในการเปรียบเทียบเนื้อหา ข้อความ เช่น การหาคำที่คล้ายคลึงในเอกสาร การค้นหาเอกสารที่มีเนื้อหาคคล้ายคลึงกับข้อความที่กำหนด หรือการค้นหาข่าวที่มีเนื้อหาคคล้ายคลึงกับข่าวที่สนใจ เพื่อให้ง่ายต่อการค้นหาและจัดเรียงข้อมูลในแบบที่มีความเกี่ยวข้องกัน

ตัวชี้วัดในการคำนวณความคล้ายคลึงระหว่าง word embeddings [30]

Cosine Similarity เป็นการวัดความคล้ายคลึงระหว่างเวกเตอร์สองเวกเตอร์โดยใช้ค่าคอซายของมุมระหว่างเวกเตอร์เพื่อกำหนดค่าความคล้ายคลึงระหว่างคู่อันดับของเวกเตอร์นั้น ๆ การวัดความคล้ายคลึงด้วย Cosine Similarity คำนวณจากเวกเตอร์สองเวกเตอร์โดยใช้สูตรคอซายของมุมระหว่างเวกเตอร์ ซึ่งสามารถแสดงให้เห็นถึงมุมของความคล้ายคลึงระหว่างเวกเตอร์ได้ว่าการชี้ไปในทิศทางเดียวกันหรือไม่ ค่าคล้ายคลึงตามหลัก Cosine Similarity อยู่ในช่วง -1 ถึง 1 โดยที่ค่า 1 แสดงถึงความคล้ายคลึงมากที่สุด ค่า 0 แสดงถึงความคล้ายคลึงน้อยที่สุด และค่า -1 แสดงถึงความแตกต่างและคล้ายคลึงกันที่น้อยที่สุด

$$\begin{aligned} \text{similarity}(A, B) &= \frac{A \cdot B}{\|A\| \|B\|} \\ &= \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}} \end{aligned}$$

A และ B คือ เวกเตอร์ของคำหรือประโยคที่ต้องการเปรียบเทียบ

$\|A\|$ และ $\|B\|$ คือ ค่าความยาวของเวกเตอร์

Jaccard Similarity หรือดัชนี Jaccard คือ เป็นหนึ่งในวิธีการวัดความคล้ายคลึงระหว่างประโยคหรือข้อความมักใช้กันอย่างแพร่หลายในการประมวลผลภาษาธรรมชาติ (Natural Language Processing) หรือการทำงานกับข้อความ ค่าคล้ายคลึงตามหลัก Jaccard คืออัตราส่วนของจำนวนคำที่เหมือนกันระหว่างสองประโยคหรือข้อความโดยหารด้วยผลรวมของจำนวนคำทั้งสอง คะแนนของ Jaccard Similarity จะอยู่ในช่วง 0 ถึง 1 โดยที่ 1 แทนความคล้ายคลึงมากที่สุดและ 0 แทนความคล้ายคลึงน้อยที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$J(doc_1, doc_2) = \frac{doc_1 \cap doc_2}{doc_1 \cup doc_2}$$

$doc_1 \cap doc_2$ คือ จำนวนคำที่ซ้ำกันระหว่างเอกสาร1 และเอกสาร2

$doc_1 \cup doc_2$ คือ จำนวนคำทั้งหมดของเอกสาร1 และเอกสาร2

Euclidean Distance เป็นสูตรที่ใช้กันอย่างแพร่หลายในการคำนวณระยะห่างระหว่างจุดพิกัดสองจุด โดยการลบจุดเวกเตอร์กัน ยกกำลังสองของผลลัพธ์ รวมผลลัพธ์และนำผลรวมไปหารากที่สองของผลลัพธ์ เช่นเดียวกันกับการใช้ระยะทางกับ word embedding โดยคำนวณระยะห่างระหว่างเวกเตอร์ของ embedding ของคำ โดยสมมติว่าแต่ละเวกเตอร์แทนชุดของพิกัด (coordinate)

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

$(q_i - p_i)$ คือ ความแตกต่างของพิกัด q_i และ p_i ระหว่างจุด p และจุด q ตามลำดับ

2.7 การประเมินประสิทธิภาพ (Model Evaluation)

งานวิจัยการทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและการวิเคราะห์เชิงเทคนิคเป็นการพยากรณ์ในแบบการจำแนกประเภท (Classification) ราคาขึ้น ราคาลง ราคาไม่มีแนวโน้ม โดยจะใช้เมตริกซ์ความสับสนในการประเมินประสิทธิภาพของแบบจำลอง ดังตารางที่ 2.2

ตารางที่ 2.2 เมทริกซ์ความสับสน (Confusion Matrix) ขนาด 3x3

		ค่าที่แบบจำลองทำนาย Predicted Class		
		ราคาลง Class -1	ราคาไม่มีแนวโน้ม Class 0	ราคาขึ้น Class 1
ค่าที่เกิดขึ้นจริง Actual Class	ราคาลง Class -1	TP	FN	FN
	ราคาไม่มีแนวโน้ม Class 0	FP	TN	FN
	ราคาขึ้น Class 1	FP	FN	TN

True Positive (TP) สิ่งที่แบบจำลองทำนายตรงกับสิ่งที่เกิดขึ้นจริง ทำนายว่า “จริง” และสิ่งที่เกิดขึ้น ก็คือ “จริง”

True Negative (TN) สิ่งที่แบบจำลองทำนายตรงกับสิ่งที่เกิดขึ้น ทำนายว่า “ไม่จริง” และสิ่งที่เกิดขึ้น ก็คือ “ไม่จริง”

False Positive (FP) สิ่งที่แบบจำลองทำนายไม่ตรงกับสิ่งที่เกิดขึ้น ทำนายว่า “จริง” แต่สิ่งที่เกิดขึ้น คือ “ไม่จริง”

False Negative (FN) สิ่งที่แบบจำลองทำนายไม่ตรงกับที่ที่เกิดขึ้นจริง ทำนายว่า “ไม่จริง” แต่สิ่งที่เกิดขึ้น คือ “จริง”

โดยมีค่าวัดประสิทธิภาพต่างๆ ดังต่อไปนี้ [31]

- ค่าความแม่นยำ (Accuracy) เป็นอัตราส่วนของจำนวนข้อมูลที่ถูกต้องทั้งหมดต่อจำนวนข้อมูลทั้งหมดในชุดข้อมูล ค่าความแม่นยำเป็นตัววัดที่ใช้บ่งบอกถึงความถูกต้องของการจำแนก

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- ค่าความเที่ยงตรง (Precision) เป็นอัตราส่วนของจำนวนข้อมูลที่ถูกตัวจำแนกทำนายว่าเป็น positive และตรงตามความจริง ต่อจำนวนข้อมูลที่ตัวจำแนกทำนายว่าเป็น positive

$$Precision = \frac{TP}{TP + FP}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ค่าความครบถ้วน (Recall) เป็นอัตราส่วนของจำนวนข้อมูลที่ถูกตัวจำแนกทำนายว่าเป็น positive และตรงตามความจริง ต่อจำนวนข้อมูลที่เป็น positive ในชุดข้อมูลทั้งหมด

$$Recall = \frac{TP}{TP + FN}$$

- ค่าความถ่วงดุล (F1-Score) เป็นค่าเฉลี่ยของค่าความเที่ยงตรงและค่าความเรียกคืน เป็นตัววัดที่ใช้ประเมินความสมดุลระหว่างค่าความเที่ยงตรงและค่าความเรียกคืน ค่า F1-Score สูงสุดเมื่อมีความถ่วงดุลระหว่างค่าความเที่ยงตรงและค่าความเรียกคืนที่ดีที่สุด

$$F1 - score = 2 \times \left[\frac{Precision \times Recall}{Precision + Recall} \right]$$

- ค่าเฉลี่ยมาโคร (Macro Avg) เป็นการเฉลี่ยค่าความเที่ยงตรง ค่าเรียกคืน และค่าความถ่วงดุลของทุกกลุ่มโดยเท่าเทียมกัน ซึ่งแสดงถึงความแม่นยำและครอบคลุมของแบบจำลองที่ใช้ในการจำแนกกลุ่มทั้งหมด ค่าที่ได้จะเป็นตัวบ่งชี้รวมทั้งหมดสำหรับความสามารถของแบบจำลองในการจำแนกกลุ่มทั้งหมดที่ได้รับการทดสอบ
- ค่าเฉลี่ยถ่วงน้ำหนัก (Weighted Avg) เป็นการเฉลี่ยค่าความเที่ยงตรง ค่าเรียกคืน และค่าความถ่วงดุลโดยให้น้ำหนักขึ้นอยู่กับจำนวนข้อมูลในแต่ละกลุ่ม ค่าที่ได้จะเป็นตัวบ่งชี้ความสามารถของแบบจำลองในการจำแนกกลุ่มโดยพิจารณาและให้ความสำคัญกับแต่ละกลุ่มตามจำนวนข้อมูลที่มีในแต่ละกลุ่ม

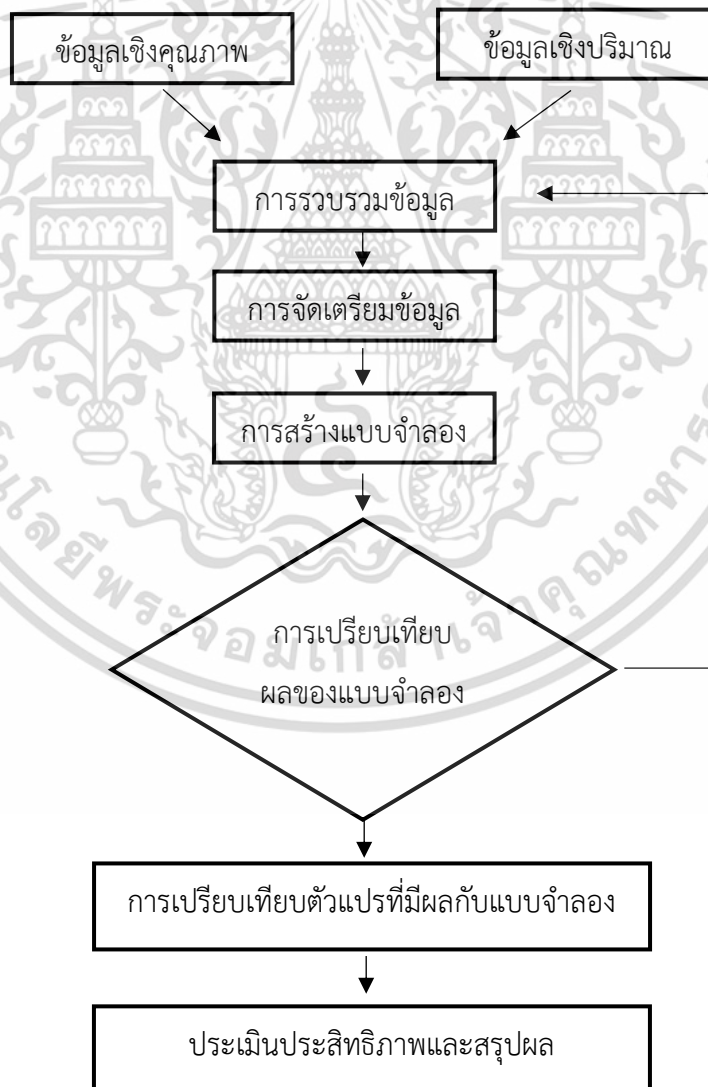
บทที่ 3

วิธีการดำเนินงานวิจัย

งานวิจัยการทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและการวิเคราะห์เชิงเทคนิค ผู้วิจัยได้มีการศึกษาค้นคว้างานวิจัยที่เกี่ยวข้องมาปรับใช้กับงานวิจัย โดยมีขั้นตอนการดำเนินงานวิจัยดังนี้

3.1 ขั้นตอนการดำเนินงาน

การทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและการวิเคราะห์เชิงเทคนิค มีลำดับขั้นตอนการดำเนินงาน ดังรูปที่ 3.1



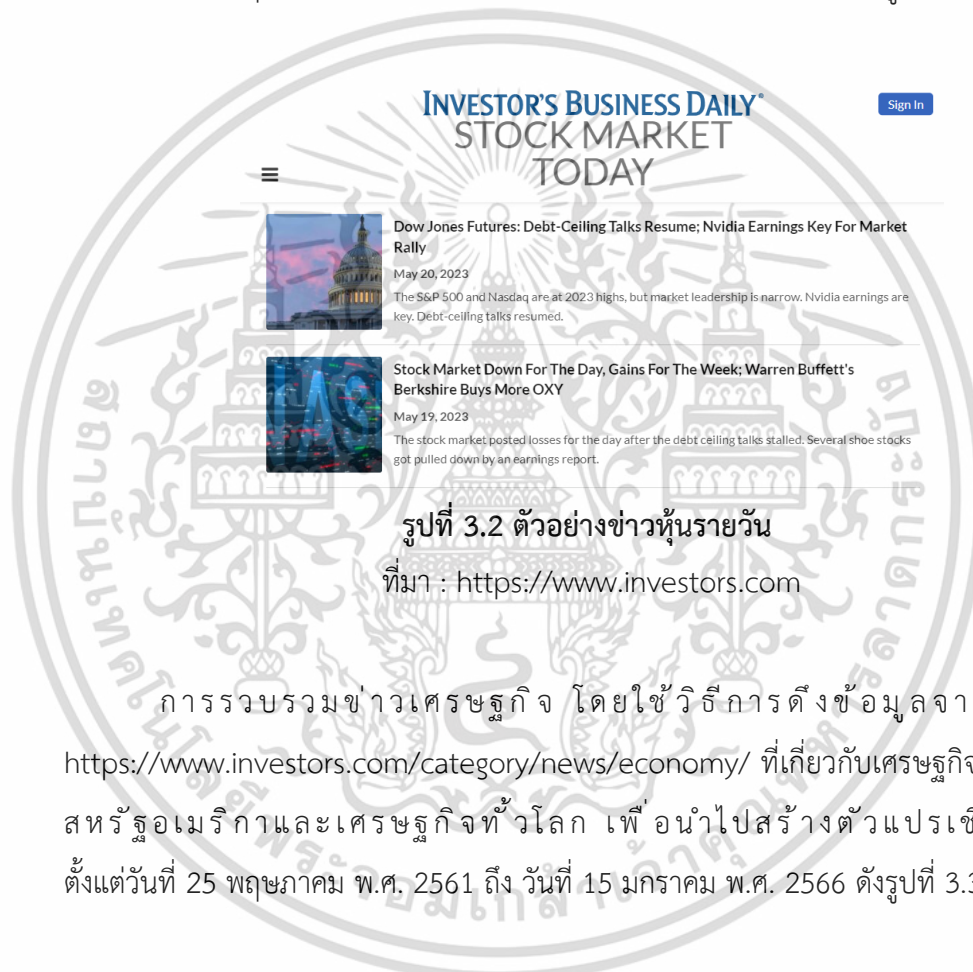
รูปที่ 3.1 ขั้นตอนการดำเนินงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 การรวบรวมข้อมูล

3.2.1 ปัจจัยข่าว

การรวบรวมข่าวที่เกี่ยวข้องกับหุ้นรายวัน โดยใช้วิธีการดึงข้อมูล (Web Scraping) จากเว็บไซต์ <https://www.investors.com/category/stock-market-today/> ที่เกี่ยวกับข่าวหุ้นรายวันของประเทศสหรัฐอเมริกา เพื่อนำไปสร้างตัวแปรเชิงคุณภาพ ตั้งแต่วันที่ 25 พฤษภาคม พ.ศ. 2561 ถึง วันที่ 15 มกราคม พ.ศ. 2566 ดังรูปที่ 3.2



รูปที่ 3.2 ตัวอย่างข่าวหุ้นรายวัน

ที่มา : <https://www.investors.com>

การรวบรวมข่าวเศรษฐกิจ โดยใช้วิธีการดึงข้อมูลจากเว็บไซต์ <https://www.investors.com/category/news/economy/> ที่เกี่ยวกับเศรษฐกิจของประเทศสหรัฐอเมริกาและเศรษฐกิจทั่วโลก เพื่อนำไปสร้างตัวแปรเชิงคุณภาพ ตั้งแต่วันที่ 25 พฤษภาคม พ.ศ. 2561 ถึง วันที่ 15 มกราคม พ.ศ. 2566 ดังรูปที่ 3.3

INVESTOR'S BUSINESS DAILY[®]
ECONOMY

New to Investors.com? [Start here!](#)

CPI Inflation Rate Falls Below 5% As Service Prices Cool; S&P 500 Fizzles
May 10, 2023
The CPI inflation rate eased in April, as did core inflation as service prices, the Fed's big worry, showed signs of cooling.

Economic Optimism Index Dives As U.S. Recession Clouds Gather
May 09, 2023
The IBD/TIPP Economic Optimism Index slid to a six-month low, as rising layoffs and high interest rates raise the risk of a U.S. recession.

รูปที่ 3.3 ตัวอย่างข่าวเศรษฐกิจ
ที่มา : <https://www.investors.com>

3.2.2 ปัจจัยเชิงเทคนิค

สำหรับการรวบรวมปัจจัยเชิงเทคนิคงานวิจัยนี้ใช้ไลบรารีของ yfinance เป็นแพ็คเกจของไพธอนที่ให้ความสะดวกในการดึงข้อมูลเพื่อรวบรวมข้อมูลของหุ้นย้อนหลังจาก Yahoo Finance API เช่น ราคาหุ้น ราคาปิด ราคาเปิด ปริมาณการซื้อขาย และตัวชี้วัดการเงินทางการเงินอื่นๆ ผู้วิจัยใช้ไลบรารี yfinance รวบรวมราคาดัชนีต่างประเทศ ราคาดัชนีในประเทศ และราคาหุ้นในประเทศรายวัน Dow Jones, Nasdaq, S&P500, SET, SET50 และราคาดัชนีอุตสาหกรรมในประเทศ (Sector Index) ตั้งแต่เดือน พฤษภาคม พ.ศ.2561 ถึงเดือน มกราคม พ.ศ.2566 ดังรูปที่ 3.4 3.5 และ 3.6 ตามลำดับ

Date	DJI_Close	NDX_Close	SPX_Close	SET_Close	DJI_Volume	NDX_Volume	SPX_Volume	SET_Volume
2023-01-13	34302.61	11541.48	3999.09	1681.73	274402048.0	8.681806e+08	2.305645e+09	5.762924e+09
2023-01-12	34189.97	11459.61	3983.17	1687.45	308605840.0	9.844340e+08	2.468086e+09	4.934908e+09
2023-01-11	33973.01	11402.52	3969.61	1685.75	302297828.0	9.564935e+08	2.353913e+09	4.778605e+09
2023-01-10	33704.10	11205.78	3919.25	1691.41	269943626.0	8.478130e+08	2.140006e+09	5.302956e+09
2023-01-09	33517.65	11108.45	3892.09	1691.12	328602398.0	1.011171e+09	2.498159e+09	4.675454e+09
...
2019-01-02	23346.24	6360.87	2510.03	1565.94	321542416.0	3.899655e+08	2.073978e+09	1.968622e+09
2018-12-31	23327.46	6329.96	2506.85	NaN	288800253.0	3.618207e+08	1.912706e+09	NaN
2018-12-28	23062.40	6285.27	2485.74	1563.88	336565447.0	3.791534e+08	2.057011e+09	2.231844e+09

รูปที่ 3.4 ตัวอย่างราคาดัชนี (รายวัน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Date	ADVANC		AOT		AWC		BANPU		BBL		BDMS		BEM	
	Close	Volume	Close	Volume	Close	Volume	Close	Volume	Close	Volume	Close	Volume	Close	Volume
2018-05-21	193.0	4067100	69.50	22197000	NaN	NaN	21.400000	41314000	193.5	3604300	26.25	39578200	7.90	19976500
2018-05-22	194.5	5502800	69.00	19517000	NaN	NaN	21.299999	43015700	194.0	5752100	26.25	45810500	7.95	17854800
2018-05-23	195.0	2744400	68.75	11752000	NaN	NaN	21.400000	59401600	193.5	3542600	26.25	59766900	7.85	17962500
2018-05-24	193.5	2705600	69.00	11994400	NaN	NaN	20.299999	134702600	192.0	2650700	26.75	58399800	8.00	46148900
2018-05-25	194.5	2870900	69.00	31610000	NaN	NaN	20.500000	48665600	193.5	2746300	27.00	59588000	8.00	24661600
...
2023-05-15	211.0	12483000	73.25	36085500	5.45	21157200.0	8.150000	87011600	162.0	11210900	29.50	18323700	8.65	42073800
2023-05-16	213.0	6325900	72.00	28372800	5.45	13320800.0	8.400000	97515800	161.0	5361900	29.50	26460200	8.65	30917600
2023-05-17	215.0	4680900	72.00	37799300	5.35	40302000.0	8.100000	95493800	159.0	5544700	29.25	37664400	8.65	39690100
2023-05-18	211.0	3567800	71.50	21507100	5.25	11403400.0	8.050000	59489600	161.5	6462600	28.75	53419400	8.35	58383100
2023-05-19	210.0	2926800	70.25	26330400	5.15	39135300.0	8.050000	93458700	160.0	4445600	28.00	49335200	8.30	36205900

รูปที่ 3.5 ตัวอย่างราคาหุ้น SET50 (รายวัน)

	AGRI.Close.T.	AUTO.Close.T.	BANK.Close.T.	COMM.Close.T.	CONMAT.Close.T.	CONS.Close.T.	ENERG.Close.T.	ETRON.Close.T.	FASHION.Close.T.	FIN.Close.T.	FOOD.Close.T.	HEALTH.Close.T.
0	198.79	579.61	514.83	44954.29	11771.03	102.22	25902.37	1594.69	730.89	3449.91	12034.38	6058.69
1	179.86	467.48	502.32	38749.67	10611.64	83.67	22755.13	1481.83	726.11	3412.09	10609.69	5580.59
2	197.10	494.61	530.20	42095.05	11439.86	91.05	24548.32	1524.47	714.67	3523.81	11730.35	5535.77
3	202.62	500.33	529.17	41888.68	11567.96	89.72	24651.26	1579.41	707.48	3561.24	11906.97	5523.77
4	185.44	485.87	514.88	40732.26	11563.32	91.23	24646.24	1484.94	706.89	3614.51	11776.13	5653.08
...
948	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
949	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
950	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
951	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
952	241.61	539.02	473.23	35472.92	9558.61	71.02	24787.75	1096.797	727.93	4753.54	12047.83	7407.35

รูปที่ 3.6 ตัวอย่างราคาดัชนีอุตสาหกรรมในประเทศ (รายวัน)

3.3 การจัดเตรียมข้อมูล

3.3.1 ปัจจัยข่าว

จากบทความข่าวที่ผู้วิจัยทำการรวบรวมเป็นข้อมูลประเภทเชิงคุณภาพ รูปที่ 3.7 และ 3.8 ตามลำดับ ดังนั้นจะทำการแปลงบทความข่าวเหล่านั้นให้อยู่ลักษณะข้อมูลเชิงปริมาณ เนื่องจากผู้วิจัยต้องการทราบว่าปัจจัยข่าวแบบไหนมีผลต่อการเปลี่ยนแปลงของราคาหุ้น ผู้วิจัยทำการสกัดบทความข่าว โดยใช้วิธีการความคล้ายคลึงกันของข้อความ (Text Similarity)

	Topic	Details	StrDate	Date	Source
0	CPI Inflation Rate Slides As Service Prices Sh...	The CPI inflation rate continued to fall sharp...	January 12, 2023	2023-01-12	Global Economic
1	Cash at cash register Tamer CPI Inflation Rate...	The CPI inflation rate should continue its rap...	January 11, 2023	2023-01-11	Global Economic
2	U.S. Economic Optimism Dims; Stock Market Inve...	Americans' view of the outlook for the U.S. ec...	January 10, 2023	2023-01-10	Global Economic
3	IBD TIPP poll IBD/TIPP Poll: Tracking The U.S....	The IBD/TIPP Economic Optimism Index gives an ...	January 10, 2023	2023-01-10	Global Economic
4	Jobs Report: Cooler Wage Growth, Weak ISM Are ...	Friday's jobs report showed solid hiring, but ...	January 06, 2023	2023-01-06	Global Economic

รูปที่ 3.7 ตัวอย่างข่าวเศรษฐกิจ

	Topic	Details	StrDate	Date	Source
0	Tesla CEO Elon Musk Market Rally Breaks Throug...	The market rally broke above key resistance th...	January 13, 2023	2023-01-13	Stock News
1	Warren Buffett Dow Jones Gains; 3 Warren Buffe...	The Dow Jones reversed higher. Tesla stock was...	January 13, 2023	2023-01-13	Stock News
2	stock market: JPMorgan Stock Market Steady In ...	Tesla was one of the worst-performing major st...	January 13, 2023	2023-01-13	Stock News
3	Dow Jones: Tesla Dow Jones Falls 250 Points As...	The Dow Jones Industrial Average fell 250 poin...	January 13, 2023	2023-01-13	Stock News
4	Tesla Model 3 Dow Jones Futures Fall: JPMorgan...	UnitedHealth earnings topped, with JPMorgan an...	January 13, 2023	2023-01-13	Stock News

รูปที่ 3.8 ตัวอย่างข่าวหุ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้วิจัยทำการแบ่งประเภทของข่าวออกเป็น 7 ประเภท ที่มีผลกับการเคลื่อนไหวของตลาด พร้อมทั้งระบุคำสำคัญ (keyword) ของข่าวแต่ละประเภท โดยผู้วิจัยอ้างอิงจากประสบการณ์ที่อยู่ในตลาด การเรียนจากนักลงทุน อาทิ คุณกระทรวง จารุศิริระ การอ่านหนังสือการลงทุน และการเรียนรู้จากการถ่ายทอดความรู้หรือประสบการณ์ของนักลงทุนผ่านสื่อต่างๆ เพื่อให้สามารถจับคู่คำสำคัญกับบทความข่าวว่าอัตราส่วนเท่าไรในการระบุประเภทของข่าวนั้น

1. ข่าวทางเศรษฐกิจ (Economic)
2. ข่าวทางการเมือง (Political Events)
3. เหตุการณ์ทางบริษัท (Company Events)
4. อารมณ์นักลงทุน (Investor Sentiment)
5. การเคลื่อนไหวของตลาด (Market Moment)
6. ภัยธรรมชาติ (Natural Disasters)
7. เหตุการณ์เฉพาะกลุ่มอุตสาหกรรม (Industry Specific Events)

เนื่องจากคำสำคัญที่ปรากฏในข่าวแต่ละประเภทค่อนข้างมีคำที่หลากหลายจึงยากในระบุคำสำคัญให้ครอบคลุมในข่าวแต่ละประเภท ผู้วิจัยจึงทำการใช้แบบจำลองพาสท์เท็กซ์ (FastText) เป็นแบบจำลองที่ถูกฝึกสอนแล้วกับข้อมูลข้อความขนาดใหญ่เป็นเครื่องมือในการประมวลผลภาษาธรรมชาติ แบบจำลองนี้สามารถตรวจสอบและวิเคราะห์ประโยคที่ให้มาและสามารถจัดหมวดหมู่คำหรือประโยคได้ตามที่ได้รับการฝึกสอน ช่วยให้แบบจำลองเข้าใจความหมายและความสัมพันธ์ระหว่างคำได้ดีขึ้น ดังรูปที่ 3.9

Topic	Details	StrDate	Date	Source	Cleaned_Topic	Economic	Political events	Company events	Investor sentiment	Natural disasters	Industry specific events	Market moment
warren buffett dow jones gains, a warren buffe...	the dow jones reversed higher, tesla stock, tes...	January 13, 2023	2023-01-13	Stock News	warren buffett dow jones gains, a warren buffe...	0.127610	0.171086	0.165763	0.183416	0.084409	0.182470	0.085891
tesla ceo elon musk market rally breaks throug...	the market rally broke above key resistance th...	January 13, 2023	2023-01-13	Stock News	tesla ceo elon musk market rally breaks throug...	0.139674	0.200522	0.184034	0.221058	0.106426	0.236849	0.104274
stock market jpmorgan stock market steady in ...	tesla was one of the worst-performing major st...	January 13, 2023	2023-01-13	Stock News	stock market jpmorgan stock market steady in ...	0.153345	0.178317	0.209959	0.230006	0.097327	0.243097	0.136209
dow jones, tesla dow jones falls 250 points as ...	the dow jones industrial average fell 250 point...	January 13, 2023	2023-01-13	Stock News	dow jones, tesla dow jones falls 250 points as ...	0.123739	0.159349	0.153086	0.152375	0.093856	0.170294	0.054202
tesla model 3 dow jones futures fall, jpmorgan an...	unitedhealth earnings topped, with jpmorgan an...	January 13, 2023	2023-01-13	Stock News	tesla model 3 dow jones futures fall, jpmorgan an...	0.154248	0.174047	0.169011	0.176259	0.095267	0.229478	0.070621
stock futures check premarket prices on dow f...	see where the stock market may be headed by ch...	January 13, 2023	2023-01-13	Stock News	stock futures check premarket prices on dow f...	0.120499	0.153312	0.147468	0.202324	0.070797	0.182338	0.067024
wall street sign u.s. stock market news	track current u.s. stock market news, and get ...	January 13, 2023	2023-01-13	Stock News	wall street sign u.s. stock market news	0.148762	0.189162	0.178628	0.171401	0.095070	0.259147	0.180956
tesla ceo elon musk dow jones rallies, tesla g...	the dow jones rallied amid inflation data, tes...	January 12, 2023	2023-01-12	Stock News	tesla ceo elon musk dow jones rallies, tesla g...	0.124767	0.171642	0.142837	0.204865	0.095753	0.190015	0.064008
aon insurance stock market adds to gains, smol...	the stock market gained momentum after turning...	January 12, 2023	2023-01-12	Stock News	aon insurance stock market adds to gains, smol...	0.149020	0.187123	0.202772	0.205690	0.090084	0.222324	0.099351
dow jones falls 150 points on cpi inflation re...	the dow jones industrial average fell 150 poin...	January 12, 2023	2023-01-12	Stock News	dow jones falls 150 points on cpi inflation re...	0.164922	0.175581	0.176489	0.164575	0.082802	0.201729	0.065691

รูปที่ 3.9 ตัวอย่างการวิเคราะห์ความคล้ายคลึงกันของข้อความ (Text Similarity)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากบทความข่าวในแต่ละวันนั้นมีจำนวนมาก ดังรูปที่ 3.10 ผู้วิจัยจึงทำการรวมคะแนนความคล้ายคลึงของคำของแต่ละประเภทให้เป็นผลคะแนนของข่าวในวันเดียว เพื่อให้ทราบว่าข่าวที่ถูกกล่าวถึงในวันนั้นส่วนใหญ่เป็นข่าวประเภทไหน ดังรูปที่ 3.11

Date	Economic	Political events	Company events	Investor sentiment	Natural disasters	Industry specific events	Market moment
2023-01-13	0.127610	0.171086	0.165763	0.183416	0.084409	0.182470	0.085891
2023-01-13	0.139671	0.200522	0.184034	0.221065	0.106426	0.236849	0.104274
2023-01-13	0.153345	0.178317	0.209959	0.230006	0.097327	0.243097	0.136209
2023-01-13	0.129739	0.159349	0.153086	0.152375	0.093856	0.170294	0.054202
2023-01-13	0.154248	0.174047	0.169011	0.176259	0.095267	0.229478	0.070621
...
2018-12-31	0.142645	0.195196	0.177703	0.253974	0.091024	0.186843	0.113446
2018-12-28	0.156834	0.207675	0.207445	0.230667	0.098175	0.209320	0.129032
2018-05-25	0.118955	0.233765	0.157811	0.193193	0.076433	0.215709	0.052013
2018-05-14	0.144080	0.216337	0.195884	0.238472	0.104887	0.219761	0.118849
2018-05-14	0.153157	0.210078	0.201144	0.226721	0.114342	0.222594	0.121869

รูปที่ 3.10 ตัวอย่างคะแนนความคล้ายคลึงกันของข้อความ (Text Similarity) ในแต่ละวัน

Date	Economic	Political events	Investor sentiment	Company events	Market moment	Natural disasters	Industry specific events
2023-01-13	0.973873	1.225795	1.336847	1.207949	0.699178	0.643153	1.503672
2023-01-12	0.807909	0.934386	0.993369	0.897060	0.438292	0.451024	1.039092
2023-01-11	0.874450	1.045317	1.116903	0.980996	0.545536	0.499788	1.086836
2023-01-10	1.007699	1.277212	1.255766	1.210920	0.749058	0.585403	1.387820
2023-01-09	0.623059	0.771440	0.887906	0.799350	0.490079	0.421207	0.855013

รูปที่ 3.11 ตัวอย่างการรวมคะแนนความคล้ายคลึงกันของคำ (Text Similarity) ในแต่ละวัน

ผู้วิจัยทำการปรับข้อมูลให้เป็นค่ามาตรฐานด้วยวิธีการ Min-Max Normalization ของคะแนนการวิเคราะห์ความคล้ายคลึงกันของข้อความใน 1 วัน ให้สัดส่วนของระดับคะแนนอยู่ในช่วง 0-1 เพื่อระบุว่ามีอัตราส่วนของของข่าวประเภทไหนเกิดขึ้นมากที่สุด ดังรูปที่ 3.12

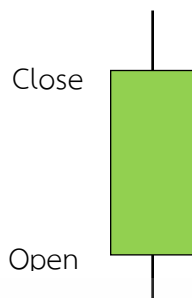
Date	Economic	Political events	Investor sentiment	Company events	Market moment	Natural disasters	Industry specific events
2023-01-13	0.107085	0.188655	0.224613	0.182877	0.018141	0.000000	0.278630
2023-01-12	0.148257	0.198988	0.222646	0.184016	0.000000	0.005107	0.240986
2023-01-11	0.141312	0.205758	0.232759	0.181498	0.017255	0.000000	0.221418
2023-01-10	0.125085	0.204916	0.198564	0.185280	0.048475	0.000000	0.237679
2023-01-09	0.106260	0.184371	0.245682	0.199064	0.036256	0.000000	0.228366
---	---	---	---	---	---	---	---
2019-01-02	0.098977	0.231739	0.221568	0.176368	0.000000	0.010215	0.261132
2019-12-31	0.101232	0.191571	0.271121	0.190650	0.053154	0.000000	0.192271
2018-12-28	0.106280	0.198398	0.240056	0.197980	0.055907	0.000000	0.201378
2018-05-25	0.097899	0.265802	0.205467	0.154724	0.000000	0.035712	0.239396

รูปที่ 3.12 อัตราส่วนข่าวที่เกิดขึ้นแต่ละประเภทของแต่ละวัน

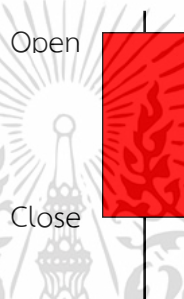
3.3.2 ปัจจัยเชิงเทคนิค

จากรายงานของผู้วิจัยมีความสนใจในการพยากรณ์ทิศทางของการเปลี่ยนแปลงราคาหุ้นในกลุ่ม SET50 เนื่องจากผู้วิจัยต้องการตัวแปรตามให้อยู่ในรูปแบบข้อความ (Categorical Data) แต่ผู้วิจัยได้ทำการรวบรวมข้อมูลตัวเลข (Numerical Data) ดังนั้นผู้วิจัยทำการแปลงข้อมูลตัวเลขเป็นข้อมูลประเภทข้อความ โดยใช้ราคาเปิด (Open price) และราคาปิด (Close price) ในการระบุทิศทางของการเปลี่ยนแปลงของราคาหุ้น

- ราคาปิดมากกว่าราคาเปิดเท่ากับราคาหุ้นขึ้น ดังรูปที่ 3.13
- ราคาปิดน้อยกว่าราคาเปิดเท่ากับราคาหุ้นลง ดังรูปที่ 3.14
- ราคาปิดเท่ากับราคาเปิดเท่ากับราคาหุ้นไม่มีการเปลี่ยนแปลง ดังรูปที่ 3.15



รูปที่ 3.13 การเปลี่ยนแปลงราคาหุ้นในทิศทางขึ้น



รูปที่ 3.14 การเปลี่ยนแปลงราคาหุ้นในทิศทาง

Open = Close

รูปที่ 3.15 การเปลี่ยนแปลงราคาหุ้นที่ไม่มี

ผู้วิจัยทำการสร้างปัจจัยเชิงเทคนิคต่างๆ สำหรับเป็นตัวแปรอิสระในการฝึกสอนแบบจำลองดังนี้

1. การคำนวณค่าสหสัมพันธ์ระหว่าง 2 ตัวแปร ของชุดข้อมูลในรูปแบบของหน้าต่างเคลื่อน (Rolling Correlation) เพื่อวิเคราะห์ความสัมพันธ์ระหว่างราคาหุ้นในช่วงเวลาที่ต่างกัน หรือในการวิเคราะห์รูปแบบการเคลื่อนไหวของข้อมูลที่มีลักษณะเปลี่ยนแปลงตามเวลา
 - กำหนดระยะเวลา (window size) ที่ต้องการใช้ในการคำนวณค่าความสัมพันธ์ระหว่างตัวแปร ซึ่งจะเป็นจำนวนช่วงเวลาที่ถูกรับรวมในการคำนวณ

- คำนวณค่าความสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y โดยใช้สูตร correlation coefficient ดังนี้

$$\text{Correlation} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}$$

เมื่อ $\text{cov}(x, y)$ คือ ค่าความแปรปรวนร่วม ระหว่างตัวแปร x และตัวแปร y

σ_x คือ ส่วนเบี่ยงเบนมาตรฐานของตัวแปร x

σ_y คือ ส่วนเบี่ยงเบนมาตรฐานของตัวแปร y

- เก็บค่าความสัมพันธ์ (Correlation) ที่ได้ในแต่ละช่วงเวลา และเลื่อนตามระยะเวลา (Rolling) เพื่อคำนวณค่าความสัมพันธ์ในช่วงถัดไป ผู้วิจัยมีการสร้างตัวแปรค่าความสัมพันธ์เคลื่อนที่ (Rolling Correlation) อาทิ ดัชนี DJI กับ ดัชนี NDX, ดัชนี DJI กับ ดัชนี SPX, ดัชนี DJI กับ ดัชนี SET, ดัชนี SET กับหุ้นกลุ่ม SET50, ดัชนี SET50 กับหุ้นกลุ่ม SET50, ดัชนี SET50 กับ ดัชนี SET



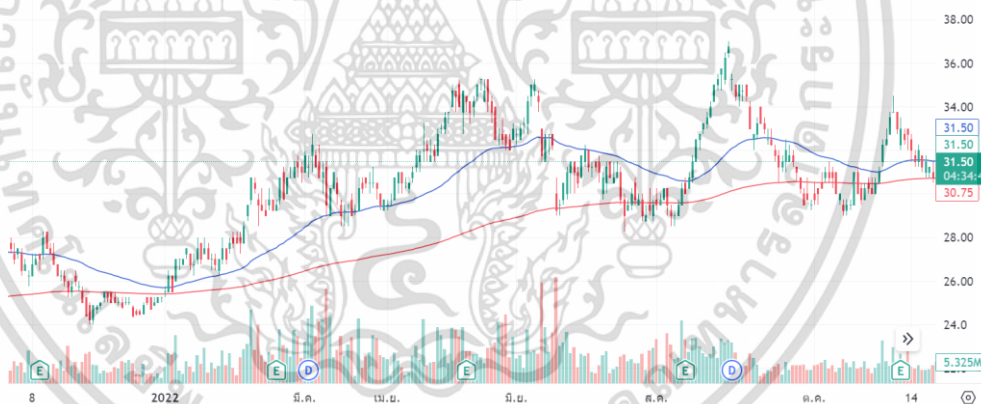
รูปที่ 3.16 ตัวอย่างการคำนวณ Rolling Correlation ของ S&P500 และ Crude Oil Future

ที่มา : <https://www.codearmo.com/python-tutorial/moving-averages-correlations-pandas>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. การระบุแนวโน้มหลัก (Primary Trend) โดยใช้เส้น EMA (Exponential Moving Average) เป็นตัวบ่งชี้ทางเทคนิคที่ใช้ในการระบุเทรนของหุ้น โดยให้น้ำหนักมากกับข้อมูลล่าสุด ซึ่ง EMA มีการตอบสนองต่อการเปลี่ยนแปลงของราคาได้เร็วขึ้น ดังนั้น EMA มีความสามารถในการจับแนวโน้มของราคาที่เกิดขึ้นในช่วงระยะเวลาเฉพาะ ผู้วิจัยใช้เส้น EMA 50 วัน หรือ 10 สัปดาห์สะท้อนการเคลื่อนไหวของราคาหุ้นระยะกลาง (เส้นสีน้ำเงิน) กับเส้น EMA 200 วัน หรือ 40 สัปดาห์สะท้อนการเคลื่อนไหวของราคาหุ้นระยะยาว (เส้นสีแดง) ในการระบุแนวโน้มหลักของหุ้น

- เส้น EMA 50 วัน มากกว่า เส้น EMA 200 วัน หมายถึงราคาหุ้นในระยะยาวยังเป็นขาขึ้น ดังรูปที่ 3.17
- เส้น EMA 50 วัน น้อยกว่า เส้น EMA 200 วัน หมายถึงราคาหุ้นในระยะยาวยังเป็นขาลง ดังรูปที่ 3.18



รูปที่ 3.17 แนวโน้มหลักขาขึ้น



รูปที่ 3.18 แนวโน้มหลักขาลง

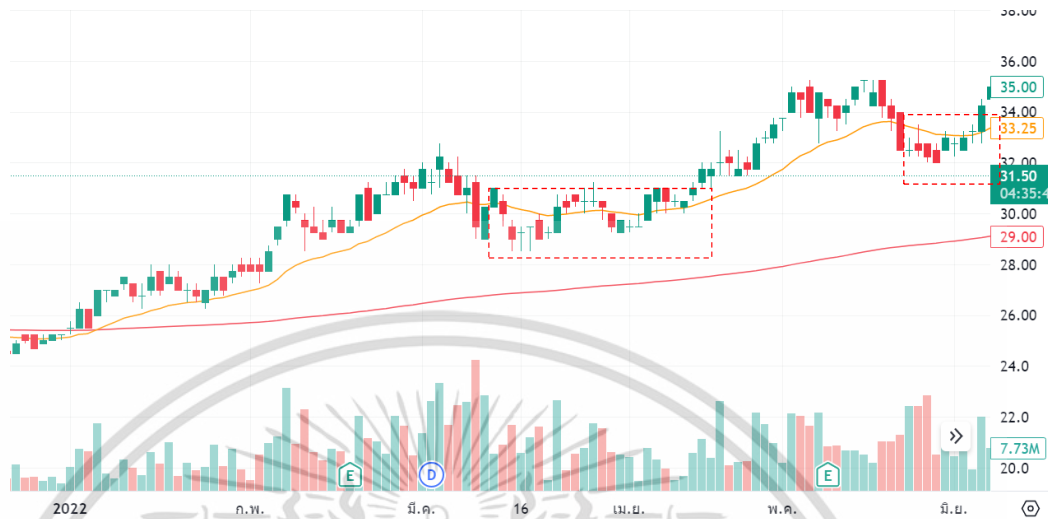
3. การระบุแนวโน้มระยะสั้น (Minor Trend or Short_Trend) โดยใช้เส้น EMA (Exponential Moving Average) เหมือนกับการระบุแนวโน้มหลัก ผู้วิจัยใช้เส้น EMA 20 วัน หรือ 4 สัปดาห์สะท้อนการเคลื่อนไหวของราคาหุ้นระยะสั้น (เส้นสีส้ม) กับราคาหุ้น ณ ปัจจุบัน ในการระบุแนวโน้มระยะสั้นของหุ้น

- ราคาหุ้น ณ ปัจจุบัน มากกว่า เส้น EMA 20 วัน หมายถึงราคาหุ้นในระยะสั้นเป็นขาขึ้น ดังรูปที่ 3.19
- ราคาหุ้น ณ ปัจจุบัน น้อยกว่า เส้น EMA 20 วัน หมายถึงราคาหุ้นในระยะสั้นเป็นขาลง ดังรูปที่ 3.20



รูปที่ 3.19 แนวโน้มระยะสั้นขาขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.20 แนวโน้มระยะสั้นขาลง

4. การคำนวณความต่างหรือระยะห่างของเส้น EMA (EMA Distance) เพื่อหาช่วงที่กำลังจะเกิดแนวโน้มใหม่ ผู้วิจัยใช้เส้น EMA 50 วัน กับเส้น EMA 200 วัน ในการระบุจุดเริ่มต้นของแนวโน้ม ดังรูปที่ 3.21

$$EMA\ Distance(\%) = \frac{EMA_{50} - EMA_{200}}{EMA_{200}} \times 100$$

เมื่อ EMA_{50} คือ ค่า Exponential Moving Average (EMA) ที่คำนวณโดยใช้ระยะเวลา 50 วัน

EMA_{200} คือ ค่า Exponential Moving Average (EMA) ที่คำนวณโดยใช้ระยะเวลา 200 วัน

- ระยะห่างเส้น EMA 50 วัน กับเส้น EMA 200 วัน มีระยะห่างที่กว้าง หมายถึงแนวโน้มของราคาที่แข็งแกร่ง
- ระยะห่างเส้น EMA 50 วัน กับเส้น EMA 200 วัน มีระยะห่างที่แคบ หมายถึงแนวโน้มของราคาที่อ่อนแอ



รูปที่ 3.21 ระยะห่างของเส้น EMA และราคา

5. การระบุจุดที่มีนัยสำคัญของราคา (Critical Point) แสดงถึงช่วงของราคา ณ ปัจจุบันมีการย่อลงมาหรือชนเส้น EMA 50 วัน หรือ EMA 200 วัน เมื่อราคาเข้าใกล้หรืออยู่บนเส้น EMA 50 วัน หรือ EMA 200 วัน ราคาจะมีการเคลื่อนไหวไปตามแนวโน้มหลัก
- สำหรับแนวโน้มหลักขึ้น เส้น EMA 50 มากกว่า EMA 200 และราคาจะมีการย่อลงมาเส้น EMA 50 หรือ EMA 200 ราคาจะมีการติดตัวขึ้นไป ดังรูปที่ 3.22
 - สำหรับแนวโน้มขาลง เมื่อเส้น EMA 50 น้อยกว่า EMA 200 และราคาจะมีการชนเส้น EMA 50 หรือ EMA 200 ราคาจะมีการติดตัวลงไป ดังรูปที่ 3.23



รูปที่ 3.22 จุดที่มีนัยสำคัญสำหรับแนวโน้มหลักขาขึ้น

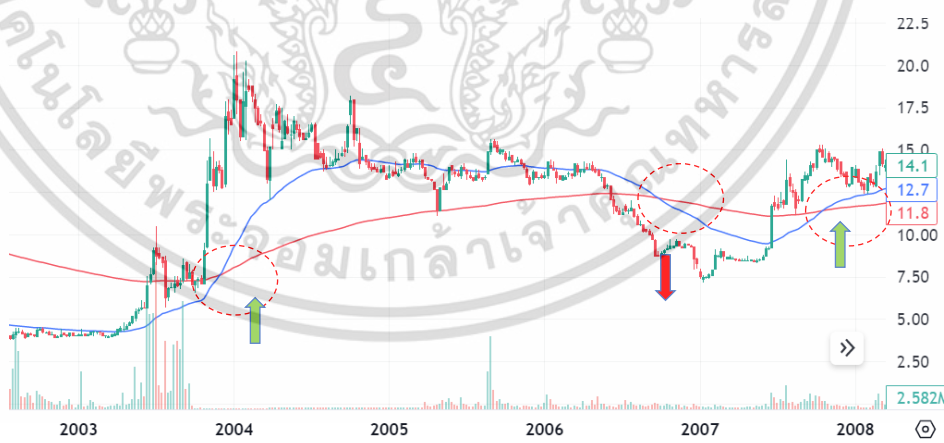
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.23 จุดที่มีนัยสำคัญสำหรับแนวโน้มหลักขาลง

6. การเปลี่ยนแปลงแนวโน้มหลัก (Changing of Trends) ผู้วิจัยใช้เส้น EMA 50 กับ EMA 200 ในการหาการเปลี่ยนแปลงแนวโน้มหลัก ดังรูปที่ 3.24

- Golden Cross คือ ราคาหุ้นกำลังจะเปลี่ยนแนวโน้มขาขึ้น เส้น EMA 50 ตัดขึ้นกับเส้น EMA 200
- Death Cross คือ ราคาหุ้นกำลังจะเปลี่ยนแนวโน้มขาลง เส้น EMA 50 ตัดลงกับเส้น EMA 200



รูปที่ 3.24 การเปลี่ยนแปลงแนวโน้มหลัก

7. การระบุจำนวนจุดที่มีนัยสำคัญของราคา (Critical Point Occurrence) แสดงถึงราคาใช้เวลาในการพักตัวหรืออยู่ในกรอบ sideways นานเท่าไร ก่อนที่ราคาจะมีเคลื่อนไหวออก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากกรอบ ผู้วิจัยใช้เส้น EMA 50 กับ เส้น EMA 200 ในการตรวจสอบว่าราคามีการแตะเส้นเหล่านี้จำนวนกี่ครั้ง ดังรูปที่ 3.25



รูปที่ 3.25 การเปลี่ยนแปลงราคาหุ้นที่ไม่มีทิศทาง

8. การระบุจำนวนความผิดปกติของปริมาณการซื้อขาย (Volume Anomaly) แสดงถึงมีนักลงทุนให้ความสนใจกับหุ้นตัวนี้จนเกิดปริมาณการซื้อขายที่ผิดปกติจากปกติ ผู้วิจัยใช้คะแนนมาตรฐาน (Z-score) ในการหาจำนวนปริมาณที่ผิดปกติ

$$Z = \frac{(X - \bar{X})}{S.D.} > threshold$$

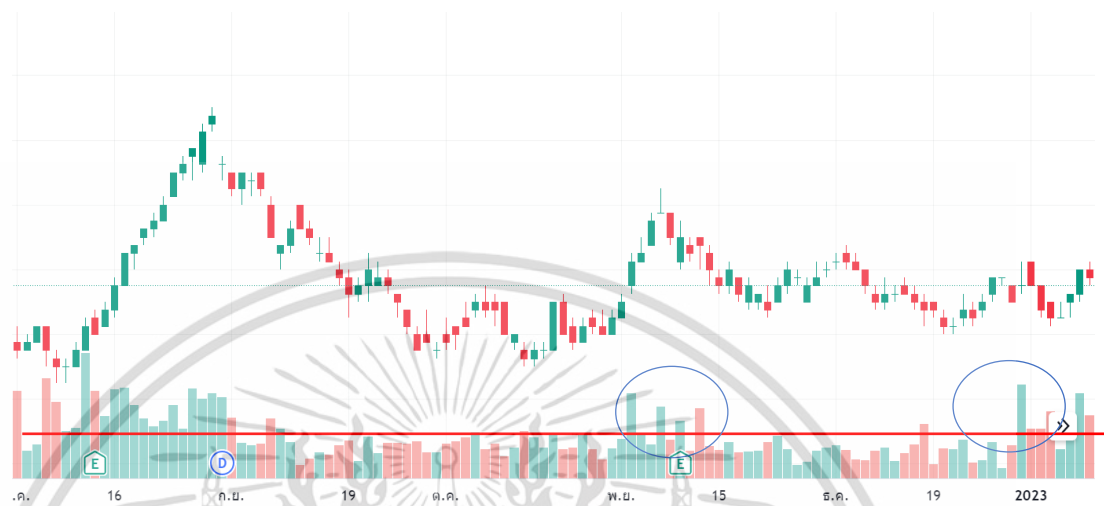
เมื่อ X คือ ปริมาณการซื้อขาย ณ ช่วงเวลานั้น

\bar{X} คือ ปริมาณการซื้อขายเฉลี่ยในกรอบเวลา (rolling window)

$S.D.$ คือ ความเบี่ยงเบนมาตรฐานของปริมาณการซื้อขายในกรอบเวลา (rolling window)

threshold คือ ค่าที่กำหนดเป็นอันดับขั้นต่ำสำหรับค่าคะแนนมาตรฐาน (Z-score) เพื่อกำหนดว่าปริมาณการซื้อขายผิดปกติหรือไม่ ผู้วิจัยใช้ค่า 0.6-0.8 โดยการใช้ค่า **threshold** จึงควรมีการสำรวจข้อมูลของหุ้นในตัวนั้นเบื้องต้น เพื่อให้ค่าเหมาะสมกับปริมาณการซื้อขายของหุ้นตัวนั้น

โดยตั้งเงื่อนไขว่าถ้าคะแนนมาตรฐาน (Z-score) มากกว่า threshold (เส้นสีแดง) แสดงถึงปริมาณการซื้อขายที่ผิดปกติ ดังรูปที่ 3.26



รูปที่ 3.26 ปริมาณการซื้อขายที่ผิดปกติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 ตัวแปรที่ใช้ในการวิเคราะห์

3.4.1 ตัวแปรตาม (Dependent Variable)

ตารางที่ 3.1 ตัวแปรตาม

ลำดับ	ชื่อตัวแปร	คำอธิบาย	ประเภทข้อมูล
1	*_TARGET	ทิศทางการเปลี่ยนแปลงราคา 1: ราคาขึ้น แห่งเทียนสีเขียว 0: ราคาสูง แห่งเทียนสีแดง -1: ราคาไม่มีการเปลี่ยนแปลง ไม่มี แห่งเทียน	นามบัญญัติ (Nominal)

*หุ้น SET50

3.4.2 ตัวแปรอิสระ (Independent Variable)

ตารางที่ 3.2 ตัวแปรอิสระปัจจัยข่าว

ชื่อตัวแปร	คำอธิบาย	ค่าที่เกี่ยวข้อง	ประเภทข้อมูล
ข่าวทางเศรษฐกิจ (Economic)	เป็นข่าวที่เกี่ยวข้องกับสถานการณ์เศรษฐกิจทั่วไป เช่น ข้อมูลการเจริญเติบโตของ GDP, อัตราการเงินเฟ้อ, อัตราเงินเฟ้อ, อัตราดอกเบี้ย, สถิติการว่างงาน เป็นต้น ข่าวที่เกี่ยวข้องกับเศรษฐกิจสามารณมีผลต่อทิศทางการเคลื่อนไหวของตลาดหุ้นได้โดยตรง	<ul style="list-style-type: none"> ● GDP ● Inflation ● Interest rates ● Economy ● CPI ● PPI ● Employment rate ● Unemployment rate ● Labor force participation rate 	นามบัญญัติ (Nominal)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

		<ul style="list-style-type: none"> ● Federal Reserve ● Monetary policy ● Fiscal policy ● GNP ● National debt ● Trade balance ● Consumer confidence ● PMI ● Retail sales ● Housing starts ● Building permits ● Consumer spending ● Business investment ● Corporate profits ● Productivity ● Economic growth ● Recession ● Recovery ● Government 	
--	--	---	--

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

		<p>spending</p> <ul style="list-style-type: none"> ● Net exports 	
<p>ข่าวทางการเมือง (Political events)</p>	<p>เป็นข่าวที่เกี่ยวข้องกับเหตุการณ์ทางการเมือง เช่น การเลือกตั้งทางการเมือง, นโยบายเศรษฐกิจที่เกี่ยวข้องกับธุรกิจหรือตลาดหุ้น, การประชุมสุดยอดระดับประเทศหรือนานาชาติ เป็นต้น ข่าวทางการเมืองสามารถสร้างผลกระทบต่อทิศทางของตลาดหุ้นได้</p>	<ul style="list-style-type: none"> ● Election ● Policy change ● Geopolitical tension ● Trade policy ● Voting ● Polls ● Ballots ● Candidates ● Voters ● Presidential ● Policy change ● Tariffs ● Trade war ● Sanctions ● Geopolitical tension ● Conflict ● War ● Terrorism ● Embargo ● Diplomacy 	<p>นามบัญญัติ (Nominal)</p>
<p>เหตุการณ์ทางบริษัท</p>	<p>เป็นข่าวที่เกี่ยวข้องกับกิจกรรมของบริษัท</p>	<ul style="list-style-type: none"> ● Earnings report ● Earnings call 	<p>นามบัญญัติ (Nominal)</p>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(Company events)	<p>ซึ่งสามารถมีผลต่อราคาหุ้นของบริษัทนั้น ๆ เช่น</p> <p>ผลประกอบการ,</p> <p>การเปิดเผยข้อมูลทางการเงิน,</p> <p>การเปลี่ยนแปลงในผู้บริหาร เป็นต้น</p>	<ul style="list-style-type: none"> ● Quarterly results ● Annual results ● Financial results ● Revenue ● Sales ● Top line ● Profits ● Bottom line ● Net income ● Earnings per share ● Losses ● Negative earnings ● Loss per share ● Guidance ● Outlook ● Forecast ● Dividend ● Dividend payout 	
<p>อารมณ์นักลงทุน (Investor sentiment)</p>	<p>เป็นข่าวที่เกี่ยวข้องกับจิตวิญญาณและทัศนคติของนักลงทุน</p> <p>อารมณ์นักลงทุนสามารถ</p>	<ul style="list-style-type: none"> ● Bullish ● Bearish ● Optimistic 	<p>นามบัญญัติ (Nominal)</p>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	<p>เป็นปัจจัยที่ส่งผลต่อความเสี่ยงและความเชื่อมั่นของตลาดหุ้น เช่น การเปลี่ยนแปลงในความเชื่อมั่นของนักลงทุนต่อตลาดหุ้น หรือสถานการณ์ทางการเงินที่สร้างความไม่แน่นอน</p>	<ul style="list-style-type: none"> ● Pessimistic ● Rally ● Crash ● Hopeful ● Worried ● Uncertain ● Volatile ● Fearful ● Greedy ● Fear/greed 	
<p>การเคลื่อนไหวของตลาด (Market moment)</p>	<p>เป็นข่าวที่เกี่ยวข้องกับเหตุการณ์ที่เกิดขึ้นในตลาดหุ้นโดยตรง เช่น การเปิดตลาด, การปิดตลาด, การเทรตที่มีปริมาณมากขึ้นหรือน้อยลง, การเคลื่อนไหวของราคาหุ้นในเวลาสั้น ๆ เป็นต้น</p>	<ul style="list-style-type: none"> ● Bull market ● Bear market ● Market correction ● Market volatility ● Market rally ● Market downturn ● Market trend ● Market upswing ● Market downswing 	<p>นามบัญญัติ (Nominal)</p>
<p>ภัยธรรมชาติ (Natural disasters)</p>	<p>เป็นข่าวที่เกี่ยวข้องกับภัยธรรมชาติ เช่น พายุ, แผ่นดินไหว, น้ำท่วม เป็นต้น</p>	<ul style="list-style-type: none"> ● Earthquake ● Tsunami ● Hurricane 	<p>นามบัญญัติ (Nominal)</p>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	ภัยธรรมชาติสามารถมีผลต่อบริษัทและเศรษฐกิจทั่วไปซึ่งส่งผลต่อตลาดหุ้นได้	<ul style="list-style-type: none"> ● Tornado ● Flood 	
เหตุการณ์เฉพาะกลุ่มอุตสาหกรรม (Industry specific events)	เป็นข่าวที่เกี่ยวข้องกับเหตุการณ์ที่เกิดขึ้นในกลุ่มอุตสาหกรรมเฉพาะ เช่น การประกาศผลิตภัณฑ์ใหม่, นวัตกรรมในอุตสาหกรรม, การเปลี่ยนแปลงในกฎหมายที่เกี่ยวข้องกับอุตสาหกรรม เป็นต้น	<ul style="list-style-type: none"> ● Biotech ● Pharmaceutical ● Healthcare ● Energy ● Oil and gas ● Technology ● Software ● Hardware ● Telecom ● Internet ● Ecommerce ● Retail ● Automotive ● Transportation ● Aviation ● Shipping ● Logistics ● Real estate ● Construction ● Hospitality ● Travel 	นามบัญญัติ (Nominal)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.3 ตัวแปรอิสระปัจจัยเชิงเทคนิครายวัน

ลำดับ	ชื่อตัวแปร	คำอธิบาย	ประเภทข้อมูล
1	**_CLOSE_PRICE	ราคาปิด	อัตราส่วน (Ratio)
2	**_VOLUME	ปริมาณการซื้อขาย	อัตราส่วน (Ratio)
3	**_PRICE_DIRECTION	ทิศทางการเปลี่ยนแปลงราคา แบ่งเป็น 3 สถานะ 1: ราคาขึ้น 0: ราคาลง -1: ราคาไม่มีการ เปลี่ยนแปลง	นามบัญญัติ (Nominal)
4	***_CLOSE_PRICE_T-1	ราคาปิดของวันก่อนหน้า 1 วัน	อัตราส่วน (Ratio)
5	***_VOLUME_T-1	ปริมาณการซื้อขายของวันก่อน หน้า 1 วัน	อัตราส่วน (Ratio)
6	ROLLING_CORRELATION_DJI_N DX	การวิเคราะห์ความสัมพันธ์ ระหว่าง 2 ตัวแปรใน ช่วงเวลาต่อเนื่อง โดย คำนวณค่าสหสัมพันธ์ ระหว่างตัวแปรราคาดัชนี Nasdaq กับราคาดัชนี Dow Jones ในช่วงเวลา 20 วัน	อัตราส่วน (Ratio)
7	ROLLING_CORRELATION_ _DJI_SPX	การวิเคราะห์ความสัมพันธ์ ระหว่าง 2 ตัวแปรใน ช่วงเวลาต่อเนื่อง โดย คำนวณค่าสหสัมพันธ์	อัตราส่วน (Ratio)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

		ระหว่างตัวแปรราคาดัชนี S&P500 กับราคาดัชนี Dow Jones ในช่วงเวลา 20 วัน	
8	ROLLING_CORRELATION _DJI_SET_T-1	การวิเคราะห์ความสัมพันธ์ระหว่าง 2 ตัวแปรในช่วงเวลาต่อเนื่อง โดยคำนวณค่าสหสัมพันธ์ระหว่างตัวแปรราคาดัชนี SET ย้อนหลัง 1 วัน กับราคาดัชนี Dow Jones ในช่วงเวลา 20 วัน	อัตราส่วน (Ratio)
9	*** ROLLING_CORRELATION STOCK_SET	การวิเคราะห์ความสัมพันธ์ระหว่าง 2 ตัวแปรในช่วงเวลาต่อเนื่อง โดยคำนวณค่าสหสัมพันธ์ระหว่างตัวแปรราคาหุ้นรายตัวในกลุ่ม SET50 กับราคาดัชนี SET ในช่วงเวลา 20 วัน	อัตราส่วน (Ratio)
10	*** ROLLING_CORRELATION STOCK_SET50	การวิเคราะห์ความสัมพันธ์ระหว่าง 2 ตัวแปรในช่วงเวลาต่อเนื่อง โดยคำนวณค่าสหสัมพันธ์ระหว่างตัวแปรราคาหุ้นรายตัวในกลุ่ม SET50 กับราคาดัชนี SET50 ในช่วงเวลา 20 วัน	อัตราส่วน (Ratio)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

11	ROLLING_CORRELATION SET_SET50	การวิเคราะห์ความสัมพันธ์ ระหว่าง 2 ตัวแปรใน ช่วงเวลาต่อเนื่อง โดย คำนวณค่าสหสัมพันธ์ ระหว่างตัวแปรราคาดัชนี SET50 กับราคาดัชนี SET ในช่วงเวลา 20 วัน	อัตราส่วน (Ratio)
12	***_EMA200	เส้นค่าเฉลี่ยราคาเคลื่อนที่ 200 วัน ใช้ในการวิเคราะห์ และติดตามแนวโน้มของ ราคาหรือดัชนีหลักทรัพย์ โดยให้น้ำหนักมากกว่าในวัน ล่าสุด เพื่อให้มีการตอบสนอง ที่ดีต่อการเปลี่ยนแปลงใน ราคาในช่วงล่าสุด	อัตราส่วน (Ratio)
13	***_EMA50	เส้นค่าเฉลี่ยราคาเคลื่อนที่ 50 วัน ใช้ในการวิเคราะห์ และติดตามแนวโน้มของ ราคาหรือดัชนีหลักทรัพย์ โดยให้น้ำหนักมากกว่าในวัน ล่าสุด เพื่อให้มีการตอบสนอง ที่ดีต่อการเปลี่ยนแปลงใน ราคาในช่วงล่าสุด	อัตราส่วน (Ratio)
14	***_EMA20	เส้นค่าเฉลี่ยราคาเคลื่อนที่ 20 วัน ใช้ในการวิเคราะห์ และติดตามแนวโน้มของ ราคาหรือดัชนีหลักทรัพย์	อัตราส่วน (Ratio)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

		โดยให้น้ำหนักมากกว่าในวัน ล่าสุด เพื่อให้มีการตอบสนอง ที่ต่อการเปลี่ยนแปลงใน ราคาในช่วงล่าสุด	
15	***_RSI	ตัวชี้วัดทางเทคนิคที่ใช้ในการ วิเคราะห์และติดตามสถานะ การซื้อขายของหลักทรัพย์ เป็นตัวชี้วัดที่เชื่อมโยงกับ ความแข็งแกร่งของการ เคลื่อนไหวราคา	อัตราส่วน (Ratio)
16	***_MACD	ตัวชี้วัดทางเทคนิคที่ใช้ในการ วิเคราะห์แนวโน้มและ สัญญาณการซื้อขายของ หลักทรัพย์	อัตราส่วน (Ratio)
17	***_SIGNAL_LINE	เส้นค่าเฉลี่ยเคลื่อนที่ (Exponential Moving Average, EMA) ที่คำนวณ จากค่า MACD ในระยะเวลา ที่กำหนด (โดยทั่วไปใช้ ระยะเวลา 9 วัน) โดยช่วยใน การวิเคราะห์แนวโน้มและ สัญญาณการซื้อขายใน MACD	อัตราส่วน (Ratio)
18	***_MACD_SIGNAL	สัญญาณของ MACD แบ่ง ออกเป็น 3 สถานะ 1. Bullish:	นามบัญญัติ (Nominal)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

		<p>แทนสถานะการซื้อขายที่เชื่อว่า มีแนวโน้มที่ตีขึ้นหรือเป็น กำลังขึ้น</p> <p>กล่าวคือเมื่อเกิดการตัดกันของ เส้น MACD</p> <p>ขึ้นไปเหนือเส้นสัญญาณ (Signal Line) และค่า MACD_Diff เป็นบวก</p> <p>2. Bearish:</p> <p>แทนสถานะการซื้อขายที่เชื่อว่า มีแนวโน้มที่แยลงหรือเป็น กำลังลง</p> <p>กล่าวคือเมื่อเกิดการตัดกันของ เส้น MACD</p> <p>ลงมาใต้เส้นสัญญาณ (Signal Line) และค่า MACD_Diff เป็นลบ</p> <p>3. Neutral:</p> <p>แทนสถานะการซื้อขายที่ไม่ อยู่ในเงื่อนไข Bullish หรือ Bearish</p> <p>กล่าวคือเมื่อไม่เกิดการตัดกัน ของเส้น MACD</p> <p>กับเส้นสัญญาณ (Signal Line) หรือค่า MACD_Diff เป็นศูนย์หรือใกล้เคียงศูนย์</p>	
19	***COUNT_VOLUME_ ANOMALY	<p>จำนวนการเกิดปริมาณการซื้อ ขายที่</p> <p>ผิดปกติในช่วงเวลา 20 วัน</p>	เรียงลำดับ (Ordinal)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

20	***_GOLDER_CROSS	<p>สัญญาณการเปลี่ยนแนวโน้ม เมื่อเส้น EMA50 ขึ้นไปเหนือเส้น EMA200 แบ่งเป็น 2 สถานะ</p> <ol style="list-style-type: none"> 1. True = Golden Cross 2. False = Death Cross 	<p>นามบัญญัติ (Nominal)</p>
21	***_PRIMARY_TREND	<p>แนวโน้มหลักของราคาหุ้นจากร เส้น EMA50 และ EMA200 แบ่งเป็น 3 สถานะ</p> <ol style="list-style-type: none"> 1. Up trend = ขาขึ้น 2. Down trend = ขาลง 3. Sideways trend = ไม่มีแนวโน้ม 	<p>นามบัญญัติ (Nominal)</p>
22	***_PRICE_TREND	<p>แนวโน้มของราคาหุ้น ณ ปัจจุบันจากราคาปิดของหุ้น และเส้น EMA50 กับ EMA200 แบ่งเป็น 7 สถานะ</p> <ol style="list-style-type: none"> 1. Up Trend = ราคาหุ้นมีแนวโน้มขึ้น 2. Weak Up Trend = ราคาหุ้นมีแนวโน้มอ่อนแอใน การขึ้น 3. Down Trend Coming = ราคาหุ้นมีแนวโน้มเปลี่ยนแปลง ลงจาก ขาขึ้นเป็นขาลง 	<p>นามบัญญัติ (Nominal)</p>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

		<p>4. Down Trend = ราคาหุ้นมีแนวโน้มลง</p> <p>5. Weak Down Trend = ราคาหุ้นมีแนวโน้มอ่อนในกา รลง</p> <p>6. Up Trend Coming = ราคาหุ้นมีแนวโน้มเปลี่ยนแปลง ลงจาก ขาลงเป็นขาขึ้น</p> <p>7. Sideways Trend = ราคาหุ้นไม่มีแนวโน้ม</p>	
23	***_SHORT_TREND	<p>แนวโน้มระยะสั้นของราคาหุ้น ณ ปัจจุบันจากราคาปิดของหุ้น และเส้น EMA20</p> <p>แบ่งเป็น 2 สถานะ</p> <p>1. Strong = ราคาหุ้นเหนือกว่าเส้น EMA20</p> <p>2. Weak = ราคาหุ้นต่ำกว่าเส้น EMA20</p>	นามบัญญัติ (Nominal)
24	***_EMA_DISTANCE	ระยะความห่างของเส้น EMA50 และเส้น EMA200	อัตราส่วน (Ratio)
25	***_CHANGING_OF_TRENDS	การเปลี่ยนแปลงของแนวโน้มจาก ความห่างของเส้น EMA50 และเส้น EMA200	นามบัญญัติ (Nominal)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

26	***_CRITICAL_POINT	ราคาปิดของหุ้นที่มีการสัมผัสเส้น EMA แบ่งเป็น 2 สถานะ 1. EMA50 Critical Point = ราคาหุ้นสัมผัสเส้น EMA50 2. EMA200 Critical Point = ราคาหุ้นสัมผัสเส้น EMA200	นามบัญญัติ (Nominal)
27	***_CRITICAL_POINT_OCCURRENCE	จำนวนครั้งที่ราคาหุ้นสัมผัสเส้น EMA50 และ EMA200	เรียงลำดับ (Ordinal)

โดยปัจจัยทางเทคนิคเหล่านี้ผู้วิจัยทำการคำนวณในช่วงเดือน มกราคม พ.ศ.2566

** ดัชนีต่างประเทศ

*** ดัชนีและหุ้น SET50 ในประเทศ

ตารางที่ 3.4 ตัวแปรอิสระดัชนีราคาต่างประเทศและดัชนีราคาในประเทศรายวัน

ลำดับ	ชื่อตัวแปร	คำอธิบาย	ประเภทข้อมูล
1	DJI	ดัชนี Dow Jones ที่ประกอบด้วยหุ้นขนาดใหญ่ 30 บริษัทในตลาดหลักทรัพย์ New York และตลาดหลักทรัพย์ Nasdaq	อัตราส่วน (Ratio)
2	NDX	ดัชนี Nasdaq ที่สะท้อนความเคลื่อนไหวของ 100 บริษัทที่ใหญ่ที่สุดที่จดทะเบียนในตลาดหลักทรัพย์ Nasdaq	อัตราส่วน (Ratio)
3	SPX	ดัชนี S&P500 ที่สะท้อนความเคลื่อนไหวของบริษัทขนาดใหญ่ 500 แห่งที่จดทะเบียนในตลาดหลักทรัพย์ในสหรัฐ	อัตราส่วน (Ratio)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4	SET	ดัชนี SET ที่สะท้อนความเคลื่อนไหวของราคาหลักทรัพย์ทั้งหมดในประเทศไทย	อัตราส่วน (Ratio)
5	SET50	ดัชนี SET50 ที่สะท้อนความเคลื่อนไหวตลาดหลักทรัพย์ราคาหุ้น 50 ตัว	อัตราส่วน (Ratio)
6	AGRI	ดัชนีกลุ่มธุรกิจการเกษตร	อัตราส่วน (Ratio)
7	AUTO	ดัชนีกลุ่มยานยนต์	อัตราส่วน (Ratio)
8	BANK	ดัชนีกลุ่มธนาคาร	อัตราส่วน (Ratio)
9	COMM	ดัชนีกลุ่มพาณิชย์	อัตราส่วน (Ratio)
10	CONMAT	ดัชนีกลุ่มวัสดุก่อสร้าง	อัตราส่วน (Ratio)
11	CONS	ดัชนีกลุ่มบริการรับเหมาก่อสร้าง	อัตราส่วน (Ratio)
12	HOME	ดัชนีกลุ่มของใช้ในครัวเรือนและสำนักงาน	อัตราส่วน (Ratio)
13	ICT	ดัชนีกลุ่มเทคโนโลยีสารสนเทศและการสื่อสาร	อัตราส่วน (Ratio)
14	MEDIA	ดัชนีกลุ่มสื่อและสิ่งพิมพ์	อัตราส่วน (Ratio)
15	PROP	ดัชนีกลุ่มพัฒนาอสังหาริมทรัพย์	อัตราส่วน (Ratio)
16	STEEL	ดัชนีกลุ่มเหล็กและผลิตภัณฑ์โลหะ	อัตราส่วน (Ratio)
17	TRANS	ดัชนีกลุ่มขนส่งและโลจิสติกส์	อัตราส่วน (Ratio)
18	ENERG	ดัชนีกลุ่มพลังงานและสาธารณูปโภค	อัตราส่วน (Ratio)
19	ETRON	ดัชนีกลุ่มชิ้นส่วนอิเล็กทรอนิกส์	อัตราส่วน (Ratio)
20	FASHION	ดัชนีกลุ่มแฟชั่น	อัตราส่วน (Ratio)
21	FIN	ดัชนีกลุ่มเงินทุนและหลักทรัพย์	อัตราส่วน (Ratio)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

22	FOOD	ดัชนีกลุ่มอาหารและเครื่องดื่ม	อัตราส่วน (Ratio)
23	HEALTH	ดัชนีกลุ่มการแพทย์	อัตราส่วน (Ratio)
24	IMM	ดัชนีกลุ่มวัสดุอุตสาหกรรมและเครื่องจักร	อัตราส่วน (Ratio)
25	INSUR	ดัชนีกลุ่มประกันภัยและประกันชีวิต	อัตราส่วน (Ratio)
26	PAPER	ดัชนีกลุ่มกระดาษและวัสดุการพิมพ์	อัตราส่วน (Ratio)
27	PERSON	ดัชนีกลุ่มของใช้ส่วนตัวและเวชภัณฑ์	อัตราส่วน (Ratio)
28	PETRO	ดัชนีกลุ่มปิโตรเคมีและเคมีภัณฑ์	อัตราส่วน (Ratio)
29	PF&REIT	ดัชนีกลุ่มกองทุนรวมอสังหาริมทรัพย์และกองทรัสต์เพื่อการลงทุนในอสังหาริมทรัพย์	อัตราส่วน (Ratio)
30	PKG	ดัชนีกลุ่มบรรจุภัณฑ์	อัตราส่วน (Ratio)
31	PROF	ดัชนีกลุ่มบริการเฉพาะกิจ	อัตราส่วน (Ratio)
32	TOURISM	ดัชนีกลุ่มการท่องเที่ยวและสันทนาการ	อัตราส่วน (Ratio)

โดยดัชนีราคาเหล่านี้ผู้วิจัยทำการคำนวณในช่วงเดือน มกราคม พ.ศ.2566

3.5 การสร้างแบบจำลอง

สำหรับงานวิจัยนี้ผู้วิจัยได้ทำการสร้างแบบจำลองทั้งหมด 5 แบบจำลอง ได้แก่

ซัพพอร์ตตเวกเตอร์แมชชีน (Support Vector Machine) ป่าไม้สุ่ม (Random Forest) อดาบูสต์ (AdaBoost) เกรเดียนท์บูสต์ติง (Gradient Boosting) แคตบูสต์ (CatBoost) และทำการรวบรวมข้อมูลสำหรับการสร้างแบบจำลองตั้งแต่เดือน พฤษภาคม พ.ศ. 2561 ถึง เดือนมกราคม พ.ศ. 2566 จำนวน 631 วัน โดยยึดวันเปิดทำการของตลาดหลักทรัพย์ในประเทศไทยเป็นหลัก แบ่งข้อมูลชุดสอน (Training Data) จำนวน 504 วันหรือคิดเป็นร้อยละ 80 ตั้งแต่เดือน พฤษภาคม พ.ศ.2561 ถึง กรกฎาคม พ.ศ.2565 และข้อมูลชุดทดสอบ (Testing Data) จำนวน 126 วันหรือคิดเป็นร้อยละ 20 ตั้งแต่เดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6 เครื่องมือที่ใช้ในการวิเคราะห์

ผู้วิจัยใช้โปรแกรมภาษาไพธอน (Python) ซึ่งเป็นภาษาโปรแกรม มีความยืดหยุ่นสูง เหมาะกับการวิเคราะห์ข้อมูลและสร้างแบบจำลอง โดยใช้ไลบรารีที่จำเป็น ดังตารางที่ 3.5

ตารางที่ 3.5 ไลบรารีสำหรับการวิเคราะห์ข้อมูล

Library	คำอธิบาย
Pandas	ใช้ในการจัดการและวิเคราะห์ข้อมูลที่มีโครงสร้างแบบตาราง (Tabular Data) อย่างมีประสิทธิภาพ ซึ่งช่วยให้ง่ายต่อการอ่านและเขียนข้อมูลจากแหล่งข้อมูลต่างๆ นอกจากนี้ยังมีฟังก์ชันที่ช่วยในการจัดการข้อมูลในรูปแบบตาราง เช่น การเลือกข้อมูลตามเงื่อนไข การกรองข้อมูล การจัดกลุ่ม และการคำนวณข้อมูลสถิติพื้นฐาน เป็นต้น
Numpy	สำหรับการคำนวณทางคณิตศาสตร์ โดยมีฟังก์ชันทางคณิตศาสตร์ที่หลากหลาย เช่น การเพิ่ม การลบ การคูณ และการหารของอาร์เรย์
Yfinance	ใช้ในการดึงข้อมูลการเงินและราคาหลักทรัพย์จากเว็บไซต์ Yahoo Finance โดยสามารถดึงข้อมูลทางการเงิน เช่น ราคาหุ้น ปริมาณการซื้อขาย และการเปลี่ยนแปลงราคาในช่วงเวลาที่ต้องการ
Scikit-Learn	สำหรับสร้างแบบจำลองการเรียนรู้ของเครื่องจักร โดยมีฟังก์ชันที่รวมถึงการทำนาย การจัดกลุ่ม และการจำแนกข้อมูล ซึ่งให้รองรับการสร้างและปรับแต่งแบบจำลองเครื่องจักร (Machine Learning) ได้ ง่ายสะดวก มีแบบจำลองหลากหลายเช่น การจัดกลุ่มที่ใช้กับข้อมูลที่ไม่มีการเฝ้าติดตาม (Unsupervised Learning) และการจำแนกที่ใช้กับข้อมูลที่มีการติดตามและข้อมูลเป้าหมาย (Supervised Learning)
Fasttext	ไลบรารีสำหรับการจัดการกับการจำแนกหรือการเรียนรู้ข้อมูลลักษณะข้อความอย่างมีประสิทธิภาพ FastText ให้บริการอินเทอร์เฟซที่ใช้ทำงานง่ายสำหรับการฝึกแบบจำลองการจำแนกข้อความ (Text Classification) การฝังคำ (Word Embeddings) และแบบจำลองความคล้ายคลึงข้อความ (Text Similarity)
Catboost	การสร้างแบบจำลองเครื่องจักรที่มีประสิทธิภาพสูงและสามารถจัดการกับข้อมูลที่มีคุณลักษณะแบบหมวดหมู่ได้ดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Plotly	ใช้ในการสร้างกราฟและการแสดงผลข้อมูลที่สวยงามและมีประสิทธิภาพสูง สามารถสร้างกราฟที่หลากหลายรูปแบบได้ เช่น กราฟเส้น กราฟแท่ง กราฟแผนภูมิวงกลม และกราฟแผนที่ สามารถแสดงข้อมูลที่ซับซ้อนได้อย่างชัดเจน
--------	--



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและการอภิปรายผล

สำหรับการวิเคราะห์การทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและการวิเคราะห์เชิงเทคนิคในการสร้างแบบจำลองทั้งหมด 5 แบบจำลอง ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ป่าไม้สุ่ม (Random Forest) อดาบู้สต์ (AdaBoost) เกรเดียนท์บู้สต์ดิง (Gradient Boosting) แคตบู้สต์ (CatBoost) สำหรับการอภิปรายผลบทที่ 4 ทางผู้วิจัยได้เลือกหุ้นทั้งหมด 5 ตัว จาก 50 ตัว ในช่วงเดือน มกราคม พ.ศ.2566 โดยพิจารณาตามมูลค่าหลักทรัพย์ตามราคาตลาดเพื่อนำมาแสดงตัวอย่างการวิเคราะห์ผลการวิจัยเป็นบางตัวเท่านั้น ดังตารางที่ 4.1

ตารางที่ 4.1 รายชื่อหุ้นใน SET50 ทั้งหมด 5 ตัว ตามมูลค่าหลักทรัพย์ตามราคาตลาด

ลำดับ	ชื่อหุ้น	มูลค่าหลักทรัพย์ตามราคาตลาด (ล้านบาท)
1	DELTA	1,163,183
2	AOT	999,999
3	PTT	885,453
4	ADVANC	630,532
5	GULF	569,058

การแบ่งชุดข้อมูลสอนกับชุดข้อมูลทดสอบสำหรับการสร้างแบบจำลองผู้วิจัยมีการปรับสัดส่วนตัวแปรตาม (Stratify) ทั้ง 3 คลาส ได้แก่ ราคาสูง ราคาขึ้น ราคาไม่มีแนวโน้ม เพื่อให้การกระจายตัวของคลาสตัวแปรตามยังคงมีสัดส่วนที่เหมือนกันทั้งในชุดข้อมูลสอนและชุดข้อมูลทดสอบ สิ่งนี้เป็นประโยชน์กับชุดข้อมูลที่ไม่สมดุล (Imbalanced Data) ช่วยให้แบบจำลองเรียนรู้และทดสอบประสิทธิภาพความแม่นยำได้อย่างถูกต้องจากการรักษาความสมดุลของคลาสตัวแปรตาม ดังตารางที่ 4.3 4.7 4.11 4.15 และ 4.19 ตามลำดับ ผู้วิจัยแบ่งข้อมูลสอนแบบจำลอง (Training Data) ทั้งหมด 505 วัน ช่วงเดือน พฤษภาคม พ.ศ.2561 ถึง กรกฎาคม พ.ศ.2565 ข้อมูลทดสอบ (Testing Data) ทั้งหมด 126 วัน ช่วงเดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 พร้อมกับตัวแปรอิสระทั้งหมด 230 ตัวแปร กับ 5

แบบจำลอง โดยกำหนดพารามิเตอร์จากค่าตั้งต้นของแบบจำลองและปรับพารามิเตอร์ของแต่ละแบบจำลองให้เหมาะสมกับปัญหาการจำแนกประเภท (Classification) ดังตารางที่ 4.2

ตารางที่ 4.2 การตั้งค่าพารามิเตอร์สำหรับแบบจำลอง

แบบจำลอง	พารามิเตอร์
Random Forest	class_weight = balanced n_estimators = 100 max_depth = 5
Gradient Boosting	learning_rate = 0.1 n_estimators = 100 max_depth = 5
SVM	kernel= rbf C=1 gamma=scale
CatBoost	l2_leaf_reg=9 learning_rate=0.1 n_estimators=100 max_depth=5
AdaBoost	n_estimators=100 learning_rate=0.1

4.1 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น DELTA

การแบ่งข้อมูลของหุ้น DELTA สำหรับการสร้างและทดสอบแบบจำลอง แบ่งได้ดังนี้ ชุดข้อมูลสอน 505 วันหรือร้อยละ 80 ของข้อมูลต้นฉบับ ตั้งแต่เดือนพฤษภาคม พ.ศ.2561 ถึง กรกฎาคม พ.ศ. 2565 มีราคาลง 230 วันหรือร้อยละ 45.43 ราคาขึ้น 218 วันหรือร้อยละ 43.05 ราคาไม่มีแนวโน้ม 59 วันหรือร้อยละ 11.50 ส่วนชุดข้อมูลทดสอบ 126 วันหรือร้อยละ 20 ของข้อมูลต้นฉบับ ตั้งแต่เดือนสิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 มีราคาลง 57 วันหรือร้อยละ 44.88 ราคาขึ้น 55 วันหรือร้อยละ 43.30 ราคาไม่มีแนวโน้ม 14 วันหรือร้อยละ 11.81 ดังตารางที่ 4.3

ตารางที่ 4.3 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น DELTA

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	ราคาลง (วัน)	ราคาขึ้น (วัน)	ราคาไม่มีแนวโน้ม (วัน)	รวม (วัน)
ชุดข้อมูลต้นฉบับ (100%)	286 (ร้อยละ 45.32)	271 (ร้อยละ 43.01)	74 (ร้อยละ 11.56)	631
ข้อมูลสอน (80%)	230 (ร้อยละ 45.43)	218 (ร้อยละ 43.05)	59 (ร้อยละ 11.50)	505
ข้อมูลทดสอบ (20%)	57 (ร้อยละ 44.88)	55 (ร้อยละ 43.30)	14 (ร้อยละ 11.81)	126

สำหรับผลของแบบจำลองพบว่า แบบจำลองอดาบูสต์ให้ค่าความแม่นยำชุดข้อมูลสอนร้อยละ 73 และชุดข้อมูลทดสอบร้อยละ 69 ให้ค่าใกล้เคียงกัน (Best Fit) ดังตารางที่ 4.4 ผู้วิจัยจึงเลือกแบบจำลองอดาบูสต์เนื่องจากแบบจำลองสามารถทำงานได้ดีทั้งชุดข้อมูลสอนและข้อมูลทดสอบ และทำการประมาณการหมุน (Cross-Validation) โดยแบ่งข้อมูลเป็น 10 ส่วนเท่าๆ กัน เพื่อประเมินและทดสอบความแม่นยำของแบบจำลองก่อนที่จะนำแบบจำลองไปใช้ทำนายข้อมูลจริง ช่วยเพิ่มความน่าเชื่อถือของผลลัพธ์แบบจำลองจากการข้อมูลหลายๆ กลุ่มแทนการใช้ข้อมูลเพียงกลุ่มเดียวในการประเมิน หลังทำการประมาณการหมุนกับชุดข้อมูลสอนทั้งหมดจะได้ค่าความแม่นยำเฉลี่ยร้อยละ 68.25

ตารางที่ 4.4 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น DELTA

แบบจำลอง	ค่าความแม่นยำบนชุดข้อมูลสอน (ร้อยละ)	ค่าความแม่นยำบนชุดข้อมูลทดสอบ (ร้อยละ)
Random Forest	87	66
Gradient Boosting	100	70
SVM	80	69
CatBoost	80	67
AdaBoost	73	69

ผู้วิจัยได้ทำวิเคราะห์ตัวแปรอิสระที่สำคัญ (Feature Impotence) โดยใช้วิธีการเลือกตัวแปร (Feature Selection) แบบ Embedded จากแบบจำลองอดาบูสต์ เนื่องจากแบบจำลองนี้ถูกพัฒนาจากแนวคิดต้นไม้ตัดสินใจการใช้วิธีการเลือกตัวแปรแบบ Embedded ในการวิเคราะห์ความสำคัญของตัว

แปรในแบบจำลองอดาปตัสต์เป็นวิธีที่ช่วยให้เราสามารถระบุความสำคัญของแต่ละตัวแปรที่มีผลต่อผลลัพธ์ของแบบจำลองได้ แนวคิดของวิธีการเลือกตัวแปรแบบ Embedded นี้เกิดจากการพัฒนาแบบจำลองอดาปตัสต์ที่เป็นการรวมกันระหว่างหลายๆ แบบจำลองย่อย โดยส่วนหนึ่งของแบบจำลองย่อยที่ใช้ในกระบวนการเลือกตัวแปรเรียกว่า Decision Stumps ซึ่งเป็นต้นไม้ตัดสินใจที่มีความลึกเพียงหนึ่งระดับเท่านั้น ในกระบวนการเลือกตัวแปรแบบ Embedded ด้วยวิธีการจากแบบจำลองอดาปตัสต์ค่าความสำคัญของแต่ละตัวแปรจะถูกใช้ในการกำหนดน้ำหนักของตัวอย่างที่ใช้ในการสอนแบบจำลองย่อยในรอบถัดไป หากตัวแปรใดมีค่าความสำคัญสูง จะถูกให้น้ำหนักสูงขึ้น หมายถึงตัวแปรดังกล่าวมีอิทธิพลต่อผลลัพธ์มากกว่าตัวแปรอื่นๆ ในแบบจำลองนั้น ผู้วิจัยทำการเลือกตัวแปรที่สำคัญ 5 อันดับแรกที่มีคะแนนสูงสุดของแบบจำลองกับการพยากรณ์ทิศทางราคาของ DELTA ดังตารางที่ 4.5

ตารางที่ 4.5 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ DELTA

ลำดับ	ตัวแปร	ประเภท	ค่าตัวแปรที่สำคัญ
1	DELTA_SIGNAL_LINE	ปัจจัยเชิงเทคนิค	0.51
2	DELTA_RSI	ปัจจัยเชิงเทคนิค	0.42
3	ROLLING_CORRELATIOH_DJI_NDX	ปัจจัยเชิงเทคนิค	0.42
4	INVESTOR_SENTIMENT	ปัจจัยข่าว	0.42
5	SET50_VOLUME	ปัจจัยเชิงเทคนิค	0.33

ผู้วิจัยทำการทดสอบกับตัวแปรอิสระที่สำคัญ 5 อันดับแรกกับชุดข้อมูลทดสอบ ตั้งแต่เดือนสิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 เพื่อวัดประสิทธิภาพของแบบจำลอง พบว่าการใช้ตัวแปรอิสระที่สำคัญ 3 อันดับแรกให้ค่าความแม่นยำร้อยละ 51.97 ดังตารางที่ 4.6

ตารางที่ 4.6 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันดับแรกในการพยากรณ์ทิศทางของราคาของ DELTA

จำนวนตัวแปรอิสระ	ชื่อตัวแปรอิสระ	ความแม่นยำ (ร้อยละ)
1	DELTA_SIGNAL_LINE	48.82
2	DELTA_SIGNAL_LINE DELTA_RSI	50.39
3	DELTA_SIGNAL_LINE	51.97

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	DELTA_RSI ROLLING_CORRELATIOH_DJI_NDX	
4	DELTA_SIGNAL_LINE DELTA_RSI ROLLING_CORRELATIOH_DJI_NDX INVESTOR_SENTIMENT	47.24
5	DELTA_SIGNAL_LINE DELTA_RSI ROLLING_CORRELATIOH_DJI_NDX INVESTOR_SENTIMENT SET50_VOLUME	45.67

ประสิทธิภาพการทำนายของแบบจำลอง AdaBoost กับชุดข้อมูลทดสอบ

ตารางเมทริกซ์สับสนและประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ DELTA ช่วงเดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 ดังรูปที่ 4.1 และรูปที่ 4.2

สำหรับคลาสราคาสูง (Class -1):

- True Positive (TP): มีข้อมูลทั้งหมด 15 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาสูง (Class -1)
- True Negative (TN): มีข้อมูลทั้งหมด 56 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาสูง (Class -1)
- False Positive (FP): มีข้อมูลทั้งหมด 10 ตัวที่ทำนายผิดว่าเป็นคลาสราคาสูง (Class -1)
- False Negative (FN): มีข้อมูลทั้งหมด 46 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาสูง (Class -1)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาสูง (Class -1) ทั้งหมดร้อยละ 60 เป็นความจริงว่าเป็นคลาสราคาสูง (Class -1)
- Recall: จากข้อมูลที่เป็นคลาสราคาสูง (Class -1) ทั้งหมดร้อยละ 25 ถูกทำนายว่าเป็นคลาสราคาสูง (Class -1) ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 35

สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0):

- True Positive (TP): มีข้อมูลทั้งหมด 1 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- True Negative (TN): มีข้อมูลทั้งหมด 108 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
- False Positive (FP): มีข้อมูลทั้งหมด 5 ตัวที่ทำนายผิดว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- False Negative (FN): มีข้อมูลทั้งหมด 13 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 17 เป็นความจริงว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- Recall: จากข้อมูลที่เป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 7 ถูกทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 10

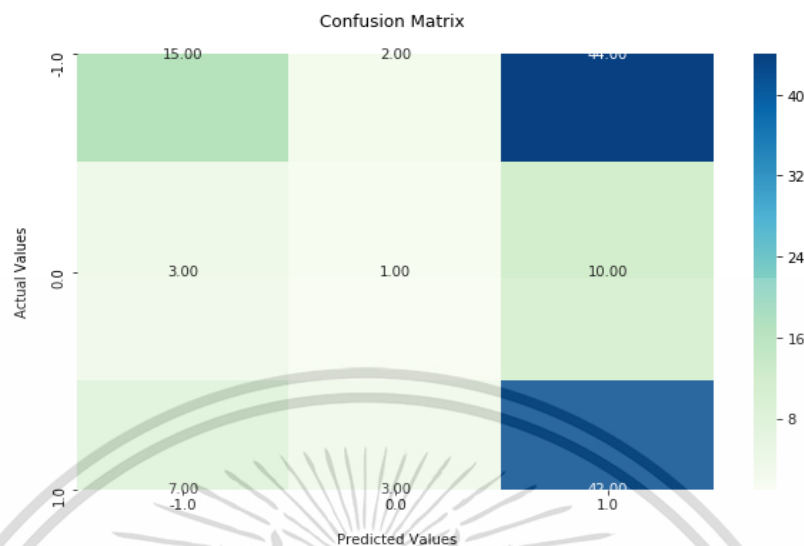
สำหรับคลาสราคาขึ้น (Class 1):

- True Positive (TP): มีข้อมูลทั้งหมด 42 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาขึ้น (Class 1)
- True Negative (TN): มีข้อมูลทั้งหมด 21 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาขึ้น (Class 1)
- False Positive (FP): มีข้อมูลทั้งหมด 54 ตัวที่ทำนายผิดว่าเป็นคลาสราคาขึ้น (Class 1)
- False Negative (FN): มีข้อมูลทั้งหมด 10 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาขึ้น (Class 1)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 44 เป็นความจริงว่าเป็นคลาสราคาขึ้น (Class 1)

- Recall: จากข้อมูลที่เป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 81 ถูกทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 57



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.1 ตารางเมทริกซ์สับสนของการพยากรณ์ราคา DELTA

	precision	recall	f1-score	support
-1.0	0.60	0.25	0.35	61
0.0	0.17	0.07	0.10	14
1.0	0.44	0.81	0.57	52
accuracy			0.46	127
macro avg	0.40	0.38	0.34	127
weighted avg	0.49	0.46	0.41	127

รูปที่ 4.2 ประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ DELTA

การตีความผลลัพธ์จากแบบจำลอง

- ค่า Precision หมายถึงความแม่นยำในการทำนายว่าเป็นกลุ่มที่ต้องการ สำหรับคลาสราคาลง (Class -1) แม่นยำที่สุดก็คือร้อยละ 60 สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0) และคลาสราคาขึ้น (Class 1) ค่า Precision ต่ำกว่าคลาสราคาลง (Class -1)
- ค่า Recall หมายถึงความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง สำหรับคลาสราคาขึ้น (Class 1) มีค่า Recall สูงที่สุดที่ประมาณร้อยละ 80.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ค่า F1-score คือค่าความสมดุลระหว่าง Precision และ Recall โดยสำหรับคลาสราคาขึ้น (Class 1) มีค่า F1-score สูงแสดงถึงความแม่นยำและความสามารถในการตรวจจับข้อมูลในคลาสที่เป็นความจริง
- Precision แบบค่าเฉลี่ยมาโคร ให้ความแม่นยำร้อยละ 40.13 ของค่าเฉลี่ย Precision ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Recall แบบค่าเฉลี่ยมาโคร ให้ความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง ร้อยละ 37.50 ของค่าเฉลี่ย Recall ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- F1-score แบบค่าเฉลี่ยมาโคร ให้ค่าความสมดุลระหว่าง Precision และ Recall ร้อยละ 33.88 ของค่าเฉลี่ย F1-score ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Precision แบบค่าเฉลี่ยถ่วงน้ำหนัก คือความแม่นยำร้อยละ 48.56 ของค่าเฉลี่ย Precision ของทุกกลุ่มโดยคิดรวมกับจำนวนข้อมูลในแต่ละกลุ่ม
- Recall แบบค่าเฉลี่ยถ่วงน้ำหนักคือความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริงร้อยละ 45.66 ของค่าเฉลี่ย Recall ของทุกกลุ่มโดยคิดรวมกับจำนวนข้อมูลในแต่ละกลุ่ม
- F1-score แบบค่าเฉลี่ยถ่วงน้ำหนัก ให้ค่าความสมดุลระหว่าง Precision และ Recall ร้อยละ 41.09 ของค่าเฉลี่ย F1-score ของทุกกลุ่มโดยคิดรวมกับจำนวนข้อมูลในแต่ละกลุ่ม

4.2 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น AOT

การแบ่งข้อมูลของหุ้น AOT สำหรับการสร้างและทดสอบแบบจำลอง แบ่งได้ดังนี้ ชุดข้อมูลสอน 505 วันหรือร้อยละ 80 ของข้อมูลต้นฉบับ ตั้งแต่เดือนพฤษภาคม พ.ศ.2561 ถึง กรกฎาคม พ.ศ.2565 มีราคาดลง 222 วันหรือร้อยละ 43.84 ราคาขึ้น 204 วันหรือร้อยละ 40.27 ราคาไม่มีแนวโน้ม 81 วันหรือร้อยละ 15.87 ส่วนชุดข้อมูลทดสอบ 126 วันหรือร้อยละ 20 ของข้อมูลต้นฉบับ ตั้งแต่เดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 มีราคาดลง 56 วันหรือร้อยละ 44.09 ราคาขึ้น 51 วันหรือร้อยละ 40.15 ราคาไม่มีแนวโน้ม 20 วันหรือร้อยละ 15.74 ดังตารางที่ 4.7

ตารางที่ 4.7 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น AOT

	ราคาดลง (วัน)	ราคาขึ้น (วัน)	ราคาไม่มีแนวโน้ม (วัน)	รวม (วัน)
ชุดข้อมูลต้นฉบับ (100%)	277 (ร้อยละ 43.89)	254 (ร้อยละ 40.25)	100 (ร้อยละ 15.84)	631
ข้อมูลสอน (80%)	222 (ร้อยละ 43.84)	204 (ร้อยละ 40.27)	81 (ร้อยละ 15.87)	505
ข้อมูลทดสอบ (20%)	56 (ร้อยละ 44.09)	51 (ร้อยละ 40.15)	20 (ร้อยละ 15.74)	126

สำหรับผลของแบบจำลองพบว่า แบบจำลองอดาบูสต์ให้ค่าความแม่นยำชุดข้อมูลสอนร้อยละ 69 และชุดข้อมูลทดสอบร้อยละ 69 ให้ค่าใกล้เคียงกัน ดังตารางที่ 4.8 ผู้วิจัยจึงเลือกแบบจำลอง AdaBoost ในการทำ Cross-Validation โดยแบ่งข้อมูลเป็น 10 ส่วนเท่าๆ กัน เพื่อสร้างและทดสอบแบบจำลอง ก่อนที่จะนำแบบจำลองไปใช้ทำนายข้อมูล แบบจำลองนี้เมื่อทดสอบกับชุดข้อมูลสอนทั้งหมดจะได้ค่าความแม่นยำเฉลี่ยร้อยละผู้วิจัยจึงเลือกแบบจำลองอดาบูสต์ เนื่องจากแบบจำลองสามารถทำงานได้ดีทั้งชุดข้อมูลสอนและชุดข้อมูลทดสอบ และทำการประมาณการหมุน โดยแบ่งข้อมูลเป็น 10 ส่วนเท่าๆ กัน เพื่อประเมินและทดสอบความแม่นยำของแบบจำลองก่อนที่จะนำแบบจำลองไปใช้ทำนายข้อมูลจริง ช่วยเพิ่มความน่าเชื่อถือของผลลัพธ์แบบจำลองจากการข้อมูลหลายๆ กลุ่มแทนการใช้ข้อมูลเพียงกลุ่มเดียวในการประเมิน หลังทำการประมาณการหมุนกับชุดข้อมูลสอนทั้งหมดจะได้ค่าความแม่นยำเฉลี่ยร้อยละ 65

ตารางที่ 4.8 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น AOT

แบบจำลอง	ค่าความแม่นยำบนชุดข้อมูลสอน (ร้อยละ)	ค่าความแม่นยำบนชุดข้อมูลทดสอบ (ร้อยละ)
Random Forest	91	66
Gradient Boosting	100	63
SVM	76	69
CatBoost	75	69
AdaBoost	69	69

ผู้วิจัยได้ทำวิเคราะห์ตัวแปรอิสระที่สำคัญโดยใช้วิธีการเลือกตัวแปรแบบ Embedded จากแบบจำลองอดาบูสต์ เนื่องจากแบบจำลองนี้ถูกพัฒนาจากแนวคิดต้นไม้ตัดสินใจ การใช้วิธีการเลือกตัวแปรแบบ Embedded ในการวิเคราะห์ความสำคัญของตัวแปรในแบบจำลองอดาบูสต์เป็นวิธีที่ช่วยให้เราสามารถระบุความสำคัญของแต่ละตัวแปรที่มีผลต่อผลลัพธ์ของแบบจำลองได้ แนวคิดของวิธีการเลือกตัวแปรแบบ Embedded นี้เกิดจากการพัฒนาแบบจำลองอดาบูสต์ที่เป็นการรวมกันระหว่างหลายๆแบบจำลองย่อย โดยส่วนหนึ่งของแบบจำลองย่อยที่ใช้ในกระบวนการเลือกตัวแปรเรียกว่า Decision Stumps ซึ่งเป็นต้นไม้ตัดสินใจที่มีความลึกเพียงหนึ่งระดับเท่านั้น ในกระบวนการเลือกตัวแปรแบบ Embedded ด้วยวิธีการจากแบบจำลองอดาบูสต์ค่าความสำคัญของแต่ละตัวแปรจะถูกใช้ในการกำหนดน้ำหนักของตัวอย่างที่ใช้ในการสอนแบบจำลองย่อยในรอบถัดไป หากตัวแปรใดมีค่าความสำคัญสูง จะถูกให้น้ำหนักสูงขึ้น หมายถึงตัวแปรดังกล่าวมีอิทธิพลต่อผลลัพธ์มากกว่าตัวแปรอื่นๆ ในแบบจำลองนั้น ผู้วิจัยทำการเลือกตัวแปรที่สำคัญ 5 อันดับแรกที่มีคะแนนสูงสุดของแบบจำลองกับการพยากรณ์ทิศทางราคาของ AOT ดังตารางที่ 4.9

ตารางที่ 4.9 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ AOT

ลำดับ	ตัวแปร	ประเภท	ค่าตัวแปรที่สำคัญ
1	TRANS_INDEX	ปัจจัยเชิงเทคนิค	59.76
2	SET_PRICE_DIRECTION	ปัจจัยเชิงเทคนิค	1.73
3	AOT_VOLUME	ปัจจัยเชิงเทคนิค	0.93
4	COMPANY_EVENT	ปัจจัยข่าว	0.74
5	ICT_VOLUME	ปัจจัยเชิงเทคนิค	0.62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้วิจัยทำการทดสอบกับตัวแปรอิสระที่สำคัญ 5 อันดับแรกกับชุดข้อมูลทดสอบ ตั้งแต่เดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 เพื่อวัดประสิทธิภาพของแบบจำลอง พบว่าการใช้ตัวแปร อิสระที่สำคัญ 2 อันดับแรกให้ค่าความแม่นยำร้อยละ 65.4 ดังตารางที่ 4.10

ตารางที่ 4.10 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันดับแรกในการพยากรณ์ทิศทาง ของราคาของ AOT

จำนวนตัวแปรอิสระ	ชื่อตัวแปรอิสระ	ความแม่นยำ (ร้อยละ)
1	TRANS_INDEX	63.5
2	TRANS_INDEX SET_PRICE_DIRECTION	65.4
3	TRANS_INDEX SET_PRICE_DIRECTION AOT_VOLUME	62.6
4	TRANS_INDEX SET_PRICE_DIRECTION AOT_VOLUME COMPANY_EVENT	62.6
5	TRANS_INDEX SET_PRICE_DIRECTION AOT_VOLUME COMPANY_EVENT ICT_VOLUME	62.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประสิทธิภาพการทำนายของแบบจำลอง AdaBoost กับชุดข้อมูลทดสอบ

ตารางเมทริกซ์สับสนและประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ AOT ช่วงเดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 ดังรูปที่ 4.3 และรูปที่ 4.4

สำหรับคลาสราคาลง (Class -1):

- True Positive (TP): มีข้อมูลทั้งหมด 47 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาลง (Class -1)
- True Negative (TN): มีข้อมูลทั้งหมด 43 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาลง (Class -1)
- False Positive (FP): มีข้อมูลทั้งหมด 26 ตัวที่ทำนายผิดว่าเป็นคลาสราคาลง (Class -1)
- False Negative (FN): มีข้อมูลทั้งหมด 11 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาลง (Class -1)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาลง (Class -1) ทั้งหมดร้อยละ 64 เป็นความจริงว่าเป็นคลาสราคาลง (Class -1)
- Recall: จากข้อมูลที่เป็นคลาสราคาลง (Class -1) ทั้งหมดร้อยละ 81 ถูกทำนายว่าเป็นคลาสราคาลง (Class -1) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 72

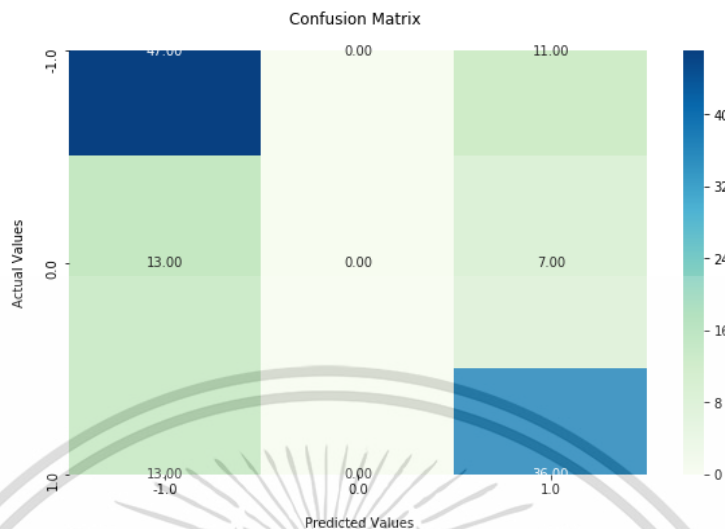
สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0):

- True Positive (TP): มีข้อมูลทั้งหมด 0 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- True Negative (TN): มีข้อมูลทั้งหมด 107 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
- False Positive (FP): มีข้อมูลทั้งหมด 0 ตัวที่ทำนายผิดว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- False Negative (FN): มีข้อมูลทั้งหมด 20 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)

- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 0 เป็นความจริงว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- Recall: จากข้อมูลที่เป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 0 ถูกทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 0

สำหรับคลาสราคาขึ้น (Class 1):

- True Positive (TP): มีข้อมูลทั้งหมด 36 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาขึ้น (Class 1)
- True Negative (TN): มีข้อมูลทั้งหมด 60 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาขึ้น (Class 1)
- False Positive (FP): มีข้อมูลทั้งหมด 18 ตัวที่ทำนายผิดว่าเป็นคลาสราคาขึ้น (Class 1)
- False Negative (FN): มีข้อมูลทั้งหมด 13 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาขึ้น (Class 1)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 67 เป็นความจริงว่าเป็นคลาสราคาขึ้น (Class 1)
- Recall: จากข้อมูลที่เป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 73.46 ถูกทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 69.90



รูปที่ 4.3 ตารางเมทริกซ์สับสนของการพยากรณ์ราคา AOT

	precision	recall	f1-score	support
-1.0	0.64	0.81	0.72	58
0.0	0.00	0.00	0.00	20
1.0	0.67	0.73	0.70	49
accuracy			0.65	127
macro avg	0.44	0.52	0.47	127
weighted avg	0.55	0.65	0.60	127

รูปที่ 4.4 ตารางประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ AOT

การตีความผลลัพธ์จากแบบจำลอง

- ค่า Precision หมายถึงความแม่นยำในการทำนายว่าเป็นกลุ่มที่ต้องการ สำหรับคลาสราคาขึ้น (Class 1) แม่นยำที่สุดก็คือร้อยละ 67 สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0) และคลาสราคาลง (Class -1) ค่า Precision ต่ำกว่าคลาสราคาขึ้น (Class 1)
- ค่า Recall หมายถึงความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง สำหรับคลาสราคาลง (Class -1) มีค่า Recall สูงที่สุดที่ประมาณร้อยละ 81
- ค่า F1-score คือค่าความสมดุลระหว่าง Precision และ Recall โดยสำหรับคลาสราคาลง (Class -1) มีค่า F1-score สูงแสดงถึงความแม่นยำและความสามารถในการตรวจจับข้อมูลในคลาสที่เป็นความจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- Precision แบบค่าเฉลี่ยมาโคร ให้ความแม่นยำ ร้อยละ 44.34 ค่าเฉลี่ย Precision ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Recall แบบค่าเฉลี่ยมาโคร ให้ความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง ร้อยละ 51.50 ของค่าเฉลี่ย Recall ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- F1-score แบบค่าเฉลี่ยมาโคร ให้ค่าความสมดุลระหว่าง Precision และ Recall ร้อยละ 47.20 ของ ค่าเฉลี่ย F1-score ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Precision แบบค่าเฉลี่ยถ่วงน้ำหนัก ให้ความแม่นยำ ร้อยละ 55 ค่าเฉลี่ย Precision ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม
- Recall แบบค่าเฉลี่ยถ่วงน้ำหนักคือความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง ร้อยละ 65 ของค่าเฉลี่ย Recall ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม
- F1-score แบบค่าเฉลี่ยถ่วงน้ำหนัก ให้ค่าความสมดุลระหว่าง Precision และ Recall ร้อยละ 60 ของค่าเฉลี่ย F1-score ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม

4.3 ผลการวิจัยการพยากรณ์ทิศทางารเปลี่ยนแปลงราคาหุ้น PTT

การแบ่งข้อมูลของหุ้น PTT สำหรับการสร้างและทดสอบแบบจำลอง แบ่งได้ดังนี้ ชุดข้อมูลสอน 505 วันหรือร้อยละ 80 ของข้อมูลต้นฉบับ ตั้งแต่เดือนพฤษภาคม พ.ศ.2561 ถึง กรกฎาคม พ.ศ.2565 มีราคาราลง 211 วันหรือร้อยละ 41.66 ราคาขึ้น 162 วันหรือร้อยละ 31.94 ราคาไม่มีแนวโน้ม 134 วันหรือร้อยละ 26.38 ส่วนชุดข้อมูลทดสอบ 126 วันหรือร้อยละ 20 ของข้อมูลต้นฉบับ ตั้งแต่เดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 มีราคาราลง 53 วันหรือร้อยละ 41.73 ราคาขึ้น 40 วันหรือร้อยละ 31.49 ราคาไม่มีแนวโน้ม 34 วันหรือร้อยละ 26.77 ดังตารางที่ 4.11

ตารางที่ 4.11 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น PTT

	ราคาราลง (วัน)	ราคาขึ้น (วัน)	ราคาไม่มีแนวโน้ม (วัน)	รวม (วัน)
ชุดข้อมูลต้นฉบับ (100%)	263 (ร้อยละ 41.67)	201 (ร้อยละ 31.85)	167 (ร้อยละ 26.46)	631

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ข้อมูลสอน (80%)	211 (ร้อยละ 41.66)	162 (ร้อยละ 31.94)	134 (ร้อยละ 26.38)	505
ข้อมูลทดสอบ (20%)	53 (ร้อยละ 41.73)	40 (ร้อยละ 31.49)	34 (ร้อยละ 26.77)	126

สำหรับผลของแบบจำลองพบว่า แบบจำลองอดาบустีให้ค่าความแม่นยำชุดข้อมูลสอนร้อยละ 68 และชุดข้อมูลทดสอบร้อยละ 64 ให้ค่าใกล้เคียงกัน ดังตารางที่ 4.12 ผู้วิจัยจึงเลือกแบบจำลองอดาบустี เนื่องจากแบบจำลองสามารถทำงานได้ดีทั้งชุดข้อมูลสอนและข้อมูลทดสอบ และทำการประมาณการหมุน โดยแบ่งข้อมูลเป็น 10 ส่วนเท่าๆ กัน เพื่อประเมินและทดสอบความแม่นยำของแบบจำลองก่อนที่จะนำแบบจำลองไปใช้ทำนายข้อมูลจริง ช่วยเพิ่มความน่าเชื่อถือของผลลัพธ์แบบจำลองจากการข้อมูลหลายๆ กลุ่มแทนการใช้ข้อมูลเพียงกลุ่มเดียวในการประเมิน หลังทำการประมาณการหมุนกับชุดข้อมูลสอนทั้งหมดจะได้ค่าความแม่นยำเฉลี่ยร้อยละ 53

ตารางที่ 4.12 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น PTT

แบบจำลอง	ค่าความแม่นยำบนชุดข้อมูล สอน (ร้อยละ)	ค่าความแม่นยำบนชุดข้อมูลทดสอบ (ร้อยละ)
Random Forest	87	52
Gradient Boosting	100	52
SVM	75	57
CatBoost	80	55
AdaBoost	68	64

ผู้วิจัยได้ทำวิเคราะห์ตัวแปรอิสระที่สำคัญโดยใช้วิธีการเลือกตัวแปรแบบ Embedded จากแบบจำลองอดาบู้สต์ เนื่องจากแบบจำลองนี้ถูกพัฒนาจากแนวคิดต้นไม้ตัดสินใจ การใช้วิธีการเลือกตัวแปรแบบ Embedded ในการวิเคราะห์ความสำคัญของตัวแปรในแบบจำลองอดาบู้สต์เป็นวิธีที่ช่วยให้เราสามารถระบุความสำคัญของแต่ละตัวแปรที่มีผลต่อผลลัพธ์ของแบบจำลองได้ แนวคิดของวิธีการเลือกตัวแปรแบบ Embedded นี้เกิดจากการพัฒนาแบบจำลองอดาบู้สต์ที่เป็นการรวมกันระหว่างหลายๆแบบจำลองย่อย โดยส่วนหนึ่งของแบบจำลองย่อยที่ใช้ในกระบวนการเลือกตัวแปรเรียกว่า Decision Stumps ซึ่งเป็นต้นไม้ตัดสินใจที่มีความลึกเพียงหนึ่งระดับเท่านั้น ในกระบวนการเลือกตัวแปรแบบ Embedded ด้วยวิธีการจากแบบจำลองอดาบู้สต์ค่าความสำคัญของแต่ละตัวแปรจะถูกใช้ในการกำหนดน้ำหนักของตัวอย่างที่ใช้ในการสอนแบบจำลองย่อยในรอบถัดไป หากตัวแปรใดมีค่าความสำคัญสูง จะถูกให้น้ำหนักสูงขึ้น หมายถึงตัวแปรดังกล่าวมีอิทธิพลต่อผลลัพธ์มากกว่าตัวแปรอื่นๆ ในแบบจำลองนั้น ผู้วิจัยทำการเลือกตัวแปรที่สำคัญ 5 อันดับแรกที่มีคะแนนสูงสุดของแบบจำลองกับการพยากรณ์ทิศทางราคาของ PTT ดังตารางที่ 4.13

ตารางที่ 4.13 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ PTT

ลำดับ	ตัวแปร	ประเภท	ค่าตัวแปรที่สำคัญ
1	PTT_VOLUME	ปัจจัยเชิงเทคนิค	0.53
2	FASRION_VOLUME	ปัจจัยเชิงเทคนิค	0.42
3	ROLLING_CORRELATION_DJI_SET	ปัจจัยเชิงเทคนิค	0.33
4	BANK_VOLUME	ปัจจัยเชิงเทคนิค	0.33
5	PTT_RSI	ปัจจัยเชิงเทคนิค	0.33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้วิจัยทำการทดสอบกับตัวแปรอิสระที่สำคัญ 5 อันดับแรกกับชุดข้อมูลทดสอบตั้งแต่เดือนสิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 เพื่อวัดประสิทธิภาพของแบบจำลอง พบว่าการใช้ตัวแปรอิสระที่สำคัญ 5 อันดับแรกให้ค่าความแม่นยำร้อยละ 48.82 ดังตารางที่ 4.14

ตารางที่ 4.14 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันดับแรกในการพยากรณ์ทิศทางของราคาของ PTT

จำนวนตัวแปรอิสระ	ชื่อตัวแปรอิสระ	ความแม่นยำ (ร้อยละ)
1	PTT_VOLUME	39.37
2	PTT_VOLUME FASRION_VOLUME	36.22
3	PTT_VOLUME FASRION_VOLUME ROLLING_CORRELATION_DJI_SET	35.43
4	PTT_VOLUME FASRION_VOLUME ROLLING_CORRELATION_DJI_SET BANK_VOLUME	37.01
5	PTT_VOLUME FASRION_VOLUME ROLLING_CORRELATION_DJI_SET BANK_VOLUME PTT_RSI	48.82

ประสิทธิภาพการทำนายของแบบจำลอง AdaBoost กับชุดข้อมูลทดสอบ

ตารางเมตริกซ์สับสนและประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ PTT ช่วงเดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 ดังรูปที่ 4.5 และรูปที่ 4.6

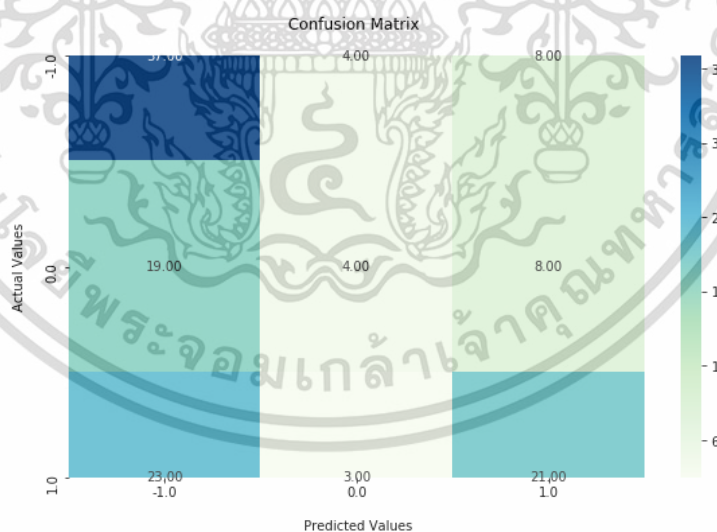
สำหรับคลาสราคา (Class -1):

- True Positive (TP): มีข้อมูลทั้งหมด 37 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคา (Class -1)

- True Negative (TN): มีข้อมูลทั้งหมด 36 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาสูง (Class -1)
 - False Positive (FP): มีข้อมูลทั้งหมด 42 ตัวที่ทำนายผิดว่าเป็นคลาสราคาสูง (Class -1)
 - False Negative (FN): มีข้อมูลทั้งหมด 12 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาสูง (Class -1)
 - Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาสูง (Class -1) ทั้งหมดร้อยละ 47 เป็นความจริงว่าเป็นคลาสราคาสูง (Class -1)
 - Recall: จากข้อมูลที่เป็นคลาสราคาสูง (Class -1) ทั้งหมดร้อยละ 76 ถูกทำนายว่าเป็นคลาสราคาสูง (Class -1) ถูกต้อง
 - F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 58
- สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0):
- True Positive (TP): มีข้อมูลทั้งหมด 4 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
 - True Negative (TN): มีข้อมูลทั้งหมด 89 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
 - False Positive (FP): มีข้อมูลทั้งหมด 7 ตัวที่ทำนายผิดว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
 - False Negative (FN): มีข้อมูลทั้งหมด 27 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
 - Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 36 เป็นความจริงว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
 - Recall: จากข้อมูลที่เป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 13 ถูกทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ถูกต้อง
 - F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 19
- สำหรับคลาสราคาขึ้น (Class 1):

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- True Positive (TP): มีข้อมูลทั้งหมด 21 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาขึ้น (Class 1)
- True Negative (TN): มีข้อมูลทั้งหมด 64 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาขึ้น (Class 1)
- False Positive (FP): มีข้อมูลทั้งหมด 16 ตัวที่ทำนายผิดว่าเป็นคลาสราคาขึ้น (Class 1)
- False Negative (FN): มีข้อมูลทั้งหมด 26 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาขึ้น (Class 1)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 57 เป็นความจริงว่าเป็นคลาสราคาขึ้น (Class 1)
- Recall: จากข้อมูลที่เป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 45 ถูกทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 50



รูปที่ 4.5 ตารางเมทริกซ์สับสนของการพยากรณ์ราคา PTT

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

	precision	recall	f1-score	support
-1.0	0.47	0.76	0.58	49
0.0	0.36	0.13	0.19	31
1.0	0.57	0.45	0.50	47
accuracy			0.49	127
macro avg	0.47	0.44	0.42	127
weighted avg	0.48	0.49	0.45	127

รูปที่ 4.6 ประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ PTT

การตีความผลลัพธ์จากแบบจำลอง

- ค่า Precision หมายถึงความแม่นยำในการทำนายว่าเป็นกลุ่มที่ต้องการ สำหรับคลาสราคาขึ้น (Class 1) แม่นยำที่สุดก็คือร้อยละ 57 สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0) และคลาสราคาลง (Class -1) ค่า Precision ต่ำกว่าคลาสราคาขึ้น (Class 1)
- ค่า Recall หมายถึงความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง สำหรับคลาสราคาลง (Class -1) มีค่า Recall สูงที่สุดที่ประมาณร้อยละ 76
- ค่า F1-score คือค่าความสมดุลระหว่าง Precision และ Recall โดยสำหรับคลาสราคาลง (Class -1) มีค่า F1-score สูงแสดงถึงความแม่นยำและความสามารถในการตรวจจับข้อมูลในคลาสที่เป็นความจริง
- Precision แบบค่าเฉลี่ยมาโคร ให้ความแม่นยำ ร้อยละ 47 ของค่าเฉลี่ย Precision ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Recall แบบค่าเฉลี่ยมาโคร ให้ความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง ร้อยละ 44 ของค่าเฉลี่ย Recall ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- F1-score แบบค่าเฉลี่ยมาโคร ให้ค่าความสมดุลระหว่าง Precision และ Recall ร้อยละ 42 ของค่าเฉลี่ย F1-score ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Precision แบบค่าเฉลี่ยถ่วงน้ำหนัก คือความแม่นยำ ร้อยละ 48 ค่าเฉลี่ย Precision ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม

- Recall แบบค่าเฉลี่ยถ่วงน้ำหนักคือความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง ร้อยละ 49 ของค่าเฉลี่ย Recall ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม
- F1-score แบบค่าเฉลี่ยถ่วงน้ำหนัก ให้ค่าความสมดุลระหว่าง Precision และ Recall ร้อยละ 45 ของค่าเฉลี่ย F1-score ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม

4.4 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น ADVANC

การแบ่งข้อมูลของหุ้น ADVANC สำหรับการสร้างและทดสอบแบบจำลอง แบ่งได้ดังนี้ ชุดข้อมูลสอน 50 วันหรือร้อยละ 80 ของข้อมูลต้นฉบับ ตั้งแต่เดือนพฤษภาคม พ.ศ.2561 ถึง กรกฎาคม พ.ศ.2565 มีราคาลง 230 วันหรือร้อยละ 45.43 ราคาขึ้น 218 วันหรือร้อยละ 43.05 ราคาไม่มีแนวโน้ม 59 วันหรือร้อยละ 11.50 ส่วนชุดข้อมูลทดสอบ 126 วันหรือร้อยละ 20 ของข้อมูลต้นฉบับ ตั้งแต่เดือนสิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 มีราคาลง 57 วันหรือร้อยละ 44.88 ราคาขึ้น 55 วันหรือร้อยละ 43.30 ราคาไม่มีแนวโน้ม 15 วันหรือร้อยละ 11.81 ดังตารางที่ 4.15

ตารางที่ 4.15 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น ADVANC

	ราคาลง (วัน)	ราคาขึ้น (วัน)	ราคาไม่มีแนวโน้ม (วัน)	รวม (วัน)
ชุดข้อมูลต้นฉบับ (100%)	286 (ร้อยละ 45.32)	271 (ร้อยละ 43.01)	74 (ร้อยละ 11.56)	631
ข้อมูลสอน (80%)	230 (ร้อยละ 45.43)	218 (ร้อยละ 43.05)	59 (ร้อยละ 11.50)	505
ข้อมูลทดสอบ (20%)	57 (ร้อยละ 44.88)	55 (ร้อยละ 43.30)	15 (ร้อยละ 11.81)	126

สำหรับผลของแบบจำลองพบว่า แบบจำลองอดาปุสต์ให้ค่าความแม่นยำชุดข้อมูลสอนร้อยละ 71 และชุดข้อมูลทดสอบร้อยละ 67 ให้ค่าใกล้เคียงกัน ดังตารางที่ 4.16 ผู้วิจัยจึงเลือกแบบจำลองอดาปุสต์ เนื่องจากแบบจำลองสามารถทำงานได้ดีทั้งชุดข้อมูลสอนและข้อมูลทดสอบ และทำการประมาณการหมุน โดยแบ่งข้อมูลเป็น 10 ส่วนเท่าๆ กัน เพื่อประเมินและทดสอบความแม่นยำของแบบจำลองก่อนที่จะนำแบบจำลองไปใช้ทำนายข้อมูลจริง ช่วยเพิ่มความน่าเชื่อถือของผลลัพธ์แบบจำลองจากการข้อมูลหลายๆ

กลุ่มแทนการใช้ข้อมูลเพียงกลุ่มเดียวในการประเมิน หลังทำการประมาณการหมุนกับชุดข้อมูลสอน ทั้งหมดจะได้ค่าความแม่นยำเฉลี่ยร้อยละ 57

ตารางที่ 4.16 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น ADVANC

แบบจำลอง	ค่าความแม่นยำบนชุดข้อมูลสอน (ร้อยละ)	ค่าความแม่นยำบนชุดข้อมูลทดสอบ (ร้อยละ)
Random Forest	90	61
Gradient Boosting	100	61
SVM	76	67
CatBoost	77	65
AdaBoost	71	67

ผู้วิจัยได้ทำวิเคราะห์ตัวแปรอิสระที่สำคัญโดยใช้วิธีการเลือกตัวแปร แบบ Embedded จากแบบจำลองอดาบูสต์เนื่องจากแบบจำลองนี้ถูกพัฒนาจากแนวคิดต้นไม้ตัดสินใจการใช้วิธีการเลือกตัวแปรแบบ Embedded ในการวิเคราะห์ความสำคัญของตัวแปรในแบบจำลองอดาบูสต์เป็นวิธีที่ช่วยให้เราสามารถระบุความสำคัญของแต่ละตัวแปรที่มีผลต่อผลลัพธ์ของแบบจำลองได้ แนวคิดของวิธีการเลือกตัวแปรแบบ Embedded นี้เกิดจากการพัฒนาแบบจำลองอดาบูสต์ที่เป็นการรวมกันระหว่างหลายๆแบบจำลองย่อย โดยส่วนหนึ่งของแบบจำลองย่อยที่ใช้ในกระบวนการเลือกตัวแปรเรียกว่า Decision Stumps ซึ่งเป็นต้นไม้ตัดสินใจที่มีความลึกเพียงหนึ่งระดับเท่านั้น ในกระบวนการเลือกตัวแปรแบบ Embedded ด้วยวิธีการจากแบบจำลองอดาบูสต์ค่าความสำคัญของแต่ละตัวแปรจะถูกใช้ในการกำหนดน้ำหนักของตัวอย่างที่ใช้ในการสอนแบบจำลองย่อยในรอบถัดไป หากตัวแปรใดมีความสำคัญสูง จะถูกให้น้ำหนักสูงขึ้น หมายถึงตัวแปรดังกล่าวมีอิทธิพลต่อผลลัพธ์มากกว่าตัวแปรอื่นๆ ในแบบจำลองนั้น ผู้วิจัยทำการเลือกตัวแปรที่สำคัญ 5 อันดับแรกที่มีคะแนนสูงสุดของแบบจำลองกับการพยากรณ์ทิศทางราคาของ ADVANC ดังตารางที่ 4.17

ตารางที่ 4.17 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ ADVANC

ลำดับ	ตัวแปร	ประเภท	ค่าตัวแปรที่สำคัญ
1	ADVANC_VOLUME	ปัจจัยเชิงเทคนิค	0.51
2	AGI_VOLUME	ปัจจัยเชิงเทคนิค	0.51
3	ADVANC_RSI	ปัจจัยเชิงเทคนิค	0.51
4	INDUSTRY_EVENT	ปัจจัยเชิงข่าว	0.42
5	ROLLING_CORREALATION_SET50_PKG	ปัจจัยเชิงเทคนิค	0.30

ผู้วิจัยทำการทดสอบกับตัวแปรอิสระที่สำคัญ 5 อันดับแรกกับชุดข้อมูลทดสอบ ตั้งแต่เดือนสิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 เพื่อวัดประสิทธิภาพของแบบจำลอง พบว่าการใช้ตัวแปรอิสระที่สำคัญ 3 อันดับแรกให้ค่าความแม่นยำร้อยละ 51.97 ดังตารางที่ 4.18

ตารางที่ 4.18 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันดับแรกในการพยากรณ์ทิศทางของราคาของ ADVANC

จำนวนตัวแปรอิสระ	ชื่อตัวแปรอิสระ	ความแม่นยำ (ร้อยละ)
1	ADVANC_VOLUME	41.73
2	ADVANC_VOLUME AGI_VOLUME	40.94
3	ADVANC_VOLUME AGI_VOLUME ADVANC_RSI	51.97
4	ADVANC_VOLUME AGI_VOLUME ADVANC_RSI INDUSTRY_EVENT	49.61
5	ADVANC_VOLUME AGI_VOLUME ADVANC_RSI	48.82

	INDUSTRY_EVENT	
	ROLLING_CORREALATION_SET50_PKG	

ประสิทธิภาพการทำนายของแบบจำลอง AdaBoost กับชุดข้อมูลทดสอบ

ตารางเมตริกซ์สับสนและประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ ADVANC ช่วงเดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 ดังรูปที่ 4.7 และตารางที่ 4.8

สำหรับคลาสราคาลง (Class -1):

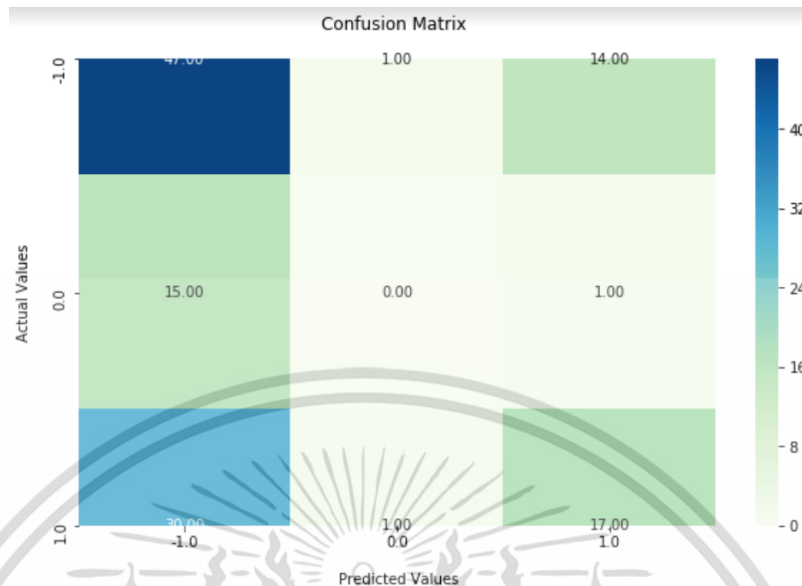
- True Positive (TP): มีข้อมูลทั้งหมด 47 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาลง (Class -1)
- True Negative (TN): มีข้อมูลทั้งหมด 19 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาลง (Class -1)
- False Positive (FP): มีข้อมูลทั้งหมด 45 ตัวที่ทำนายผิดว่าเป็นคลาสราคาลง (Class -1)
- False Negative (FN): มีข้อมูลทั้งหมด 15 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาลง (Class -1)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาลง (Class -1) ทั้งหมดร้อยละ 51 เป็นความจริงว่าเป็นคลาสราคาลง (Class -1)
- Recall: จากข้อมูลที่เป็นคลาสราคาลง (Class -1) ทั้งหมดร้อยละ 76 ถูกทำนายว่าเป็นคลาสราคาลง (Class -1) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 61

สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0):

- True Positive (TP): มีข้อมูลทั้งหมด 0 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- True Negative (TN): มีข้อมูลทั้งหมด 108 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
- False Positive (FP): มีข้อมูลทั้งหมด 2 ตัวที่ทำนายผิดว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- False Negative (FN): มีข้อมูลทั้งหมด 16 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
 - Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 0 เป็นความจริงว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
 - Recall: จากข้อมูลที่เป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 0 ถูกทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ถูกต้อง
 - F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 0
- สำหรับคลาสราคาขึ้น (Class 1):
- True Positive (TP): มีข้อมูลทั้งหมด 17 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาขึ้น (Class 1)
 - True Negative (TN): มีข้อมูลทั้งหมด 63 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาขึ้น (Class 1)
 - False Positive (FP): มีข้อมูลทั้งหมด 15 ตัวที่ทำนายผิดว่าเป็นคลาสราคาขึ้น (Class 1)
 - False Negative (FN): มีข้อมูลทั้งหมด 31 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาขึ้น (Class 1)
 - Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 53 เป็นความจริงว่าเป็นคลาสราคาขึ้น (Class 1)
 - Recall: จากข้อมูลที่เป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 35 ถูกทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ถูกต้อง
 - F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 42



รูปที่ 4.7 ตารางเมทริกซ์สับสนของการพยากรณ์ราคา ADVNAC

	precision	recall	f1-score	support
-1.0	0.51	0.76	0.61	62
0.0	0.00	0.00	0.00	16
1.0	0.53	0.35	0.42	48
accuracy			0.51	126
macro avg	0.35	0.37	0.35	126
weighted avg	0.45	0.51	0.46	126

รูปที่ 4.8 ประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ ADVANC

การตีความผลลัพธ์จากแบบจำลอง

- ค่า Precision หมายถึงความแม่นยำในการทำนายว่าเป็นกลุ่มที่ต้องการ สำหรับคลาสราคาขึ้น (Class 1) แม่นยำที่สุดก็คือร้อยละ 53 สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0) และคลาสราคาลง (Class -1) ค่า Precision ต่ำคลาสราคาขึ้น (Class 1)
- ค่า Recall หมายถึงความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง สำหรับคลาสราคาขึ้น (Class 1) มีค่า Recall สูงที่สุดที่ประมาณร้อยละ 76

- ค่า F1-score คือค่าความสมดุลระหว่าง Precision และ Recall โดยสำหรับคลาสราคา
ลง (Class -1) มีค่า F1-score สูงแสดงถึงความแม่นยำและความสามารถในการตรวจจับ
ข้อมูลในคลาสที่เป็นความจริง
- Precision แบบค่าเฉลี่ยมาโคร ให้ความแม่นยำ ร้อยละ 35 ของค่าเฉลี่ย Precision ของ
ทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Recall แบบค่าเฉลี่ยมาโคร ให้ความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความ
จริง ร้อยละ 37 ของค่าเฉลี่ย Recall ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- F1-score แบบค่าเฉลี่ยมาโคร ให้ค่าความสมดุลระหว่าง Precision และ Recall ร้อย
ละ 35 ของค่าเฉลี่ย F1-score ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Precision แบบค่าเฉลี่ยถ่วงน้ำหนัก คือความแม่นยำ ร้อยละ 45 ของค่าเฉลี่ย Precision
ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม
- Recall แบบค่าเฉลี่ยถ่วงน้ำหนักคือความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็น
ความจริง ร้อยละ 51 ของค่าเฉลี่ย Recall ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่
ละกลุ่ม
- F1-score แบบค่าเฉลี่ยถ่วงน้ำหนัก ให้ค่าความสมดุลระหว่าง Precision และ Recall
ร้อยละ 46 ของค่าเฉลี่ย F1-score ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม

4.5 ผลการวิจัยการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น GULF

การแบ่งข้อมูลของหุ้น DELTA สำหรับการสร้างและทดสอบแบบจำลอง แบ่งได้ดังนี้ ชุดข้อมูล
สอน 505 วันหรือร้อยละ 80 ของข้อมูลต้นฉบับ ตั้งแต่เดือนพฤษภาคม พ.ศ.2561 ถึง กรกฎาคม พ.ศ.
2565 มีราคาลง 292 วันหรือร้อยละ 46.27 ราคาขึ้น 253 วันหรือร้อยละ 40.09 ราคาไม่มีแนวโน้ม 86
วันหรือร้อยละ 13.62 ส่วนชุดข้อมูลทดสอบ 126 วันหรือร้อยละ 20 ของข้อมูลต้นฉบับ ตั้งแต่เดือน
สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 มีราคาลง 59 วันหรือร้อยละ 46.45 ราคาขึ้น 51 วันหรือร้อย
ละ 40.15 ราคาไม่มีแนวโน้ม 17 วันหรือร้อยละ 13.38 ดังตารางที่ 4.19

ตารางที่ 4.19 ชุดข้อมูลในการสร้างแบบจำลองของหุ้น GULF

	ราคาลง (วัน)	ราคาขึ้น (วัน)	ราคาไม่มีแนวโน้ม (วัน)	รวม (วัน)
ชุดข้อมูลต้นฉบับ (100%)	292 (ร้อยละ 46.27)	253 (ร้อยละ 40.09)	86 (ร้อยละ 13.62)	631
ข้อมูลสอน (80%)	234 (ร้อยละ 46.23)	203 (ร้อยละ 40.07)	70 (ร้อยละ 13.69)	505
ข้อมูลทดสอบ (20%)	59 (ร้อยละ 46.45)	51 (ร้อยละ 40.15)	17 (ร้อยละ 13.38)	126

สำหรับผลของแบบจำลองพบว่า แบบจำลองอดาบуст์ให้ค่าความแม่นยำชุดข้อมูลสอนร้อยละ 65 และชุดข้อมูลทดสอบร้อยละ 59 ให้ค่าใกล้เคียงกัน ดังตารางที่ 4.20 ผู้วิจัยจึงเลือกแบบจำลองอดาบуст์ เนื่องจากแบบจำลองสามารถทำงานได้ดีทั้งชุดข้อมูลสอนและข้อมูลทดสอบ และทำการประมาณการหมุน โดยแบ่งข้อมูลเป็น 10 ส่วนเท่าๆ กัน เพื่อประเมินและทดสอบความแม่นยำของแบบจำลองก่อนที่จะนำแบบจำลองไปใช้ทำนายข้อมูลจริง ช่วยเพิ่มความน่าเชื่อถือของผลลัพธ์แบบจำลองจากการข้อมูลหลายๆ กลุ่มแทนการใช้ข้อมูลเพียงกลุ่มเดียวในการประเมิน หลังทำการประมาณการหมุนกับชุดข้อมูลสอนทั้งหมดจะได้ค่าความแม่นยำเฉลี่ยร้อยละ 57

ตารางที่ 4.20 ผลของแบบจำลองสำหรับการพยากรณ์ทิศทางการเปลี่ยนแปลงราคาหุ้น GULF

แบบจำลอง	ค่าความแม่นยำบนชุดข้อมูล สอน (ร้อยละ)	ค่าความแม่นยำบนชุดข้อมูลทดสอบ (ร้อยละ)
Random Forest	87	54
Gradient Boosting	100	50
SVM	73	57
CatBoost	77	57
AdaBoost	65	59

ผู้วิจัยได้ทำวิเคราะห์ตัวแปรอิสระที่สำคัญ โดยใช้วิธีการเลือกตัวแปรแบบ Embedded จากแบบจำลองอดาบู้สต์ เนื่องจากแบบจำลองนี้ถูกพัฒนาจากแนวคิดต้นไม้ตัดสินใจ การใช้วิธีการเลือกตัวแปรแบบ Embedded ในการวิเคราะห์ความสำคัญของตัวแปรในแบบจำลองอดาบู้สต์เป็นวิธีที่ช่วยให้เราสามารถระบุความสำคัญของแต่ละตัวแปรที่มีผลต่อผลลัพธ์ของแบบจำลองได้ แนวคิดของวิธีการเลือกตัวแปรแบบ Embedded นี้เกิดจากการพัฒนาแบบจำลองอดาบู้สต์ที่เป็นการรวมกันระหว่างหลายๆแบบจำลองย่อย โดยส่วนหนึ่งของแบบจำลองย่อยที่ใช้ในกระบวนการเลือกตัวแปรเรียกว่า Decision Stumps ซึ่งเป็นต้นไม้ตัดสินใจที่มีความลึกเพียงหนึ่งระดับเท่านั้น ในกระบวนการเลือกตัวแปรแบบ Embedded ด้วยวิธีการจากแบบจำลองอดาบู้สต์ ค่าความสำคัญของแต่ละตัวแปรจะถูกใช้ในการกำหนดน้ำหนักของตัวอย่างที่ใช้ในการสอนแบบจำลองย่อยในรอบถัดไป หากตัวแปรใดมีค่าความสำคัญสูง จะถูกให้น้ำหนักสูงขึ้น หมายถึงตัวแปรดังกล่าวมีอิทธิพลต่อผลลัพธ์มากกว่าตัวแปรอื่นๆ ในแบบจำลองนั้น ผู้วิจัยทำการเลือกตัวแปรที่สำคัญ 5 อันดับแรกที่มีคะแนนสูงสุดของแบบจำลองกับการพยากรณ์ทิศทางราคาของ GULF ดังตารางที่ 4.21

ตารางที่ 4.21 ตัวแปรอิสระที่สำคัญพยากรณ์ทิศทางของราคาของ GULF

ลำดับ	ตัวแปร	ประเภท
1	GULF_VOLUME	ปัจจัยเชิงเทคนิค
2	POLITICAL_EVENT	ปัจจัยเชิงข่าว
3	GULF_RSI	ปัจจัยเชิงเทคนิค
4	FASHION_VOLUME	ปัจจัยเชิงเทคนิค
5	ROLLING_CORRELATION_SET50_TOURISM	ปัจจัยเชิงเทคนิค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผู้วิจัยทำการทดสอบกับตัวแปรอิสระที่สำคัญ 5 อันดับแรกกับชุดข้อมูลทดสอบ ตั้งแต่เดือนสิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 เพื่อวัดประสิทธิภาพของแบบจำลอง พบว่าการใช้ตัวแปรอิสระที่สำคัญ 3 อันดับแรกให้ค่าความแม่นยำร้อยละ 55.12 ดังตารางที่ 4.22

ตารางที่ 4.22 การเปรียบเทียบประสิทธิภาพตัวแปรอิสระที่สำคัญ 5 อันดับแรกในการพยากรณ์ทิศทางของราคาของ GULF

จำนวนตัวแปรอิสระ	ชื่อตัวแปรอิสระ	ความแม่นยำ (ร้อยละ)
1	GULF_VOLUME	44.88
2	GULF_VOLUME POLITICAL_EVENT	47.24
3	GULF_VOLUME POLITICAL_EVENT GULF_RSI	55.12
4	GULF_VOLUME POLITICAL_EVENT GULF_RSI FASHION_VOLUME	53.54
5	GULF_VOLUME POLITICAL_EVENT GULF_RSI FASHION_VOLUME ROLLING_CORRELATION_SET50_T OURISM	52.76

ประสิทธิภาพการทำนายของแบบจำลอง AdaBoost กับชุดข้อมูลทดสอบ

ตารางเมตริกซ์สับสนและประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ GULF ช่วงเดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566 ดังรูปที่ 4.9 และรูปที่ 4.10

สำหรับคลาสราคาลง (Class -1):

- True Positive (TP): มีข้อมูลทั้งหมด 46 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาลง (Class -1)
- True Negative (TN): มีข้อมูลทั้งหมด 30 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาลง (Class -1)
- False Positive (FP): มีข้อมูลทั้งหมด 37 ตัวที่ทำนายผิดว่าเป็นคลาสราคาลง (Class -1)
- False Negative (FN): มีข้อมูลทั้งหมด 14 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาลง (Class -1)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาลง (Class -1) ทั้งหมดร้อยละ 55 เป็นความจริงว่าเป็นคลาสราคาลง (Class -1)
- Recall: จากข้อมูลที่เป็นคลาสราคาลง (Class -1) ทั้งหมดร้อยละ 77 ถูกทำนายว่าเป็นคลาสราคาลง (Class -1) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 64

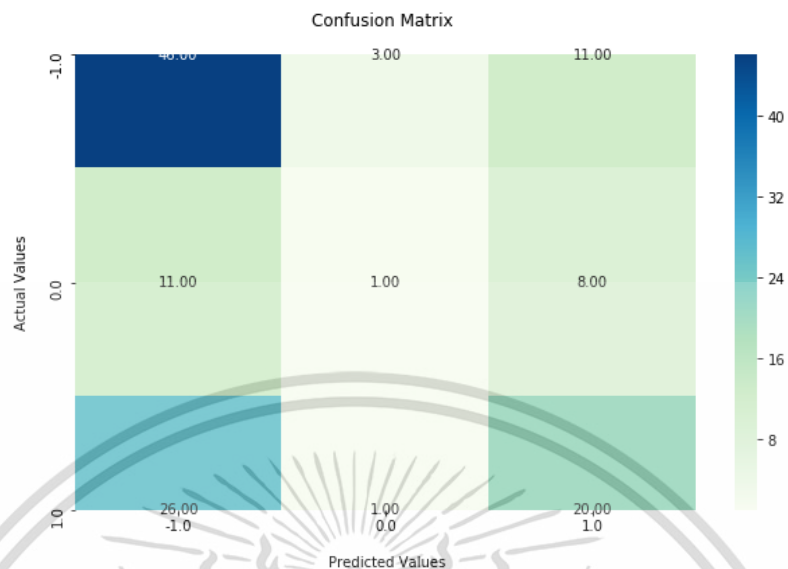
สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0):

- True Positive (TP): มีข้อมูลทั้งหมด 1 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- True Negative (TN): มีข้อมูลทั้งหมด 103 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
- False Positive (FP): มีข้อมูลทั้งหมด 4 ตัวที่ทำนายผิดว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)
- False Negative (FN): มีข้อมูลทั้งหมด 19 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาไม่มีแนวโน้ม (Class 0)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0) ทั้งหมดร้อยละ 20 เป็นความจริงว่าเป็นคลาสราคาไม่มีแนวโน้ม (Class 0)

- Recall: จากข้อมูลที่เป็นคลาสราคาไม่มีแวนโน่ (Class 0) ทั้งหมดร้อยละ 5 ถูกทำนายว่าเป็นคลาสราคาไม่มีแวนโน่ (Class 0) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 8

สำหรับคลาสราคาขึ้น (Class 1):

- True Positive (TP): มีข้อมูลทั้งหมด 20 ตัวที่ทำนายถูกต้องว่าเป็นคลาสราคาขึ้น (Class 1)
- True Negative (TN): มีข้อมูลทั้งหมด 61 ตัวที่ทำนายถูกต้องว่าไม่ใช่คลาสราคาขึ้น (Class 1)
- False Positive (FP): มีข้อมูลทั้งหมด 19 ตัวที่ทำนายผิดว่าเป็นคลาสราคาขึ้น (Class 1)
- False Negative (FN): มีข้อมูลทั้งหมด 27 ตัวที่ทำนายผิดว่าไม่ใช่คลาสราคาขึ้น (Class 1)
- Precision: จากข้อมูลที่ทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 51 เป็นความจริงว่าเป็นคลาสราคาขึ้น (Class 1)
- Recall: จากข้อมูลที่เป็นคลาสราคาขึ้น (Class 1) ทั้งหมดร้อยละ 43 ถูกทำนายว่าเป็นคลาสราคาขึ้น (Class 1) ถูกต้อง
- F1-score: ค่าความสมดุลระหว่าง Precision และ Recall คือร้อยละ 47



รูปที่ 4.9 ตารางเมทริกซ์สับสนของการพยากรณ์ทิศทางราคา GULF

	precision	recall	f1-score	support
-1.0	0.55	0.77	0.64	60
0.0	0.20	0.05	0.08	20
1.0	0.51	0.43	0.47	47
accuracy			0.53	127
macro avg	0.42	0.41	0.40	127
weighted avg	0.48	0.53	0.49	127

รูปที่ 4.10 ประสิทธิภาพของแบบจำลองสำหรับการพยากรณ์ทิศทางของราคาของ GULF

การตีความผลลัพธ์จากแบบจำลอง

- ค่า Precision หมายถึงความแม่นยำในการทำนายว่าเป็นกลุ่มที่ต้องการ สำหรับคลาสราคาลง (Class -1) แม่นยำที่สุดก็คือร้อยละ 55 สำหรับคลาสราคาไม่มีแนวโน้ม (Class 0) และคลาสราคาขึ้น (Class 1) ค่า Precision ต่ำกว่าคลาสราคาลง (Class -1)
- ค่า Recall หมายถึงความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความจริง สำหรับคลาสราคาขึ้น (Class 1) มีค่า Recall สูงที่สุดที่ประมาณร้อยละ 77

- ค่า F1-score คือค่าความสมดุลระหว่าง Precision และ Recall โดยสำหรับคลาสราคา
ลง (Class -1) มีค่า F1-score สูงแสดงถึงความแม่นยำและความสามารถในการตรวจจับ
ข้อมูลในคลาสที่เป็นความจริง
- Precision แบบค่าเฉลี่ยมาโคร ให้ความแม่นยำ ร้อยละ 42 ของค่าเฉลี่ย Precision ของ
ทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Recall แบบค่าเฉลี่ยมาโคร ให้ความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็นความ
จริง ร้อยละ 41 ของค่าเฉลี่ย Recall ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- F1-score แบบค่าเฉลี่ยมาโคร ให้ค่าความสมดุลระหว่าง Precision และ Recall ร้อย
ละ 40 ของค่าเฉลี่ย F1-score ของทุกกลุ่มเทียบกับกลุ่มทั้งหมด
- Precision แบบค่าเฉลี่ยถ่วงน้ำหนัก คือความแม่นยำ ร้อยละ 48 ของค่าเฉลี่ย Precision
ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม
- Recall แบบค่าเฉลี่ยถ่วงน้ำหนักคือความสามารถในการตรวจจับข้อมูลในกลุ่มที่เป็น
ความจริง ร้อยละ 53 ของค่าเฉลี่ย Recall ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่
ละกลุ่ม
- F1-score แบบค่าเฉลี่ยถ่วงน้ำหนัก ให้ค่าความสมดุลระหว่าง Precision และ Recall
ร้อยละ 49 ของค่าเฉลี่ย F1-score ของทุกกลุ่มโดยคิดร่วมกับจำนวนข้อมูลในแต่ละกลุ่ม

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยการทำนายทิศทางราคาหุ้นในกลุ่ม SET50 โดยใช้ข่าวและการวิเคราะห์เชิงเทคนิค ผู้วิจัยได้ทำการสร้างแบบจำลองและทดสอบกับหุ้น SET50 โดยได้ผลลัพธ์ดังนี้

5.1 สรุปผลการดำเนินงาน

จากตารางที่ 5.1 ผลการทดลองทั้ง 5 แบบจำลองกับหุ้นในกลุ่ม SET50 พบว่าแบบจำลอง ป่าไม้ สุ่ม เกรเดียนท์บูสต์ติ่ง และแคสบูสต์ พบปัญหาการเรียนรู้ที่มากเกินไป โดยแบบจำลองอดาบูสต์ สามารถพยากรณ์ทิศทางราคาได้ดีกับราคาหุ้นในกลุ่ม SET50 ทั้งหมด 41 หลักทรัพย์ และชีพพอดต์ตเวกเตอร์แมทซึนสามารถพยากรณ์ทิศทางราคาได้ดีกับราคาหุ้นในกลุ่ม SET50 ทั้งหมด 9 หลักทรัพย์ ซึ่งทั้ง 2 แบบจำลองให้ค่าความแม่นยำที่เหมาะสมทั้งชุดข้อมูลสอนและชุดข้อมูลทดสอบ โดยสรุปแบบจำลองอดาบูสต์มุ่งเน้นไปที่การปรับน้ำหนักของชุดข้อมูลระหว่างการฝึกอบรม แบบจำลองชีพพอดต์ตเวกเตอร์แมทซึนเป็นการหาเส้นไฮเปอร์เพลนที่เหมาะสมกับชุดข้อมูลในการจำแนกประเภท ส่วนแบบจำลองป่าไม้สุ่ม เป็นการรวมต้นไม้ตัดสินใจแล้วทำการโหวตคะแนนให้เสียงส่วนใหญ่หรือการหาค่าเฉลี่ย และแบบจำลองเกรเดียนท์บูสต์ติ่งจะเพิ่มตัวแบบการเรียนรู้ใหม่จากลำดับก่อนหน้าที่มีการปรับปรุงข้อผิดพลาดก่อนหน้า แบบจำลองแคสบูสต์เพิ่มความสามารถในการจัดการข้อมูลเชิงคุณภาพ แต่ละแบบจำลองมีจุดแข็งและจุดอ่อน ควรมีการพิจารณาปัญหาเกี่ยวกับแบบจำลองและลักษณะของชุดข้อมูล เพื่อให้ประสิทธิภาพของแบบจำลองดีขึ้น

ตารางที่ 5.1 ประสิทธิภาพแบบจำลองกับหุ้นในกลุ่ม SET50

แบบจำลอง	จำนวนหลักทรัพย์
Random Forest	-
Gradient Boosting	-
SVM	9 (ร้อยละ 18)
CatBoost	-
AdaBoost	41 (ร้อยละ 82)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 5.2 ตัวแปรอิสระที่สำคัญอันดับที่ 1 ของการพยากรณ์ทิศทาง การเปลี่ยนแปลงราคาหุ้นในกลุ่ม SET50 ช่วงเดือน สิงหาคม พ.ศ.2565 ถึง มกราคม พ.ศ.2566

- อันดับที่ 1 คือ SECTOR_INDEX หรือทิศทาง การเปลี่ยนแปลงของดัชนีกลุ่มอุตสาหกรรมในอัตราส่วนร้อยละ 50 โดยดัชนีกลุ่ม มีผลกับหุ้นรายตัว เนื่องจากการเคลื่อนไหวของดัชนีกลุ่มอุตสาหกรรมนั้นสะท้อนถึงกระแสเงินทุนเคลื่อนย้าย (Fund Flow) ทั้งสถาบันในประเทศ นักลงทุนต่างประเทศ กลุ่มนักลงทุนบัญชีบริษัทหลักทรัพย์ นักลงทุนในประเทศ ที่กำลังให้ความสนใจในอุตสาหกรรมนั้น
- อันดับที่ 2 คือ INDIVIDUAL_STOCK_VOLUME หรือปริมาณการซื้อขายของหุ้นรายตัวร้อยละ 22 จากที่มีกระแสเงินทุนเคลื่อนย้ายเข้ามาซื้อขายในอุตสาหกรรมต่างๆ ทำให้หุ้นรายตัวเกิดปริมาณซื้อขายที่มากขึ้นจากการที่กลุ่มนักลงทุนให้ความสนใจ
- อันดับที่ 3 คือ INDIVIDUAL_RSI หรือเครื่องมือทางเทคนิคสำหรับการวัดการเคลื่อนไหวของราคาว่ามีภาวะการซื้อที่มากเกินไปหรือการขายที่มากเกินไป เพื่อช่วยในการวางกลยุทธ์ในการซื้อขาย เมื่อเกิดภาวะซื้อที่มากเกินไป ค่า RSI มากกว่า 70 จะทำให้ราคาเกิดการเคลื่อนไหวในทิศทางขึ้นและมีแนวโน้มที่จะเกิดแรงขายจนทำให้ราคาเคลื่อนไหวในทิศทางลง ในทางตรงกันข้ามภาวะขายที่มากเกินไป ค่า RSI น้อยกว่า 30 จะทำให้ราคาเกิดการเคลื่อนไหวในทิศทางลงและมีแนวโน้มที่จะเกิดแรงซื้อทำให้ราคาเกิดการเคลื่อนไหวในทิศทางขึ้น
- อันดับที่ 4 คือ SET_INDEX หรือดัชนี SET ของประเทศไทยร้อยละ 6 เนื่องจากหุ้น SET50 มีมูลค่าหลักทรัพย์ตามราคาตลาดในสัดส่วนที่มากที่สุดทำให้การเคลื่อนไหวของดัชนี SET มีแนวโน้มที่เคลื่อนไหวในลักษณะเดียวกันจึงส่งกับหุ้นรายตัวด้วยเช่นกัน
- อันดับที่ 5 คือ NEWS หรือข่าวร้อยละ 4 โดยข่าวที่มีผลกับการเคลื่อนไหวของราคาหุ้นจะเป็นข่าวด้านเศรษฐกิจ อย่างเช่น การขึ้นดอกเบี้ย หรือข่าวการเมืองที่เกี่ยวกับเชิงนโยบาย

ตารางที่ 5.2 ตัวแปรอิสระที่สำคัญอันดับที่ 1 ของการพยากรณ์ทิศทาง การเปลี่ยนแปลงราคาหุ้นใน กลุ่ม SET50

ลำดับ	ชื่อตัวแปร	ประเภทตัวแปร	จำนวนหลักทรัพย์
1	SECTOR_INDEX	ปัจจัยเชิงเทคนิค	25 (ร้อยละ 50)
2	INDIVIDUAL_STOCK_VOLUME	ปัจจัยเชิงเทคนิค	11 (ร้อยละ 22)
3	INDIVIDUAL_RSI	ปัจจัยเชิงเทคนิค	4 (ร้อยละ 8)
4	SET_INDEX	ปัจจัยเชิงเทคนิค	3 (ร้อยละ 6)
5	NEWS	ปัจจัยข่าว	2 (ร้อยละ 4)
6	INDIVIDUAL_SECTOR_VOLUME	ปัจจัยเชิงเทคนิค	2 (ร้อยละ 4)
7	PRICE_TREND	ปัจจัยเชิงเทคนิค	1 (ร้อยละ 2)
8	SIGNAL_LINE	ปัจจัยเชิงเทคนิค	1 (ร้อยละ 2)
9	EMA_DISTNACE	ปัจจัยเชิงเทคนิค	1 (ร้อยละ 2)

5.2 ข้อเสนอแนะ

5.2.1 สำหรับงานวิจัยนี้กลยุทธ์ในการซื้อขายส่วนใหญ่จะจบภายใน 1 วันหรือ 1 แท่งเทียน จึงเหมาะกับการเก็งกำไรระยะสั้นมากกว่าการถือในระยะเวลานานๆ ซึ่งเหมาะสำหรับคนที่มีเวลาเฝ้าหน้าจอ และต้องการผลตอบแทนในระยะสั้นๆ

5.2.2 ปัจจัยทางเทคนิคที่ผู้วิจัยใช้เพียงตัวเลขของ Indicator ไม่ได้มีการพิจารณาในเรื่องของรูปแบบแท่งเทียน (Price Pattern) ในการสร้างตัวแปรอิสระสำหรับการพยากรณ์ทิศทางของราคา

5.2.3 ปัจจัยข่าวที่นำมาวิเคราะห์ผู้วิจัยนำเพียงแค่ข่าวต่างประเทศในการสร้างแบบจำลองอาจจะ มีผลกับราคาหุ้น SET50 ในบางช่วง ดังนั้นควรมีการพิจารณาปัจจัยข่าวในประเทศด้วยและการวิเคราะห์ ปัจจัยข่าวในงานวิจัยนี้จะมองเป็นรูปรวมของตลาด ไม่ได้มีการวิเคราะห์ข่าวหุ้นรายตัวในเชิงลึก อย่างเช่นงบกำไรขาดทุน หรือผลประกอบการ

5.2.4 ตัวแปรที่สำคัญจะมีการเปลี่ยนแปลงตามแต่ละช่วงเวลา หรือตามเหตุการณ์ต่างๆ จึงควรมี การทดสอบย้อนหลัง (Backtest) ในแต่ละช่วงเวลาเพื่อให้เกิดความแม่นยำในการใช้ตัวแปรในการ พยากรณ์ทิศทางของราคา

เอกสารอ้างอิง

- [1] ตลาดหลักทรัพย์. 2565. **ดัชนีราคา SET50 Index**. [Online]. Available :
<https://www.set.or.th/th/market/index/set50/profile>
- [2] ตลาดหลักทรัพย์. 2565. **รายชื่อหลักทรัพย์ใน SET50**. [Online]. Available :
<https://www.set.or.th/th/market/index/set50/overview>
- [3] Tan, T. Z., Quek, C., & See, Ng. G. 2007. “Biological brain-inspired genetic complementary learning for stock market and bank failure prediction.” *Computational Intelligence*, 23(2), 236–261.
- [4] McMillan, D. 2005. “Time variation in the cointegrating relationship between stock prices and economic activity.” *International Review of Applied Economics*, 19, 359-368.
- [5] กรุงศรี. 2565. **ดัชนี Dow Jones**. [Online]. Available :
<https://www.krungsri.com/Krungsri2020/media/KSE/Investment/knowledge/get-to-know-us-stock-indexes.PDF>
- [6] รัชณี รุ่งศรีรัตนวงศ์. 2553. “ปัจจัยที่ส่งผลต่อดัชนีราคาตลาดหลักทรัพย์ SET 50.” *บริหารธุรกิจมหาบัณฑิต สาขาวิชาการเงิน, มหาวิทยาลัยหอการค้าไทย*. 44-45.
- [7] Malik, S., Hussain, S. & Ahmed, S. 2009. ” Impact of Political Event on Trading volume and Stock Returns: The Case of KSE.” *International Review of Business Research Papers Vol. 5 No. 4 June 2009 Pp. 354- 36*.
- [8] Fama, Fisher, Jensen and Roll. 1969. “The Adjustment of Stock Prices to New Information”, *International Economic Review*, 19, 1, 1-21.
- [9] Ball and Brown. 1968. “An Empirical Evaluation of Accounting Income Numbers,” *Journal of Accounting Research*, 6:2, 159-178.
- [10] Baker, M., & Wurgler, J. 2007. “Investor sentiment in the stock market.” NBER Working Paper No. 13189. Retrieved from National Bureau of Economic Research

- [11] Bhargavi, R., Srinivas Gumparathi, and R. Anith. 2017. “Relative Strength Index for Developing Effective Trading Strategies in Constructing Optimal Portfolio.” *International Journal of Applied Engineering Research* 12: 8926–36.
- [12] Bustos, Oscar, and Alexandra Pomares-Quimbaya. 2020. “Stock market movement forecast: A Systematic review.” *Expert Systems with Applications* 156: 113464.
- [13] Perry, Marcus B. 2011. “The Weighted Moving Average Technique.” In *Wiley Encyclopedia of Operations Research and Management Science*. Hoboken: John Wiley & Sons, Inc.
- [14] Praekhaow, Puchong. 2010. “Determination of Trading Points using the Moving Average Methods.” Paper presented at the International Conference for a Sustainable Greater Mekong Subregion, Bangkok, Thailand, August 26–27; pp. 519–44.
- [15] Vaidya, Rashesh. 2018. “Stochastic and Momentum Analysis of Nepalese Stock Market.” *The Journal of Nepalese Business Studies* XI: 14–22.
- [16] Worthington A.C. & Valadkhani A. 2004. “Measuring the impact of natural disasters on capital markets: An empirical application using intervention analysis.” *Applied Economics* 36(19), 2177–2186.
- [17] Bastianin A, Conti F, Manera M. 2016. “The impacts of oil price shocks on stock market volatility: evidence from the G7 countries.” *Energy Policy* 98:160–169.
- [18] หลักทรัพย์บัวหลวง. 2017. **แท่งเทียนราคา**. [Online]. Available : <https://knowledge.bualuang.co.th/knowledge-base/candlesticks/>
- [19] หลักทรัพย์บัวหลวง. 2018. **Volume หรือ ปริมาณการซื้อขาย**. [Online]. Available : <https://knowledge.bualuang.co.th/knowledge-base/volume/>
- [20] หลักทรัพย์บัวหลวง. 2020. **รู้จัก EMA คืออะไร**. [Online]. Available : <https://knowledge.bualuang.co.th/knowledge-base/whatisema/>
- [21] หลักทรัพย์บัวหลวง. 2020. **รู้จัก MACD คืออะไร**. [Online]. Available <https://knowledge.bualuang.co.th/knowledge-base/macd/>

- [22] หลีกทรัพย์บัวหลวง. 2021. **RSI คืออะไร**. [Online]. Available :
<https://knowledge.bualuang.co.th/knowledge-base/rsi/>
- [23] Chinmay Bhalerao. 2022. **Feature selection techniques for data**. [Online].
 Available : <https://medium.com/mlearning-ai/feature-selection-techniques-for-data-57f0eacd8fa8>
- [24] Maria Gusarova. 2022. **Feature Selection Techniques**. [Online]. Available :
<https://medium.com/@data.science.enthusiast/feature-selection-techniques-forward-backward-wrapper-selection-9587f3c70cfa>
- [25] Rohith Gandhi. 2018. **Support Vector Machine - Introduction to Machine Learning Algorithms**. [Online]. Available :
<https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [26] Davis David. 2020. **Random Forest Classifier Tutorial: How to Use Tree-Based Algorithms for Machine Learning**. [Online]. Available :
<https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/>
- [27] ชิตพงษ์ กิตตินราดร. 2563. **Boosting**. [Online]. Available :
<https://guopai.github.io/ml-blog11.html>
- [28] Phiratath Nopratana Wong. 2019. **Fraud Detection : EP 3 Modeling** [Online]. Available :
<https://medium.com/siamfriend/fraud-detection-ep-3-modeling-%E0%B8%A5%E0%B9%89%E0%B8%A1%E0%B9%80%E0%B8%A3%E0%B9%87%E0%B8%A7-%E0%B9%80%E0%B8%A3%E0%B8%B5%E0%B8%A2%E0%B8%99%E0%B9%80%E0%B8%A3%E0%B9%87%E0%B8%A7-fail-fast-learn-faster-dad8194b3e55>
- [29] lukkidd. 2018. **Word Embedding and Word2Vec**. [Online]. Available :
<https://lukkidd.com/word-embedding-%E0%B9%81%E0%B8%A5%E0%B8%B0-word2vec->

%E0%B8%84%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-e60bdf6d78d3

[30] Soumayan Pal. 2021. **What is Text Similarity and How to Implement it?** [Online].

Available : <https://medium.com/msackiit/what-is-text-similarity-and-how-to-implement-it-c74c8b641883>

[31] Pagon Gatchalee. 2019 **Confusion Matrix เครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย ใน Machine learning** [Online]. Available :

<https://medium.com/@pagongatchalee/confusion-matrix->

%E0%B9%80%E0%B8%84%E0%B8%A3%E0%B8%B7%E0%B9%88%E0%B8%AD%E0%B8%87%E0%B8%A1%E0%B8%B7%E0%B8%AD%E0%B8%AA%E0%B8%B3%E0%B8%84%E0%B8%B1%E0%B8%8D%E0%B9%83%E0%B8%99%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B8%9B%E0%B8%A3%E0%B8%B0%E0%B9%80%E0%B8%A1%E0%B8%B4%E0%B8%99%E0%B8%9C%E0%B8%A5%E0%B8%A5%E0%B8%B1%E0%B8%9E%E0%B8%98%E0%B9%8C%E0%B8%82%E0%B8%AD%E0%B8%87%E0%B8%81%E0%B8%B2%E0%B8%A3%E0%B8%97%E0%B8%B3%E0%B8%99%E0%B8%B2%E0%B8%A2-%E0%B9%83%E0%B8%99machine-learning-fba6e3f9508c



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

```

import fasttext
import fasttext.util

fasttext.util.download_model('en', if_exists='ignore')
ft = fasttext.load_model('cc.en.300.bin')
df = pd.read_csv('/content/stock_news.csv')

# define the topic dictionary
topic_dict = {
    "Economic": {
        "keywords": ['gdp', 'inflation', 'interest rates', 'economy', 'CPI',
                    'PPI', 'employment rate', 'unemployment rate',
                    'labor force participation rate',
                    'Federal Reserve', 'monetary policy', 'fiscal policy', 'GNP',
                    'national debt', 'trade balance', 'consumer confidence', 'PMI', 'IPI',
                    'retail sales', 'housing starts', 'building permits', 'consumer spending',
                    'business investment', 'corporate profits', 'productivity',
                    'economic growth', 'recession', 'recovery', 'government spending', 'net
exports'],
        "subvalues": {
            "Inflation": ["lower than expected", "higher than expected"],
            "government spending": ["more spend", "less spent"]
        }
    },
    "Political events": {

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

"keywords": ['election', 'policy change', 'geopolitical tension', 'Trade policy',
            'voting', 'polls', 'ballots', 'candidates', 'voters', 'presidential',
            'policy change', 'tariffs', 'trade war', 'sanctions',
            'geopolitical tension', 'conflict', 'war', 'terrorism', 'embargo', 'diplomacy'],
"subvalues": {
    "election": ["win", "lose", "result"],
    "war": ['weapons']
}
},
"Company events": {
    "keywords": ['earnings report', 'earnings call', 'quarterly results', 'annual results',
                'financial results', 'revenue', 'sales', 'top line', 'profits', 'bottom line',
                'net income', 'earnings per share', 'losses', 'negative earnings', 'loss per
share'],
    "subvalues": {
        "election": ["win", "lose", "result"],
        "war": ['weapons']
    }
},
"Investor sentiment": {
    "keywords": ["bullish", "bearish", "optimistic", "pessimistic", "Rally", "crash", "hopeful",
                "worried", "uncertain", "volatile", "fearful", "greedy", "fear/greed"],
    "subvalues": {
        "bullish": ["positive", "Rally"],
        "bearish": ["negative", "Crash"],
    }
},
"Natural disasters": {

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

"keywords": ['Earthquake', 'Tsunami', 'Hurricane', 'Tornado', 'Flood',
             'Drought', 'Wildfire', 'Volcanic eruption', 'Landslide', 'Avalanche', 'Blizzard',
             'Cyclone', 'Hailstorm', 'Heat wave', 'Monsoon', 'Thunderstorm', 'Typhoon'],
"subvalues": {
  "Earthquake": [],
  "Tsunami": []
}
},
"Industry specific events": {
  "keywords": ['biotech', 'pharmaceutical', 'healthcare', 'energy',
              'oil and gas', 'renewable energy', 'technology', 'software',
              'hardware', 'telecom', 'internet', 'ecommerce', 'retail',
              'automotive', 'transportation', 'aviation'],
  "subvalues": {
    "biotech": [],
    "pharmaceutical": []
  }
},
"Market moment": {
  "keywords": ['Bull market', 'Bear market', 'Market correction', 'Market volatility',
              'Market rally',
              'Market downturn', 'Market trend', 'Market upswing', 'Market
              downswing'],
  "subvalues": {
    "Bull market": [],
    "Bear market": []
  }
}
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

}

# Define a function to calculate the similarity score between a list of keywords and a
sentence
def calc_similarity_score(keywords, sentence):
    sentence_vec = ft.get_sentence_vector(sentence)
    keywords_vec = [ft.get_word_vector(keyword) for keyword in keywords]
    similarities = [np.dot(sentence_vec,
keyword_vec)/(np.linalg.norm(sentence_vec)*np.linalg.norm(keyword_vec)) for
keyword_vec in keywords_vec]
    return sum(similarities) / len(similarities)

# Add a column to the dataframe with the similarity scores for each topic and keyword
for topic, topic_dict in topic_dict.items():
    keywords = topic_dict['keywords'] + [value for sublist in topic_dict['subvalues'].values()
for value in sublist]
    df[topic] = df['cleaned_topic'].apply(lambda x: calc_similarity_score(keywords, x))

df.to_csv('/content/drive/MyDrive/IS/defined_topic.csv')

defined_stock_df = pd.read_csv('defined_topic.csv')
defined_stock_df = defined_stock_df.drop(['Unnamed: 0', 'Unnamed: 0.1', 'Details',
'StrDate', 'cleaned_topic'], axis=1)
defined_stock_df.head()

# Group by date and aggregate scores and keywords
grouped_global = defined_stock_df.groupby('Date').agg({'Economic': 'sum',
'Political events': 'sum',

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
'Company events': 'sum',
'Investor sentiment': 'sum',
'Natural disasters': 'sum',
'Industry specific events': 'sum',
'Market moment': 'sum'})
```

```
grouped_global = grouped_global.sort_values(by='Date', ascending=False)
```

```
# Define the list of symbols
```

```
symbol_index = ['^DJII', '^ANDX', '^SPX', '^SET']
```

```
# Download the stock data from Stooq
```

```
stock_df = web.DataReader(symbol_index, 'stooq')[['Close', 'Open', 'Volume']]
```

```
# Compute rolling correlation of DJI with NDX, SPX, and SET
```

```
dji_close = stock_df['Close']['^DJII']
```

```
ndx_close = stock_df['Close']['^ANDX']
```

```
spx_close = stock_df['Close']['^SPX']
```

```
set_close = stock_df['Close']['^SET']
```

```
rolling_corr_ndx = dji_close.rolling(window=20, min_periods=2).corr(ndx_close)
```

```
rolling_corr_spx = dji_close.rolling(window=20, min_periods=2).corr(spx_close)
```

```
rolling_corr_set = dji_close.rolling(window=20, min_periods=2).corr(set_close)
```

```
# Add new columns to the DataFrame
```

```
stock_df['Rolling_corr_DJI_NDX'] = rolling_corr_ndx
```

```
stock_df['Rolling_corr_DJI_SPX'] = rolling_corr_spx
```

```
stock_df['Rolling_corr_DJI_SET'] = rolling_corr_set
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# Define function to determine direction based on open and close prices
def get_direction(row):
    return np.where(row['Close'] > row['Open'], 'up',
                   np.where(row['Close'] < row['Open'], 'down',
                              np.where(row['Close'] == row['Open'], 'no change', 'closed market'))))

# Create a new column 'Direction' using the get_direction function
stock_df['Direction'] = stock_df.apply(get_direction, axis=1)
stock_df.head()

stock_direction = stock_df[['Close', 'Volume', 'Direction', 'Rolling_corr_DJI_NDX',
                             'Rolling_corr_DJI_SPX', 'Rolling_corr_DJI_SET']]

stock_direction.columns = stock_direction.columns.droplevel(0)

stock_direction = stock_direction.reset_index()

stock_direction.columns = ['Date', 'DJI_Close', 'NDX_Close', 'SPX_Close', 'SET_Close',
                           'DJI_Volume', 'NDX_Volume', 'SPX_Volume', 'SET_Volume', 'Direction',
                           'Rolling_corr_DJI_NDX', 'Rolling_corr_DJI_SPX', 'Rolling_corr_DJI_SET']

# define a function to extract the values
def extract_values(row):
    return pd.Series(row['Direction'])

# apply the function to each row and add the columns to the dataframe

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

stock_direction[['^DJI', '^NDX', '^SPX', '^SET']] = stock_direction.apply(extract_values,
axis=1)

# drop the original Direction column
stock_direction = stock_direction.drop('Direction', axis=1)
stock_direction
global_stock_market = pd.merge(stock_direction, grouped_global, on='Date',
how='outer')

# create a dictionary to map string values to numerical values
map_dict = {'up': 1, 'down': -1, 'no change': 0, 'closed market': -9999}

# apply the mapping to the columns using the replace() method
global_stock_market[['^DJI', '^NDX', '^SPX', '^SET']] = global_stock_market[['^DJI', '^NDX',
'^SPX', '^SET']].replace(map_dict)

def row_ratio_scaling(row):
    row_sum = row.sum()
    return row / row_sum

# apply min-max scaling to each row
global_stock_market[['Economic', 'Political events', 'Company events', 'Investor
sentiment', 'Natural disasters', 'Industry specific events', 'Market moment']] =
global_stock_market[['Economic', 'Political events', 'Company events', 'Investor
sentiment', 'Natural disasters', 'Industry specific events', 'Market
moment']].apply(row_ratio_scaling, axis=1)

global_stock_market = global_stock_market.sort_values('Date', ascending=False)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

global_stock_market = global_stock_market.reset_index(drop=True)

def extract_date_info(df, col_name):
    # extract date components
    date_components = pd.to_datetime(df[col_name])

    # create new columns for day name, month name, and year
    df['day_name'] = date_components.dt.day_name()
    df['month_name'] = date_components.dt.month_name()
    df['year'] = date_components.dt.year
    df['day'] = date_components.dt.day
    df['quarter'] = date_components.dt.quarter

    # return the modified dataframe
    return df

df = extract_date_info(global_stock_market, 'Date')

# Delete rows where all three columns are null
df = df.dropna(subset=['^DJI', '^NDX', '^SPX', '^SET', 'Economic', 'Political events', 'Investor
sentiment', 'Company events', 'Market moment', 'Natural disasters', 'Industry specific
events'], how='any')

new_df = df.copy()
new_df = new_df.sort_values('Date', ascending=False)
new_df = new_df.reset_index(drop=True)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# select the columns of interest
cols = ['Economic', 'Political events', 'Investor sentiment', 'Company events', 'Market
moment', 'Natural disasters', 'Industry specific events']

# create a new column with the name of the column with the maximum value
new_df['max_column'] = new_df[cols].idxmax(axis=1)

# create a new column with a dummy variable indicating whether the maximum value is
present in that column
for column in cols:
    new_df[f'{column}_dummy'] = (new_df['max_column'] == column).astype(int)

# drop the 'max_column' column
new_df.drop(columns='max_column', inplace=True)

new_df = new_df.drop(columns=['Economic_dummy', 'Political events_dummy', 'Investor
sentiment_dummy', 'Company events_dummy',
                             'Market moment_dummy', 'Natural disasters_dummy', 'Industry
specific events_dummy', 'year', 'day'], axis=1)

new_df.head()

import os
import subprocess
import string
import pandas as pd
pd.set_option('display.max_rows', 500)
pd.set_option('display.max_columns', 500)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

pd.set_option('display.width', 1000)

import numpy as np
import seaborn as sns
import glob
import pickle
import requests

from bs4 import BeautifulSoup
from datetime import datetime
import re

import matplotlib.pyplot as plt
from sklearn.metrics import classification_report
import plotly.express as px
import yfinance as yf
import ta
import pandas_datareader.data as web
from finvizfinance.quote import finvizfinance
import starfishX as sx

def get_stock_data(symbols, threshold=0.8, window=20):
    data = {}
    for symbol in symbols:
        stock_data = yf.download(symbol + ".BK", period="5y", interval="1d")[['Close',
'Open', 'Volume']]
        stock_data.loc[:, 'EMA200'] = stock_data["Close"].ewm(span=200,
adjust=False).mean()
        stock_data.loc[:, 'EMA50'] = stock_data["Close"].ewm(span=50, adjust=False).mean()
        stock_data.loc[:, 'EMA20'] = stock_data["Close"].ewm(span=20, adjust=False).mean()

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
stock_data.loc[:, 'Target'] = np.where(stock_data["Close"] > stock_data["Open"], "up",
np.where(stock_data["Close"] < stock_data["Open"], "down", "no change"))
```

```
stock_data.loc[:, 'RSI'] = ta.momentum.RSIIndicator(stock_data["Close"]).rsi()
```

```
exp1 = stock_data['Close'].ewm(span=12, adjust=False).mean()
```

```
exp2 = stock_data['Close'].ewm(span=26, adjust=False).mean()
```

```
stock_data.loc[:, 'MACD'] = exp1 - exp2
```

```
stock_data.loc[:, 'Signal line'] = stock_data['MACD'].ewm(span=9,
adjust=False).mean()
```

```
stock_data.loc[:, 'MACD_Signal'] = ta.trend.MACD(stock_data['Close']).macd_signal()
```

```
stock_data.loc[:, 'MACD_Diff'] = ta.trend.MACD(stock_data['Close']).macd_diff()
```

```
stock_data.loc[:, 'MACD_Bullish_Crossover'] = (stock_data['MACD'] >
stock_data['MACD_Signal']) & (stock_data['MACD_Diff'] > 0)
```

```
stock_data.loc[:, 'MACD_Bearish_Crossover'] = (stock_data['MACD'] <
stock_data['MACD_Signal']) & (stock_data['MACD_Diff'] < 0)
```

```
stock_data.loc[:, 'MACD_Signal'] = np.where(stock_data['MACD_Bullish_Crossover'],
'Bullish', np.where(stock_data['MACD_Bearish_Crossover'], 'Bearish', 'Neutral'))
```

```
stock_data = stock_data.drop(columns=['MACD_Bullish_Crossover',
'MACD_Bearish_Crossover'], axis=1)
```

```
threshold = threshold
```

```
window = window
```

```
mean = stock_data['Volume'].rolling(window).mean()
```

```
std = stock_data['Volume'].rolling(window).std()
```

```
stock_data.loc[:, 'Volume_Z_Score'] = (stock_data['Volume'] - mean) / std
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

stock_data.loc[:, 'Volume_Anomaly'] =
np.where(np.abs(stock_data['Volume_Z_Score']) > threshold, True, False)
stock_data.loc[:, 'Volume_Anomaly_Count'] =
stock_data['Volume_Anomaly'].rolling(window).sum()

# Add the death cross and golden cross
stock_data.loc[:, 'Death_Cross'] = (stock_data['EMA50'] < stock_data['EMA200'])
stock_data.loc[:, 'Golden_Cross'] = (stock_data['EMA50'] > stock_data['EMA200'])
stock_data.loc[:, 'Primary_Trend'] = np.where(stock_data['Death_Cross'], 'Down
Trend', np.where(stock_data['Golden_Cross'], 'Up Trend', 'Sideways Trend'))
#stock_data = stock_data.drop(columns=['Death_Cross', 'Golden_Cross'], axis=1)

# Add the price direction column based on conditions
stock_data.loc[:, 'Price_Trend'] = np.where((stock_data['EMA50'] >
stock_data['EMA200']) & (stock_data['Close'] > stock_data['EMA50']) & (stock_data['Close']
> stock_data['EMA200']), 'Up Trend',
np.where((stock_data['EMA50'] > stock_data['EMA200'])
& (stock_data['Close'] < stock_data['EMA50']) & (stock_data['Close'] >
stock_data['EMA200']), 'Weak Up Trend',
np.where((stock_data['EMA50'] > stock_data['EMA200'])
& (stock_data['Close'] < stock_data['EMA50']) & (stock_data['Close'] <
stock_data['EMA200']), 'Down Trend Coming',
np.where((stock_data['EMA50'] < stock_data['EMA200'])
& (stock_data['Close'] < stock_data['EMA50']) & (stock_data['Close'] <
stock_data['EMA200']), 'Down Trend',
np.where((stock_data['EMA50'] < stock_data['EMA200'])
& (stock_data['Close'] > stock_data['EMA50']) & (stock_data['Close'] <
stock_data['EMA200']), 'Weak Down Trend',

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

np.where((stock_data['EMA50'] < stock_data['EMA200'])
& (stock_data['Close'] > stock_data['EMA50']) & (stock_data['Close'] >
stock_data['EMA200']), 'Up Trend Coming','Sideways Trend'))))

stock_data.loc[:, 'Short_Trend'] = np.where(stock_data['Close'] >
stock_data['EMA20'], 'Strong', 'Weak')

# Calculate the distance between the two EMAs
stock_data['EMA_Distance'] = abs(stock_data['EMA50'] - stock_data['EMA200']) /
((stock_data['EMA50'] + stock_data['EMA200']) / 2) * 100

# Add the column to indicate changing of trends
#stock_data.loc[:, 'Changing_of_Trends'] =
np.where((abs(stock_data['EMA_Distance']) < 1) & (stock_data['Primary_Trend'] == 'Up
Trend'), 'Up Trend -> Down Trend',
#np.where((abs(stock_data['EMA_Distance']) < 1) &
(stock_data['Primary_Trend'] == 'Down Trend'), 'Down Trend -> Up Trend', 'None'))

stock_data.loc[:, 'Changing_of_Trends'] = np.where((abs(stock_data['EMA_Distance'])
< 1), 'New trend','None')

# Write a critical point uptrend and down trend
buffer_ema50 = stock_data['EMA50'] * 0.015
buffer_ema200 = stock_data['EMA200'] * 0.015
stock_data.loc[:, 'Critical_Point'] = np.where(np.abs(stock_data['Close'] -
stock_data['EMA50']) < buffer_ema50, 'EMA50 Critical Point',
np.where(np.abs(stock_data['Close'] -
stock_data['EMA200']) < buffer_ema200, 'EMA200 Critical Point', np.NaN))

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

stock_data['Critical_Point_occurrence'] =
stock_data['Critical_Point'].str.contains('Critical Point').astype(int)
stock_data['Count_Critical_Point'] = stock_data['Critical_Point_occurrence']
stock_data.loc[:, 'Changing_of_Trends'] =
np.where(stock_data['Count_Critical_Point'] > 1, 'Sideway: ' +
stock_data['Count_Critical_Point'].astype(str), stock_data['Changing_of_Trends'])

stock_data = stock_data.drop(columns=['MACD_Diff', 'Volume_Z_Score',
'Volume_Anomaly'], axis=1)
data[symbol] = stock_data
df = pd.concat(data.values(), keys=data.keys(), axis=1)
return df

set50_df = get_stock_data(set50_name[:])

set50_df.columns = [f"{idx}_{col}" for idx, col in zip(set50_df.columns.get_level_values(0),
set50_df.columns.get_level_values(1))]

# Mering All Features
global_features = pd.read_csv('global_feats.csv')

global_features = global_features.drop('Unnamed: 0', axis=1)

global_features['Date'] = pd.to_datetime(global_features['Date'])

label_map = {'down': -1, 'no change': 0, 'up': 1}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

for stock in my_list:
    df = pd.merge(global_features, dfs[stock].reset_index(), on='Date', how='inner')
    target_col = stock + '_Target'
    df[target_col] = df[target_col].map(label_map)

    dfs[stock] = df

# Exclude SET 0:-2
for stock in my_list[0:-2]:
    # Create new features as T-1
    cols_to_shift = ['Rolling_corr_DJI_SET', stock+'_Close', stock+'_Volume',
stock+'_EMA200', stock+'_EMA50', stock+'_EMA20',
stock+'_Target', stock+'_RSI', stock+'_MACD', stock+'_Signal line',
stock+'_MACD_Signal', stock+'_Volume_Anomaly_Count', stock+'_Death_Cross',
stock+'_Golden_Cross', stock+'_Primary_Trend', stock+'_Price_Trend',
stock+'_Short_Trend', stock+'_EMA_Distance', stock+'_Changing_of_Trends',
stock+'_Critical_Point', stock+'_Critical_Point_occurrence',
stock+'_Count_Critical_Point', stock+'_Close_corr_with_SET_Close']

    for col in cols_to_shift:
        dfs[stock][f'{col}_T-1'] = dfs[stock][col].shift(-1)

# Merge dataframes by date
global_col = ['Date', 'day_name', 'month_name', 'quarter', 'DJI_Close', 'NDX_Close',
'SPX_Close', 'DJI_Volume', 'NDX_Volume', 'SPX_Volume', 'Rolling_corr_DJI_NDX',
'Rolling_corr_DJI_SPX', 'Economic', 'Political events', 'Investor sentiment', 'Company

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
events', 'Market moment', 'Natural disasters', 'Industry specific events', '^DJI', '^NDX',
'^SPX']
```

```
individual_col = ['Date', 'Rolling_corr_DJI_SET_T-1', f'{stock}_Close_T-1',
f'{stock}_Volume_T-1', f'{stock}_EMA200_T-1', f'{stock}_EMA50_T-1',
f'{stock}_EMA20_T-1', f'{stock}_Target_T-1', f'{stock}_RSI_T-1',
f'{stock}_MACD_T-1', f'{stock}_Signal line_T-1',
f'{stock}_MACD_Signal_T-1', f'{stock}_Volume_Anomaly_Count_T-1',
f'{stock}_Death_Cross_T-1', f'{stock}_Golden_Cross_T-1',
f'{stock}_Primary_Trend_T-1', f'{stock}_Price_Trend_T-1',
f'{stock}_Short_Trend_T-1', f'{stock}_EMA_Distance_T-1',
f'{stock}_Changing_of_Trends_T-1', f'{stock}_Critical_Point_T-1',
f'{stock}_Critical_Point_occurrence_T-1',
f'{stock}_Count_Critical_Point_T-1', f'{stock}_Close_corr_with_SET_Close_T-1']

dfs[stock] = pd.merge(dfs[stock][global_col], dfs[stock][individual_col], on='Date',
how='inner')
set_col = ['Date', 'SET_Close_y', 'SET_Volume_y', 'SET_EMA200', 'SET_EMA50',
'SET_EMA20', 'SET_Target', 'SET_RSI', 'SET_MACD', 'SET_Signal line', 'SET_MACD_Signal',
'SET_Volume_Anomaly_Count', 'SET_Death_Cross', 'SET_Golden_Cross',
'SET_Primary_Trend', 'SET_Price_Trend', 'SET_Short_Trend', 'SET_EMA_Distance',
'SET_Changing_of_Trends', 'SET_Critical_Point', 'SET_Critical_Point_occurrence',
'SET_Count_Critical_Point', 'SET_Close_corr_with_SET_Close']
dfs['SET'] = dfs['SET'][set_col]

# Select the columns to shift and create the shifted columns

for col in set_col:
    dfs['SET'][f'{col}_T-1'] = dfs['SET'][col].shift(-1)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# Get a list of column names that end with 'T-1' or are 'Date'
cols = [col for col in dfs['SET'].columns if col.endswith('T-1') or col == 'Date']

# Select the columns in the dataframe
setindex_df = dfs['SET'][cols]

setindex_df = setindex_df.drop('Date_T-1', axis=1)

from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import MinMaxScaler, OneHotEncoder
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier,
AdaBoostClassifier
from sklearn.svm import SVC
from sklearn.naive_bayes import GaussianNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from catboost import CatBoostClassifier
import plotly.graph_objects as go

from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import Lasso
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import StratifiedKFold

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score,
roc_auc_score, confusion_matrix

from sklearn.neural_network import MLPClassifier

from sklearn.model_selection import StratifiedShuffleSplit

from sklearn import set_config

import lightgbm as lgb

from sklearn.feature_selection import RFE

```

```

stock_symbol = 'AOT'
start_feature_row = 1
end_feature_row = 504
start_target_row = 505
end_target_row = 631

features_df =
new_dfs[stock_symbol].iloc[start_feature_row:end_feature_row].drop(['Date',
stock_symbol+'_Target_T-1'], axis=1)

target_df = new_dfs[stock_symbol][stock_symbol+'_Target_T-
1'].iloc[start_target_row:end_target_row]

```

```

def evaluate_classification_model(y_true, y_pred, y_proba=None):

```

```

    """

```

```

    Evaluate a multiclass classification model using various metrics.

```

```

    Args:

```

```

    - y_true: array-like, true labels of the data

```

```

    - y_pred: array-like, predicted labels of the data

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- y_proba: array-like, predicted probabilities of the data (default: None)

Returns:

- dictionary containing the evaluation metrics

"""

```
evaluation = {}
```

```
# Calculate evaluation metrics
```

```
evaluation['accuracy'] = accuracy_score(y_true, y_pred)
```

```
evaluation['macro_precision'] = precision_score(y_true, y_pred, average='macro')
```

```
evaluation['micro_precision'] = precision_score(y_true, y_pred, average='micro')
```

```
evaluation['weighted_precision'] = precision_score(y_true, y_pred, average='weighted')
```

```
evaluation['macro_recall'] = recall_score(y_true, y_pred, average='macro')
```

```
evaluation['micro_recall'] = recall_score(y_true, y_pred, average='micro')
```

```
evaluation['weighted_recall'] = recall_score(y_true, y_pred, average='weighted')
```

```
evaluation['macro_f1_score'] = f1_score(y_true, y_pred, average='macro')
```

```
evaluation['micro_f1_score'] = f1_score(y_true, y_pred, average='micro')
```

```
evaluation['weighted_f1_score'] = f1_score(y_true, y_pred, average='weighted')
```

```
# If predicted probabilities are given, calculate AUC-ROC
```

```
if y_proba is not None:
```

```
    evaluation['macro_auc_roc'] = roc_auc_score(y_true, y_proba, multi_class='ovo',
average='macro')
```

```
    evaluation['micro_auc_roc'] = roc_auc_score(y_true, y_proba, multi_class='ovo',
average='micro')
```

```
    evaluation['weighted_auc_roc'] = roc_auc_score(y_true, y_proba, multi_class='ovo',
average='weighted')
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# Calculate confusion matrix
confusion = confusion_matrix(y_true, y_pred)
evaluation['confusion_matrix'] = confusion.tolist()

return evaluation

# Split data into train, validation, and test sets
X_train, X_test, y_train, y_test = train_test_split(features_df, target_df, test_size=0.3,
random_state=random_seed, stratify=target_df) #stratify=target_df

# Calculate the proportion of target in the original data
prop_target_original = target_df.value_counts(normalize=True)

# Calculate the proportion of target in the training set
prop_target_train = y_train.value_counts(normalize=True)

# Calculate the proportion of target in the test set
prop_target_test = y_test.value_counts(normalize=True)

numeric_features = X_train.select_dtypes(include=['float64', 'int64', 'int32']).columns
categorical_features = X_train.select_dtypes(include=['object']).columns

# Define numeric transformer pipeline
numeric_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='median')),

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

('scaler', MinMaxScaler())
])

# Define categorical transformer pipeline
categorical_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='most_frequent')),
    ('onehot', OneHotEncoder(handle_unknown='ignore'))
])

# Combine numeric and categorical transformers into a single pipeline
preprocessor = ColumnTransformer(
    transformers=[
        ('num', numeric_transformer, numeric_features),
        ('cat', categorical_transformer, categorical_features)
    ])

results = {}
for name, clf in classifiers.items():
    print(name)
    pipeline = Pipeline(steps=[('preprocessor', preprocessor), ('classifier', clf)])

# Fit the pipeline on the training set
pipeline.fit(X_train, y_train)

# Make prediction
y_pred_test = pipeline.predict(X_test)
y_pred_train = pipeline.predict(X_train)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# Evaluate Train and Test

model_train = evaluate_classification_model(y_train, y_pred_train)

model_test = evaluate_classification_model(y_test, y_pred_test)

results[name] = {'model_train': model_train, 'model_test': model_test}

print(results)

# Cross-Validation
def plot_results(results):
    # Create lists of model names, train accuracies, and test accuracies
    model_names = []
    train_accs = []
    test_accs = []

    for model, scores in results.items():
        model_names.append(model)
        train_accs.append(scores['model_train']['accuracy'])
        test_accs.append(scores['model_test']['accuracy'])

    # Create the bar chart
    fig = go.Figure(data=[
        go.Bar(name='Model Train', x=model_names, y=train_accs, text=train_accs,
        textposition='auto'),
        go.Bar(name='Model Test', x=model_names, y=test_accs, text=test_accs,
        textposition='auto')
    ])

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# Update the layout
fig.update_layout(
    title='Model Evaluation Results',
    xaxis_title='Model',
    yaxis_title='Accuracy',
    barmode='group',
    bargap=0.15,
    bargroupgap=0.1
)

# Set y-axis tick format to two decimal places
fig.update_yaxes(tickformat=".2f")

# Add labels to the bars
fig.update_traces(texttemplate='%{y:.2f}', textposition='outside')

# Show the chart
fig.show()

selected_classifiers = {
    'SVM': SVC(random_state=random_seed, kernel='rbf', C=1, gamma='scale'),
    'CatBoost': CatBoostClassifier(verbose=0, l2_leaf_reg=9, random_state=random_seed,
learning_rate=0.1, n_estimators=100, max_depth=5),
    'AdaBoost': AdaBoostClassifier(random_state=random_seed, n_estimators=100,
learning_rate=0.1)
}

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

cv_results = []
for name, classifier in selected_classifiers.items():
    pipeline = Pipeline(steps=[('preprocessor', preprocessor), ('classifier', classifier)])

    # Fit the pipeline on the training set
    pipeline.fit(X_train, y_train)

    scores = cross_val_score(pipeline, X_train, y_train, cv=5)
    mean_accuracy = np.mean(scores)
    std_accuracy = np.std(scores)
    cv_results.append({'Classifier': name, 'Mean Accuracy': mean_accuracy, 'Std':
std_accuracy})

cv_results_df = pd.DataFrame(cv_results)
cv_results_df

correlation_df = features_df.join(target_df)

# Select the features you want to compare with the target
selected_features = ['AOT_Volume_T-1', 'Company events', 'ICT.Vol_T-1']

selected_features2 = ['TRANS_Target_T-1', 'SET_Target_T-1']

# Convert categorical target variable to numeric labels
target_labels, target_codes = np.unique(correlation_df['AOT_Target_T-1'],
return_inverse=True)
correlation_df['AOT_Target_T-1'] = target_codes

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# Calculate correlation between numerical features and categorical target
num_corr_matrix =
correlation_df[selected_features].corrwith(correlation_df['AOT_Target_T-1'])

# Calculate ANOVA between categorical features and categorical target
cat_corr_matrix = pd.DataFrame()
for column in selected_features2:
    grouped_data = correlation_df.groupby(column)['AOT_Target_T-1'].apply(lambda x:
stats.f_oneway(x, correlation_df['AOT_Target_T-1'])[0])
    cat_corr_matrix[column] = grouped_data

# Combine the correlation matrices
corr_matrix = pd.concat([num_corr_matrix, cat_corr_matrix])

# Print the correlation matrix
print(corr_matrix)

# split the data into train and test sets
X_train, X_test, y_train, y_test = train_test_split(features_df, target_df, stratify=target_df,
test_size=0.2, random_state=random_seed)

# get numeric and categorical features
numeric_features = X_train.select_dtypes(include=['int64', 'float64']).columns
categorical_features = X_train.select_dtypes(include=['object']).columns

# scale the numeric features in the training and test sets
scaler = MinMaxScaler()

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

X_train_scaled = X_train.copy()
X_test_scaled = X_test.copy()
X_train_scaled[numeric_features] = scaler.fit_transform(X_train_scaled[numeric_features])
X_test_scaled[numeric_features] = scaler.transform(X_test_scaled[numeric_features])

# encode the categorical features in the training and test sets
encoder = OneHotEncoder(sparse=False, handle_unknown='ignore')
X_train_encoded = encoder.fit_transform(X_train_scaled[categorical_features])
X_test_encoded = encoder.transform(X_test_scaled[categorical_features])

# get the names of the encoded features
encoded_feature_names = encoder.get_feature_names(categorical_features)

# create new dataframes with the encoded features and their names as column names
X_train_encoded_df = pd.DataFrame(X_train_encoded,
                                  columns=encoded_feature_names)
X_test_encoded_df = pd.DataFrame(X_test_encoded,
                                  columns=encoded_feature_names)

# concatenate the encoded dataframe with the original numerical dataframe for both
training and test sets
X_train_final =
X_train_encoded_df.join(X_train_scaled.reset_index(drop=True)[numeric_features])
X_test_final =
X_test_encoded_df.join(X_test_scaled.reset_index(drop=True)[numeric_features])

# Create an AdaBoostClassifier object
adaboost_clf = AdaBoostClassifier(random_state=random_seed, n_estimators=100)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# fit the GridSearchCV object to the training data
adaboost_clf.fit(X_train_final, y_train)

y_pred = adaboost_clf.predict(X_test_final)

# evaluate accuracy of best model on test data
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy score of best model on test data: {:.2f}".format(accuracy))

# retrieve the feature importances
feature_importance = adaboost_clf.feature_importances_

important_features = X_train_final.columns[feature_importance > 40]
print(important_features)

# Sort the feature importance scores
sorted_idx = np.argsort(feature_importance)
sorted_idx = sorted_idx[len(feature_importance) - 50:]

# Define the position of each bar
pos = np.arange(sorted_idx.shape[0]) + 0.5

# Create the plot
plt.figure(figsize=(10, 12))
plt.barh(pos, feature_importance[sorted_idx], align='center')
plt.yticks(pos, X_train_final.columns[sorted_idx], size=14)
plt.xticks(size=14)

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

plt.xlabel('Importance', size=14)
plt.title('Feature Importance', size=16)
plt.show()

feature_selected = X_train_final.columns[sorted_idx][-5][::-1]

for num_feature in range(1, 6):
    # split the data into train and test sets
    X_train, X_test, y_train, y_test =
train_test_split(features_df[feature_selected[:num_feature]], target_df, test_size=0.2,
random_state=random_seed)

    # get numeric and categorical features
    numeric_features = X_train.select_dtypes(include=['int64', 'float64']).columns
    categorical_features = X_train.select_dtypes(include=['object']).columns

    # scale the numeric features in the training and test sets
    scaler = MinMaxScaler()
    X_train_scaled = X_train.copy()
    X_test_scaled = X_test.copy()
    X_train_scaled[numeric_features] =
scaler.fit_transform(X_train_scaled[numeric_features])
    X_test_scaled[numeric_features] = scaler.transform(X_test_scaled[numeric_features])

    # encode the categorical features in the training and test sets
    encoder = OneHotEncoder(sparse=False, handle_unknown='ignore')
    X_train_encoded = encoder.fit_transform(X_train_scaled[categorical_features])
    X_test_encoded = encoder.transform(X_test_scaled[categorical_features])

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

# get the names of the encoded features
encoded_feature_names = encoder.get_feature_names(categorical_features)

# create new dataframes with the encoded features and their names as column
names
X_train_encoded_df = pd.DataFrame(X_train_encoded,
columns=encoded_feature_names)
X_test_encoded_df = pd.DataFrame(X_test_encoded,
columns=encoded_feature_names)

# concatenate the encoded dataframe with the original numerical dataframe for both
training and test sets
X_train_final =
X_train_encoded_df.join(X_train_scaled.reset_index(drop=True)[numeric_features])
X_test_final =
X_test_encoded_df.join(X_test_scaled.reset_index(drop=True)[numeric_features])

# create a CatBoostClassifier object
adaboost_clf = AdaBoostClassifier(random_state=random_seed, n_estimators=100,
learning_rate=0.1)

# fit the GridSearchCV object to the training data
adaboost_clf.fit(X_train_final, y_train)

y_pred = adaboost_clf.predict(X_test_final)
y_train_pred = adaboost_clf.predict(X_train_final)

```

```

# evaluate accuracy of the model on train and test data
test_accuracy = accuracy_score(y_test, y_pred)

print("Name of features", list(X_train.columns))
print("Number of features: {}\nAccuracy score on test data:
 {:.4f}\n".format(num_feature, test_accuracy))

print(classification_report(y_test, y_pred))

plt.figure(figsize=(10,6))
fx=sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, fmt=".2f", cmap="GnBu")
fx.set_title('Confusion Matrix \n');
fx.set_xlabel('\n Predicted Values\n')
fx.set_ylabel('Actual Values\n');
fx.xaxis.set_ticklabels(adaboost_clf.classes_)
fx.yaxis.set_ticklabels(adaboost_clf.classes_)
plt.show()

```

ประวัติผู้เขียน

ชื่อ นายศุภณัฐ ก้อนศิลา
 วัน เดือน ปีเกิด 22 ธันวาคม 2537
 ที่อยู่ปัจจุบัน 1/112 อ่อนนุช 59/1 เขตประเวศ แขวงประเวศ กรุงเทพมหานคร
 ประวัติการศึกษา (2556) เศรษฐศาสตรบัณฑิต สาขาเศรษฐศาสตร์
 (มหาวิทยาลัยศรีนครินทรวิโรฒ)
 ทุนการศึกษาที่ได้รับ ไม่มี
 ผลงานทางวิชาการ ไม่มี



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้